Tecnologia i serveis de preservació digital

XV Congrés D'Arxivística De Catalunya
28-30 Maig 2015, Lleida

José Carlos Ramalho
Dep. Informatics, University of Minho, jcr@di.uminho.pt
KEEP SOLUTIONS, jcr@keep.pt

KEEPSOLUTIONS
University of Minho SPIN-OFF

KEEP SOLUTIONS: Background

- DigitArq, CRAV (2003..[2008-2012])
- RODA (2006..[2008-…[)
- RCAAP (2008-…)
- PPA (2009)
- Open source: RODA, KOHA, DSpace, Moodle, etc.
- Scientific research
  - **SCAPE**: large scale digital preservation
  - **4C**: digital preservation costs prediction
  - **e-arK**: european reference model development

http://www.keep.pt

This work was partially supported by the SCAPE Project.
The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

What are we going to talk about?

Agenda (sort of…)

What is a Digital Object?

Any information object that can be represented by a binary digit sequence

Let’s take a look at some examples...

Science

KEEPSOLUTIONS
University of Minho SPIN-OFF

MAY 15, 1935 PHYSICAL REVIEW VOLUME 47

Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?

A. EINSTEIN, B. PODOLSKY AND N. ROSEN, *Institute for Advanced Study, Princeton, New Jersey*

(Received March 25, 1935)

In a complete theory there is an element corresponding to each element of reality. A sufficient condition for the reality of a physical quantity is the possibility of predicting it with certainty, without disturbing the system. In quantum mechanics in the case of two physical quantities described by non-commuting operators, the knowledge of one precludes the knowledge of the other. Then either (1) the description of reality given by the wave function in quantum mechanics is not complete or (2) these two quantities cannot have simultaneous reality. Consideration of the problem of making predictions concerning a system on the basis of measurements made on another system that had previously interacted with it leads to the result that if (1) is false then (2) is also false. One is thus led to conclude that the description of reality as given by a wave function is not complete.

1.

ANY serious consideration of a physical theory must take into account the distinction between the objective reality, which is independent of any theory, and the physical concepts with which the theory operates. These concepts are intended to correspond with the objective reality, and by means of these concepts we picture this reality to ourselves.

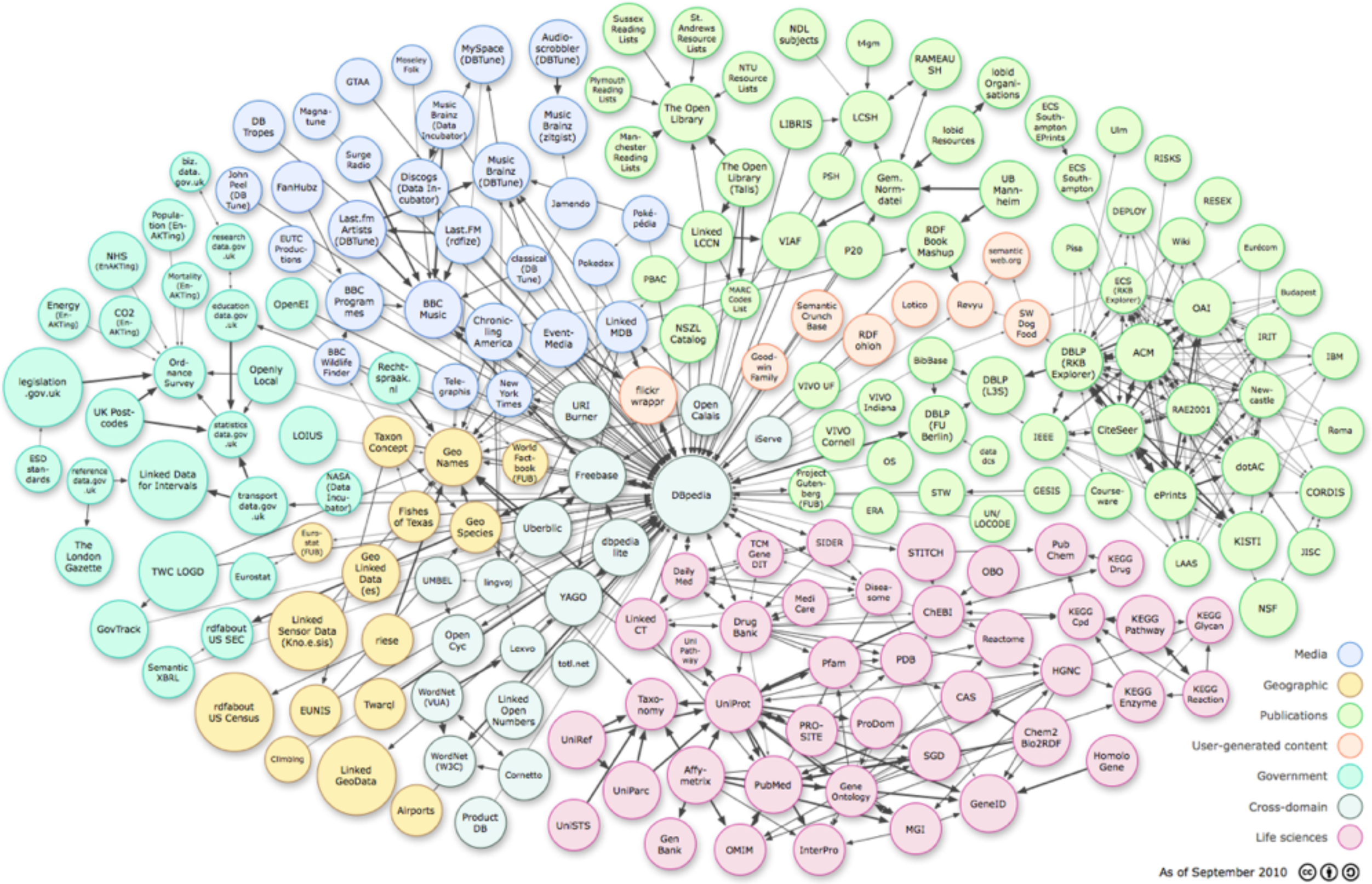
In attempting to judge the success of a physical theory, we may ask ourselves two questions: (1) "Is the theory correct?" and (2) "Is the description given by the theory complete?" It is only in the case in which positive answers may be given to both of these questions, that the concepts of the theory may be said to be satisfactory. The correctness of the theory is judged by the degree of agreement between the conclusions of the theory and human experience. Whatever the meaning assigned to the term *complete*, the following requirement for a complete theory seems to be a necessary one: *every element of the physical reality must have a counterpart in the physical theory*. We shall call this the condition of completeness. The second question is thus easily answered, as soon as we are able to decide what are the elements of the physical reality.

The elements of the physical reality cannot be determined by *a priori* philosophical considerations, but must be found by an appeal to results of experiments and measurements. A comprehensive definition of reality is, however, unnecessary for our purpose. We shall be satisfied with the following criterion, which we regard as reasonable. *If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical*

Universidade do Minho
Instituto de Educação e Psicologia

António Carlos Ribeiro da Silva

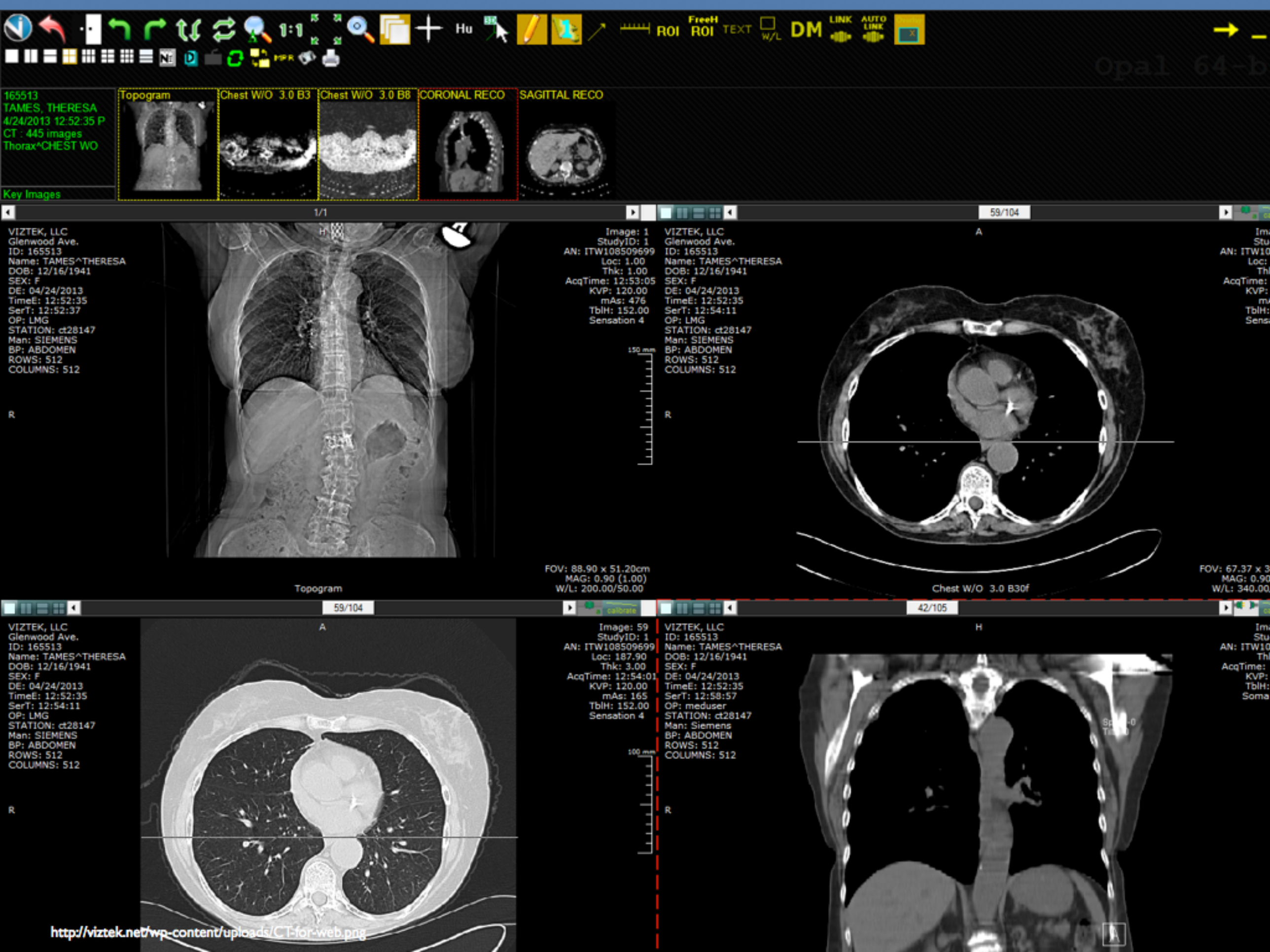
Abordagem Curricular por Competências no Ensino Superior: um estudo exploratório nos cursos de Administração, Ciências Contábeis e Economia no Estado da Bahia – Brasil


Sussex Reading Lists
St. Andrews Resource Lists
NDL subjects
tags
RAMEAU SH
lobid Organi-sations
NTU Resource Lists
Plymouth Reading Lists
The Open Library
LIBRIS
LCSH
lobid Resources
Man-chester Reading Lists
The Open Library (Talis)
PSH
Gem. Norm-datei
UB Mann-heim
ECS South-ampton EPrints
Ulm
RISKS
ECS South-ampton
DEPLOY
RESEX
Wiki
Eurécom
Pisa
Budapest
ECS (RKB Explorer)
OAI
IRIT
IBM
DBLP (RKB Explorer)
ACM
New-castle
RAE2001
CiteSeer
Roma
dotAC
CORDIS
ePrints
KISTI
JISC
LAAS
NSF
IEEE
GESIS
Course-ware
STW
UN/ LOCODE
ERA
OS
Project Guten-berg (FUB)
data dcs
VIVO Cornell
VIVO Indiana
VIVO UF
DBLP (FU Berlin)
DBLP (L3S)
BibBase
Revyu
SW Dog Food
semantic web.org
Lotico
RDF ohloh
Semantic Crunch Base
Good-win Family
RDF Book Mashup
P20
VIAF
Linked LCCN
MARC Codes List
PBAC
Poké-pédia
Pokedex
Jamendo
Music Brainz (zitgist)
Audio-scrobbler (DBTune)
MySpace (DBTune)
Music Brainz (DBTune)
Music Brainz (Data Incubator)
Moseley Folk
GTAA
Magna-tune
DB Tropes
Surge Radio
Discogs (Data In-cubator)
FanHubz
John Peel (DB Tune)
Last.fm Artists (DBTune)
Last.FM (rdfize)
EUTC Produc-tions
classical (DB Tune)
BBC Programmes
BBC Music
BBC Wildlife Finder
OpenEI
Chronic-ling America
Event-Media
Linked MDB
NSZL Catalog
New York Times
Tele-graphis
Recht-spraak.nl
flickr wrappr
URI Burner
Open Calais
Freebase
DBpedia
iServe
Geo Names
Taxon Concept
World Fact-book (FUB)
Fishes of Texas
Geo Species
Uberblic
dbpedia lite
NASA (Data Incu-bator)
Linked Data for Intervals
transport data.gov.uk
reference data.gov.uk
LOIUS
statistics data.gov.uk
Ord-nance Survey
Openly Local
legislation .gov.uk
UK Post-codes
ESD stan-dards
The London Gazette
TWC LOGD
Eurostat
Eurostat (FUB)
Geo Linked Data (es)
GovTrack
rdfabout US SEC
Linked Sensor Data (Kno.e.sis)
Semantic XBRL
rdfabout US Census
riese
EUNIS
Twarql
Climbing
Linked GeoData
Airports
UMBEL
lingvoj
YAGO
Open Cyc
WordNet (VUA)
WordNet (W3C)
Linked Open Numbers
Lexvo
totl.net
Cornetto
Product DB
Linked CT
Daily Med
TCM Gene DIT
Drug Bank
Uni Path-way
Medi Care
Disea-some
SIDER
STITCH
ChEBI
OBO
Pub Chem
KEGG Drug
KEGG Cpd
KEGG Pathway
KEGG Glycan
KEGG Enzyme
KEGG Reaction
Reactome
HGNC
Pfam
PDB
CAS
Chem2 Bio2RDF
Homolo Gene
Taxo-nomy
UniProt
PRO-SITE
ProDom
SGD
UniRef
UniParc
UniSTS
Affy-metrix
PubMed
Gene Ontology
GeneID
MGI
Gen Bank
OMIM
InterPro
Energy (En-AKTing)
CO2 (En-AKTing)
NHS (EnAKTing)
Popula-tion (En-AKTing)
Mortality (En-AKTing)
biz. data. gov.uk
research data.gov.uk
education data.gov.uk
Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences
As of September 2010

Medicine

KEEPSOLUTIONS
University of Minho SPIN-OFF

http://www.almaveterinaria.pt/admin/imagens_links/20111206183 5-imagiologia3.jpg


ROI
ROI
TEXT
W/L
DM
LINK
AUTO LINK
165513
TAMES, THERESA
4/24/2013 12:52:35 P
CT : 445 images
Thorax^CHEST WO
Key Images
Topogram
Chest W/O 3.0 B3
Chest W/O 3.0 B8
CORONAL RECO
SAGITTAL RECO
1/1
59/104
VIZTEK, LLC
Glenwood Ave.
ID: 165513
Name: TAMES^THERESA
DOB: 12/16/1941
SEX: F
DE: 04/24/2013
TimeE: 12:52:35
SerT: 12:52:37
OP: LMG
STATION: ct28147
Man: SIEMENS
BP: ABDOMEN
ROWS: 512
COLUMNS: 512
R
Image: 1
StudyID: 1
AN: ITW108509699
Loc: 1.00
Thk: 1.00
AcqTime: 12:53:05
KVP: 120.00
mAs: 476
TblH: 152.00
Sensation 4
150 mm
FOV: 88.90 x 51.20cm
MAG: 0.90 (1.00)
W/L: 200.00/50.00
VIZTEK, LLC
Glenwood Ave.
ID: 165513
Name: TAMES^THERESA
DOB: 12/16/1941
SEX: F
DE: 04/24/2013
TimeE: 12:52:35
SerT: 12:54:11
OP: LMG
STATION: ct28147
Man: SIEMENS
BP: ABDOMEN
ROWS: 512
COLUMNS: 512
R
A
Chest W/O 3.0 B30f
59/104
42/105
VIZTEK, LLC
Glenwood Ave.
ID: 165513
Name: TAMES^THERESA
DOB: 12/16/1941
SEX: F
DE: 04/24/2013
TimeE: 12:52:35
SerT: 12:54:11
OP: LMG
STATION: ct28147
Man: SIEMENS
BP: ABDOMEN
ROWS: 512
COLUMNS: 512
R
A
Image: 59
StudyID: 1
AN: ITW108509699
Loc: 187.90
Thk: 3.00
AcqTime: 12:54:01
KVP: 120.00
mAs: 165
TblH: 152.00
Sensation 4
100 mm
VIZTEK, LLC
ID: 165513
Name: TAMES^THERESA
DOB: 12/16/1941
SEX: F
DE: 04/24/2013
TimeE: 12:52:35
SerT: 12:58:57
OP: meduser
STATION: ct28147
Man: Siemens
BP: ABDOMEN
ROWS: 512
COLUMNS: 512
R
H
http://viztek.net/wp-content/uploads/CT-for-web.png

http://www.hospitalveterinario.pt/site/upload/image/Noticias/tomografia6.jpg

FileMed - Michael A. O'Donovan, MD

File Edit Tools Appointments Help License

- New Patient
- Demographics
- Clinical Chart
- Images
- Network Configuration
- Export Database
- Backup
- Exit

filemed

Demographics

Last Name, Maiden: Watts
First Name, MI: Jennifer A
Record#: 000001
Social Security #: 23784598
D.O.B. mm/dd/yyyy: 03/08/1967
Current age: 39 y.
Sex: F
Marital St.: MAR
Occupation: Teacher
Nationality: american
Insurance/Coverage: Blue Cross/Blue Shield
Insurance ID: 47815879
Address: 7235 SW 48th St
City: Miami
State/Province: FL
Zip/Postal Code: 33155
Phone: 305-666-5599
Fax: 305-666-5560
Mobile/Pager: 305-666-5015
Email: jenwatts@uol.net
Referring Physician: Dr W. Garland
Attending Physician: Dr. Herman Stewart
Date of First Visit: 07/15/2004

Photo
Add
Delete
Change

Diagnoses

ICD-9CM | ICD-10 | Paste

11/25/2004: Allergic rhinitis | Nasal polyps | Acute sinusitis.

Remove Pt | Print | Save | Cancel

Patient: Watts, Jennifer A | User: Administrator | Version 5.1 - Multiuser

http://www.suggestsoft.com/images/varsur-corporation/filemed-electronic-medical-records-emr.gif

KEEPSOLUTIONS

Engineering

KEEPSOLUTIONS
University of Minho SPIN-OFF

AutoSketch - [Home2.SKF]

File Edit View Tools Draw Inquire Database Window Help

Repeat Line Multiple F4
Line
Arc
Circle
Polyline
Polygon
Curve
Marker
Text
Dimension
Duplicate
Fill
Picture
Detail View
Guidelines
Symbol

Center, Side
Side, Side
3 Points
Center, Radius
Tangent 2 Entities
Tangent 3 Entities
Ellipse Rectangle
Ellipse Axes

Building

93%

64.04'
24.00'
24.00'
12.11'
36.66'
24.57'
12.00'
18.00'
6.57'
36.57'
24.00'
25.00'
15.00'
64.00'

Content Librarian
_ArchitecturalSampler.s
BayWindow
CabinetBase...
CabinetBase...
CabinetWall2Dr
CabinetWallC...
CarElev
CarPlan
Chair
ChairRoundB...
Compactor
DataScop_b
DeskChair
Dishwshr23lx...
Dishwshr24lx...
Dishwshr30lx...
Door2f10iSw...
Door2f4iSwg...
Door2f6iSwg...
Door2f6iSwg...
Door2f6iSwg...
Door2f8iSwg...
Door2f8iSwg...
Door2f8iSwg...
Door2fFolding
Door2fSwgL30
Door2fSwgR90
Door2PanelExt
Door2PanelInt

http://www.cadalyst.com/files/cadalyst/nodes/2005/2910/i2.jpg

Draws a circle based on the centerpoint and the radius

dsPIC Controller

U$4
SN74LVC1G3157DBVR

U$3
FT232RSSOP28

IC1

DSPIC30F4013/P

Use the onboard clock for the first version.

Programming Port

RJ11

U$2
RJ11-6N-B

Serial Port

IC3
MAX232

Full Bridge Rectifier Protected Power Input

Switching 5V Supply for PIC

U$1
LM2675-5.0

http://web.mit.edu/neltnerb/www/artwork/supply-schematic-2.png

TITLE: powersupply_v3	
Document Number:	REV:
Date: 1/25/2008 15:52:25	Sheet: 2/2

Eclipse File Edit Run Source Refactor Navigate Search Project Window Help

blapp – Java – android.graphics.Canvas – Eclipse – /Users/bill/Desktop/android/eclipse/workspace

Debug Java Br

Project Explorer

- DragViewV2
  - src
    - com.blahti.example.drag2
      - DragActivityV2.java
      - DragController.java
      - DragLayer.java
      - DragSource.java
      - DragView.java
      - DropSpot.java
      - DropTarget.java
      - MyAbsoluteLayout.java
  - gen [Generated Java Files]
  - Android 2.2
    - android.jar – /Users/bill/Des
      - android
      - android.accessibilityservi
      - android.accounts
      - android.app
      - android.app.admin
      - android.app.backup
      - android.appwidget
      - android.bluetooth
      - android.content
      - android.content.pm
      - android.content.res
      - android.database
      - android.database.sqlite
      - android.gesture
      - android.graphics
        - AvoidXfermode.class
        - Bitmap.class
        - BitmapFactory.class
        - BitmapShader.class
        - BlurMaskFilter.class
        - Camera.class
        - Canvas.class

Outline

- Canvas
  - Canvas()
  - Canvas(Bitmap)
  - Canvas(GL)
  - getGL() : GL
  - freeGlCaches() : void
  - setBitmap(Bitmap) : void
  - setViewport(int, int) : void
  - isOpaque() : boolean
  - getWidth() : int
  - getHeight() : int
  - getDensity() : int
  - setDensity(int) : void
  - MATRIX_SAVE_FLAG : int
  - CLIP_SAVE_FLAG : int
  - HAS_ALPHA_LAYER_SAVE_
  - FULL_COLOR_LAYER_SAVE_
  - CLIP_TO_LAYER_SAVE_FLA
  - ALL_SAVE_FLAG : int
  - save() : int
  - save(int) : int
  - saveLayer(RectF, Paint, int)
  - saveLayer(float, float, float
  - saveLayerAlpha(RectF, int,
  - saveLayerAlpha(float, float
  - restore() : void
  - getSaveCount() : int

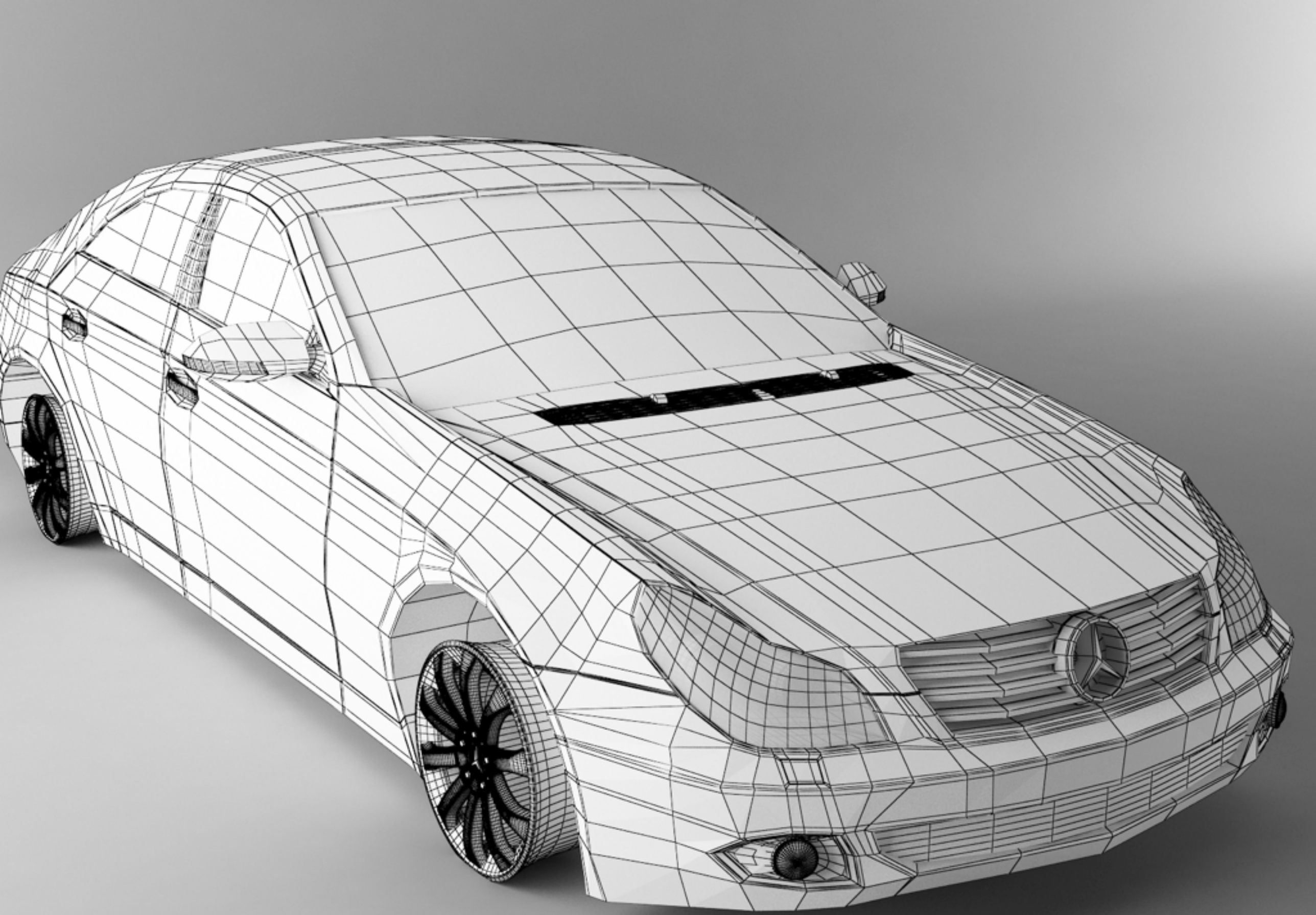
DragLayer.java | View.class | Account.class | ViewGroup.class | Canvas.class

```
/* Copyright (C) 2006 The Android Open Source Project

package android.graphics;

import android.text.TextUtils;
import android.text.SpannedString;
import android.text.SpannableString;
import android.text.GraphicsOperations;
import android.util.DisplayMetrics;

import javax.microedition.khronos.opengles.GL;

/**
 * The Canvas class holds the "draw" calls. To draw something, you need
 * 4 basic components: A Bitmap to hold the pixels, a Canvas to host
 * the draw calls (writing into the bitmap), a drawing primitive (e.g. Rect,
 * Path, text, Bitmap), and a paint (to describe the colors and styles for the
 * drawing).
 */
public class Canvas {
    // assigned in constructors, freed in finalizer
    final int mNativeCanvas;

    /*  Our native canvas can be either a raster, gl, or picture canvas.
        If we are raster, then mGL will be null, and mBitmap may or may not be
        present (our default constructor creates a raster canvas but no
        java-bitmap is). If we are a gl-based, then mBitmap will be null, and
        mGL will not be null. Thus both cannot be non-null, but its possible
        for both to be null.
    */
    private Bitmap  mBitmap;    // if not null, mGL must be null
    private GL      mGL;        // if not null, mBitmap must be null
```

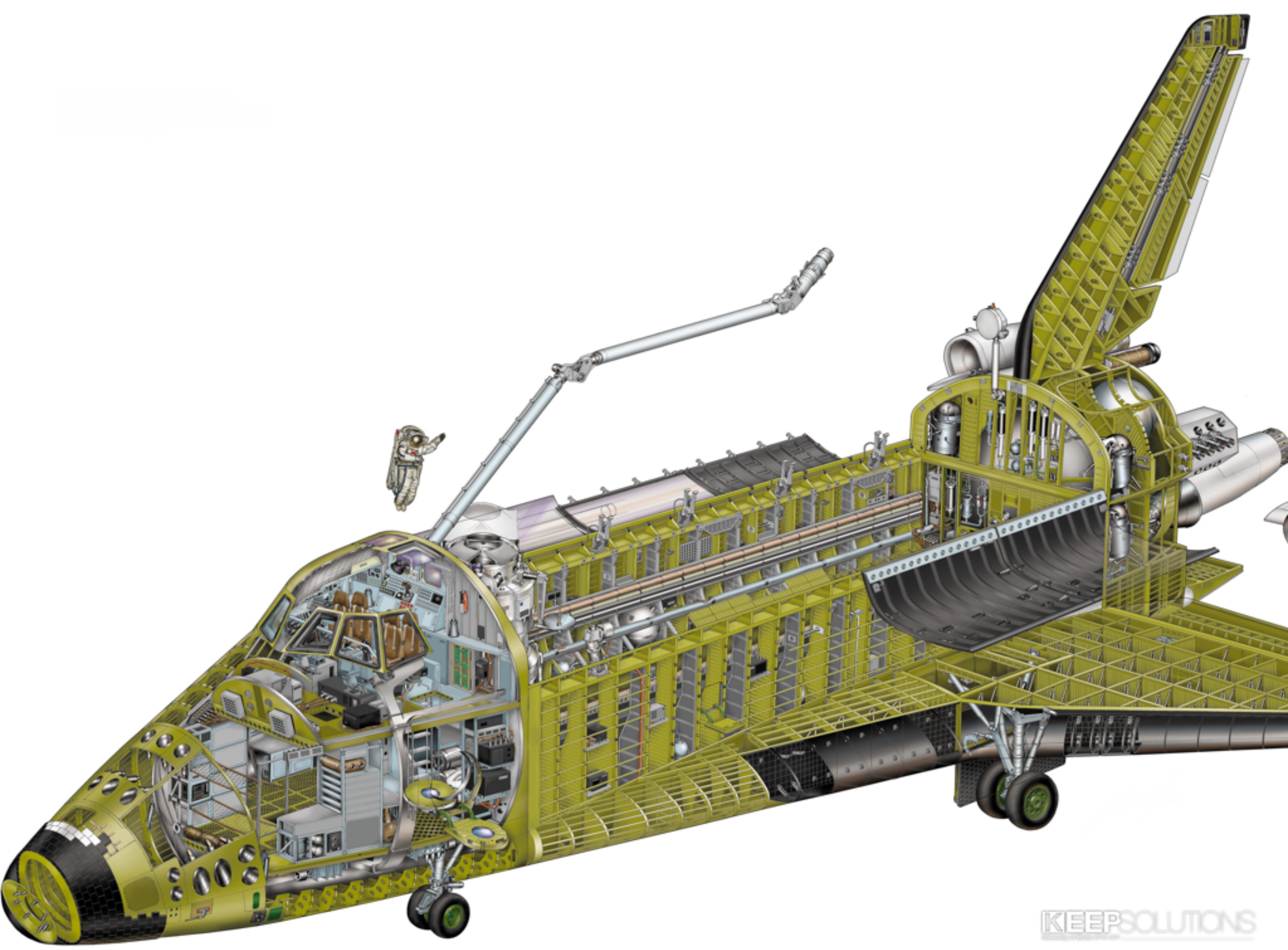
Problems | Javadoc | Declaration | Task List | Console | Search | JUnit

0 items

Description	Resource	Path



http://fc03.deviantart.net/fs71/f/2012/069/d/9/mercedes_cls_wireframe_render__by_ctl3d-d4sakwp.jpg



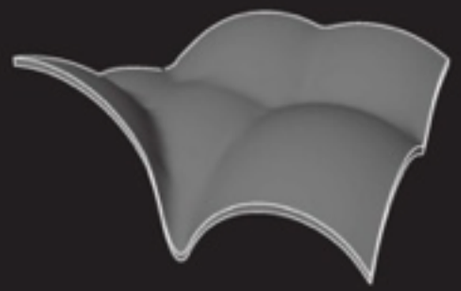
Art

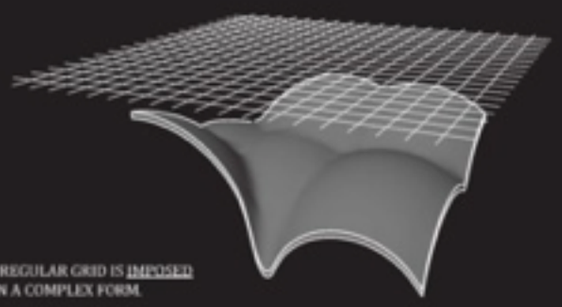
http://upload.wikimedia.org/wikipedia/commons/f/fd/An-Najah_University_media_room_Victor_2011_-1-70.jpg

THESIS ISSUE AND PROPOSAL
THE DISCRETISATION GRID OF COMPLEX FORMS

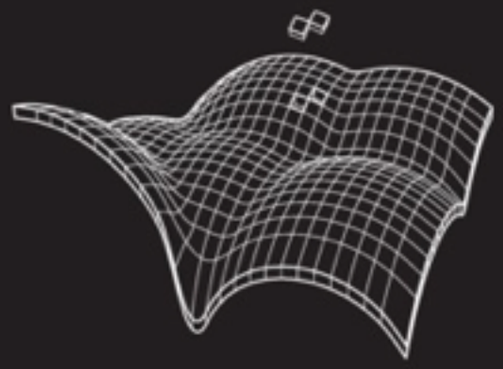
DISCRETISATION:
'...the necessity to decompose a
continuous geometric object into discrete elements.'

-AKT



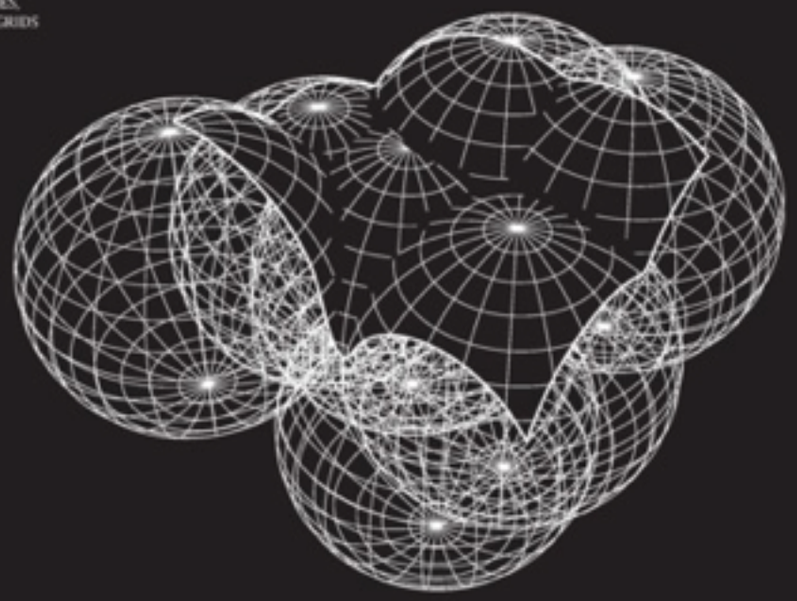


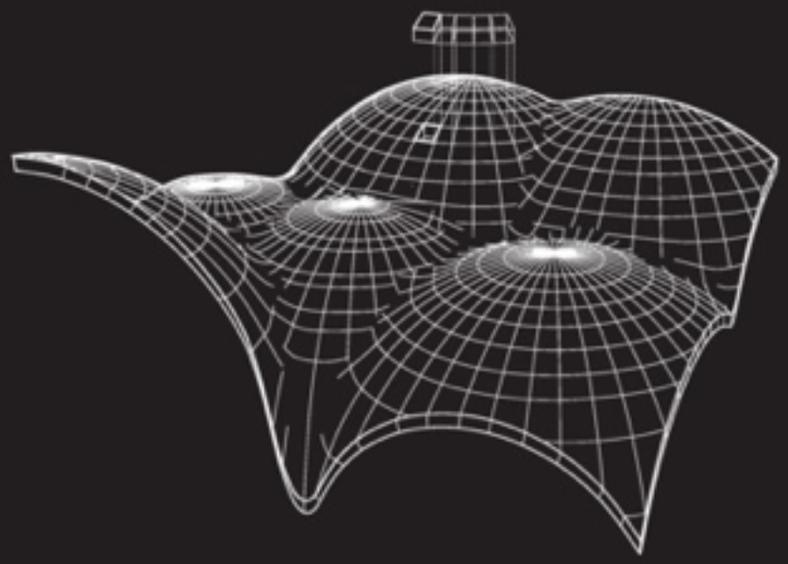
A REGULAR GRID IS IMPOSED
ON A COMPLEX FORM.



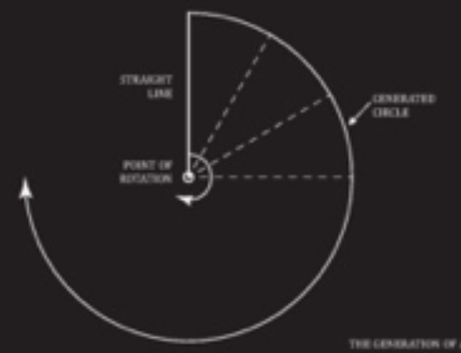
THESIS PROPOSAL:
THE GEOMETRICAL GRID

THE APPROXIMATION OF A COMPLEX SURFACE TO
A SERIES OF CIRCULAR GEOMETRIES.
'ADOPTING' THEIR DISCRETISING GRIDS
TO GENERATE ITS OWN.

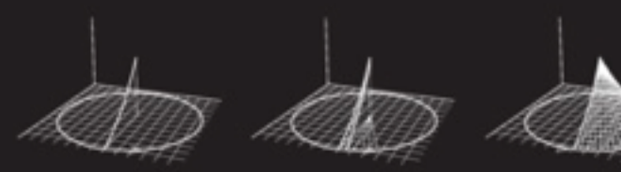




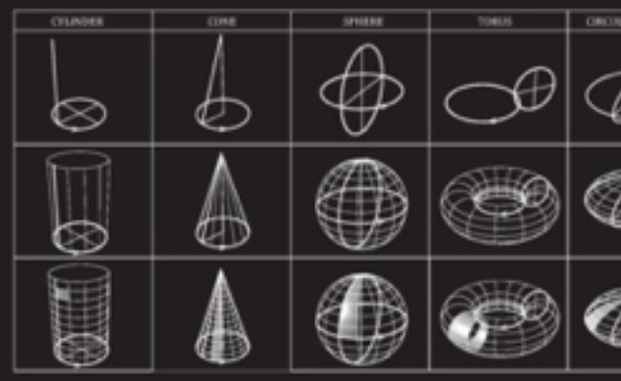
CIRCULAR GEOMETRIES:
COMPLEX SURFACES GENERATED FROM GEOMETRICAL LINES. NAMELY:
THE STRAIGHT LINE
THE CIRCLE AND
THE ELLIPSE, WITH THE CIRCLE IN THE HORIZONTAL AXIS



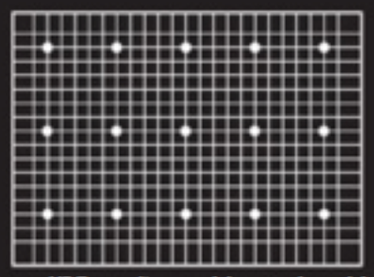
THE GENERATION OF A CI



THE CIRCLE AS THE HORIZ



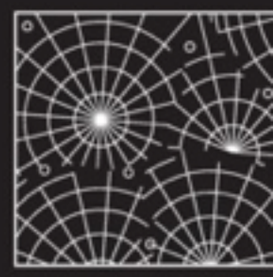
A MATRIX OF SURFACES GENER
THE CIRCLE AS THE HORIZ



THE REGULAR GRID SUGGESTS:
1 PANELS
2 A STRUCTURAL SYSTEM AND
3 SPATIAL ARRANGEMENT.

BUT:
1 ALL PANELS ARE UNIQUE
2 THE STRUCTURAL SYSTEM MAY BECOME SEPARATE FROM THE FORM
3 THE COMPLEXITY OF THE FORM IS NOT REFLECTED IN THE SPATIAL ARRANGEMENT.

THE GEOMETRICAL GRID ALLOWS:
1 A FIXED NUMBER OF MODULE TYPES
2 AN INTEGRATION OF FORM AND STRUCTURE
3 AN INTEGRATION OF SPACE AND FORM



http://25.media.tumblr.com/tumblr_m3vcwfrSeH1qzc0peo1_1280.jpg

http://cryptocrimson.com/wp-content/uploads/2014/05/aaart.jpg

http://www.creativeapplications.net/wp-content/uploads/2009/07/bodypaint.png

Industry and services

http://www.muntthee.com/wp-content/uploads/2011/10/JFE-III-Film.jpg

KEEPSOLUTIONS
University of Minho SPIN-OFF

http://www.capcom-central.com/concept/re4/LeonWireFrame.jpg

http://imagens2.publico.pt/imagens.aspx/780492?tp=UH&db=IMAGENS

http://www.agenciaproweb.com.br/clientes/fmnovotempo/images/radio-studio-2.jpg

http://pro.jvc.com/pro/pr/2009/images/TV_Worldwide_2.jpg

Memory institutions

KEEPSOLUTIONS
University of Minho SPIN-OFF

http://upload.wikimedia.org/wikipedia/commons/0/01/Dunhuang_manuscript_digitisation.jpg

SQLQuery116.sql - MISIA.master (sa (52))* - Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Help

New Query | master | Execute | Debug

Object Explorer

Connect

- dbo.TipoMedida
- dbo.TipoNivel
- dbo.TipoNivelRelacionado
- dbo.TipoNivelRelacionadoCodigo
- dbo.TipoNoticiaATipoControloAForma
- dbo.TipoNoticiaAut
- dbo.TipoOperation
- dbo.TipoOrdenacao
- dbo.TipoPertinencia
- dbo.TipoServer
- dbo.TipoSubDensidade
- dbo.TipoSuporte
- dbo.TipoTecnicaRegisto
- dbo.TipoTecnicasDeRegisto
- dbo.TipoTipologias
- dbo.TipoTradicaoDocumental
- dbo.Trustee
- dbo.TrusteeDepositoPrivilege
- dbo.TrusteeGroup
- dbo.TrusteeNivelPrivilege
- dbo.TrusteeObjetoDigitalPrivilege
- dbo.TrusteePrivilege
- dbo.TrusteeUser
- dbo.UserGroups
- dbo.WebClientActivity
- Views
- Synonyms
- Programmability
- Service Broker
- Storage
- Security
- ReportServer
- ReportServerTempDB
- Security
- Server Objects
- Replication
- AlwaysOn High Availability
- Management
- Integration Services Catalogs
- SQL Server Agent

SQLQuery117.sql - MISIA.master (sa (53)) | SQLQuery116.sql - MISIA.master (sa (52))* | SQLQuery115.sql -...Archeevo3 (sa (59))*

```
/****** Script for SelectTopNRows command from SSMS  ******/
SELECT TOP 1000 IDControloAut
IDTrusteeOperator
IDTrusteeAuthority
DataEdicao
DataAutoria
Versao
isDeleted
IDTipoNoticiaAut
Importacao
  FROM GISA.dbo.ControloAutDataDeDescricao
  WHERE DataEdicao<>DataAutoria

  SELECT TOP 100 *
  FROM GISA.dbo.ControloAutDataDeDescricao
  WHERE IDTrusteeOperator IS NOT NULL
```

100 %

Results | Messages

	IDControloAut	IDTrusteeOperator	IDTrusteeAuthority	DataEdicao	DataAutoria	Versao	isDeleted	IDTipoNoticiaAut
1	1	13	NULL	2008-01-21 14:32:28.000	2008-01-21 14:32:28.000	0x0000000000E42986	0	4
2	1	13	NULL	2008-01-21 14:38:39.000	2008-01-21 14:38:39.000	0x0000000000E42987	0	4
3	1	13	NULL	2008-01-22 09:19:48.000	2008-01-22 09:19:48.000	0x0000000000E42988	0	4
4	1	14	NULL	2008-03-14 10:07:05.000	2008-03-14 10:07:05.000	0x0000000000E42989	0	4
5	1	14	NULL	2008-09-10 12:08:10.000	2008-09-10 12:08:10.000	0x0000000000E4298A	0	4
6	1	14	NULL	2008-09-15 15:55:23.000	2008-09-15 15:55:23.000	0x0000000000E4298B	0	4
7	1	14	NULL	2008-09-15 16:00:19.000	2008-09-15 16:00:19.000	0x0000000000E4298C	0	4
8	1	14	NULL	2009-09-18 14:23:10.000	2009-09-18 14:21:45.000	0x0000000000E4298D	0	4
9	1	14	NULL	2013-12-12 11:50:46.000	2013-12-12 11:13:16.953	0x0000000001C52C14	0	4
10	3593	14	NULL	2010-09-06 14:51:53.000	2010-09-06 14:35:54.843	0x0000000000E4298E	0	1
11	3594	6	NULL	2011-10-21 18:07:25.293	2011-10-21 18:07:25.293	0x0000000001068CBF	0	5
12	3595	6	NULL	2011-10-21 18:07:25.293	2011-10-21 18:07:25.293	0x000000000106784F	0	1
13	3596	6	NULL	2011-10-21 18:07:25.293	2011-10-21 18:07:25.293	0x0000000001067850	0	1
14	3597	14	14	2008-09-10 15:21:08.000	2008-09-10 15:21:08.000	0x0000000000E4298F	0	5
15	3597	14	NULL	2013-06-04 12:02:53.000	2013-06-04 11:49:14.187	0x00000000019147A6	0	5
16	3598	6	NULL	2011-10-21 18:07:25.293	2011-10-21 18:07:25.293	0x0000000001067851	0	1

Query executed successfully.

Output

Show output from:

Ready

LOS CATALANES ACUDIRÁN A LAS URNAS EN OTOÑO

Mas adelanta las elecciones catalanas al 27 de septiembre

• El presidente catalán afirma que no habrá lista única soberanista, pero sí "una hoja de ruta compartida"

El presidente Mas durante su comparecencia. / AFP (ATLAS)

CLARA GIL DEL OLMO / PERE RÍOS | Barcelona

Mas ha anunciado este miércoles, tras reunirse durante cuatro horas con el líder de ERC, que se "ha rehecho el pacto de unidad entre los partidos y las entidades" para culminar el proceso soberanista "hasta la victoria, hasta ganar". Nacionalistas y republicanos concurrirán a las urnas por separado, aunque con "una hoja de ruta compartida que en este momento ya está muy avanzada y que se incorporará a los programas electorales". La campaña comenzará el 11 de septiembre, día de la Diada

LA DISTRIBUCIÓN DE FUERZAS EN CATALUÑA

› Así quedó el Parlament tras las elecciones de 2012

Catalonian parliamentary election, 2015

José Carlos | Página inicial

Sobre **Catalonian parliamentary election, 2015**

Gosto | Guardar | Editar

Interesse

A página Elecciones catalanas de 2015 foi fundida com esta [?]

0 pessoas gostam deste tópico

A **parliamentary election** is scheduled to be held in Catalonia, as announced on January 15, 2015 by President Artur Mas, on September 27 that same year.

The election will take place among the Catalan independence movement, that started in 2012. President Mas announced that it was his intention to turn the election into a...

Continuar a ler

De Wikipédia, a enciclopédia grátis · Editar na Wikipédia

Esta Página é gerada automaticamente com base naquilo em que os utilizadores do Facebook estão interessados e que

Home

KEEPSOLUTIONS
University of Minho SPIN-OFF

http://i.huffpost.com/gen/1713160/thumbs/o-WEDDING-VIDEO-facebook.jpg


KEEPSOLUTIONS


iPod
Music
Videos
Photos
Podcasts
Extras
Settings
Shuffle Songs
Now Playing
DAY & AGE
MENU
9:41 AM
Music
iPod
9:41 AM
FaceTime
Calendar
Photos
Camera
YouTube
Stocks
Maps
Weather
Notes
Utilities
iTunes
App Store
Game Center
Settings
Music
Mail
Safari
Videos

Why has the world surrendered to digital ?

KEEPSOLUTIONS
University of Minho SPIN-OFF

Easy to share, to transmit, to disseminate/distribute
specially using Internet

Easy to create and to edit
using authoring or editing software

Requires small amount of physical storage space
when compared to analogical artifacts

It can be duplicated without quality loss
a digital duplicate is exactly equal to its original

Avoids original object manipulation
within a scanning context it promotes the preservation of the tangible object

However...

It needs a technological context in order to be consumed

Susceptible to the fast technology obsolescence

Information can become prisoner of its own supporters and / or formats

What changes in a digital context?


TER24AGO

EDIÇÃO PORTO

PÚBLICO

www.publico.pt

SINÓNIMOS E ANTÓNIMOS

Dicionário

COM MAIS 10 VOLUMES

TRÊS MILHÕES POR DISTRIBUIR ÀS VÍTIMAS DOS INCÊNDIOS DE 2003

Verbas das campanhas de solidariedade do ano passado podem ir para casos deste ano

A joalheira de Phnom Penh

CULTURA

Polícia à procura do DNA dos ladrões dos quadros de Munch

Assessor de Rui Rio ganha mais que Jorge Sampaio

Finanças com dúvidas no reembolso de 40 milhões de IVA ao BCP

FUTEBOL

Gilberto Madail admite chamada de Derlei à selecção nacional

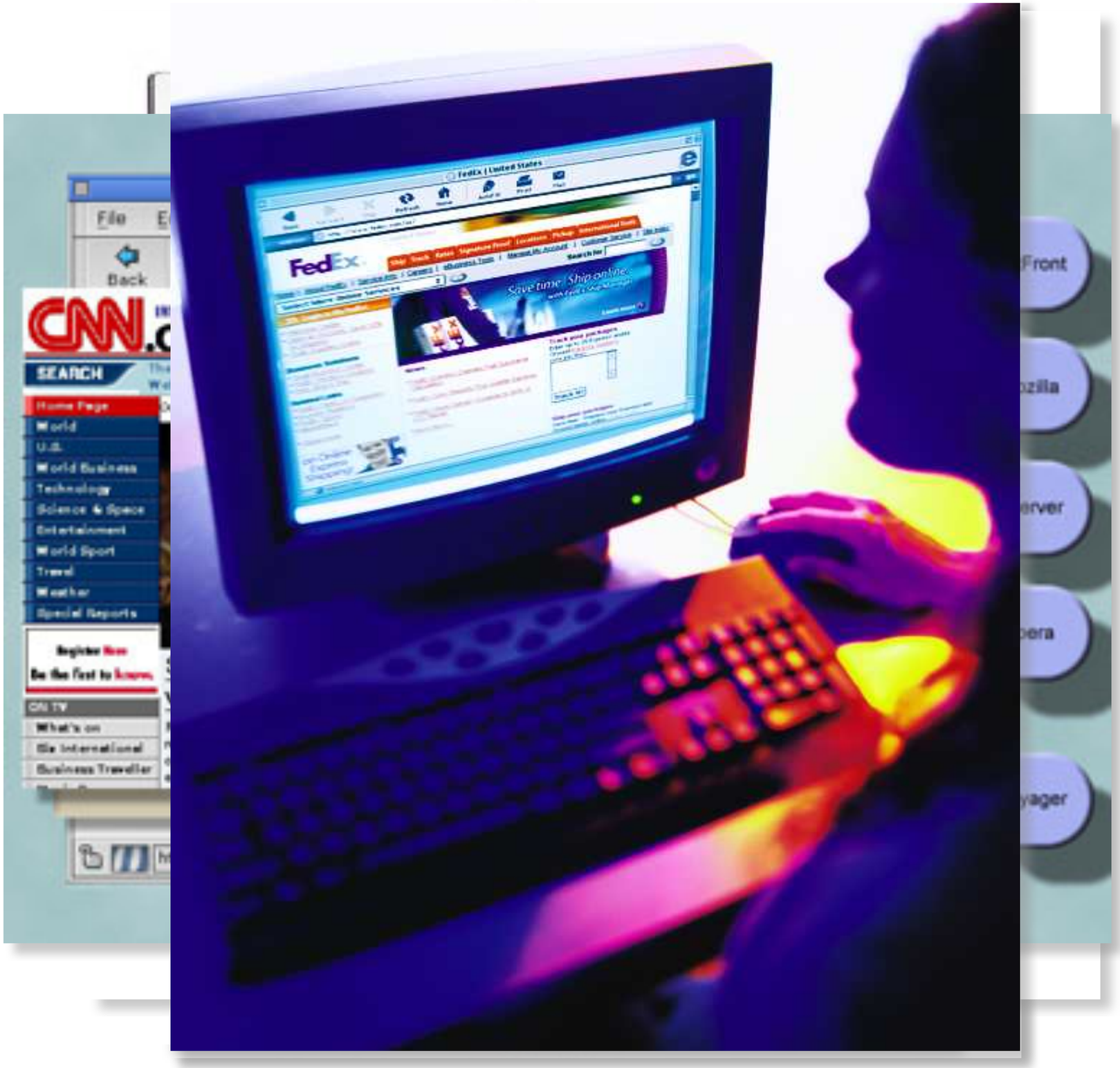
1809

Britânica Kelly Holmes acaba com reinado de Maria Mutola nos 800 metros

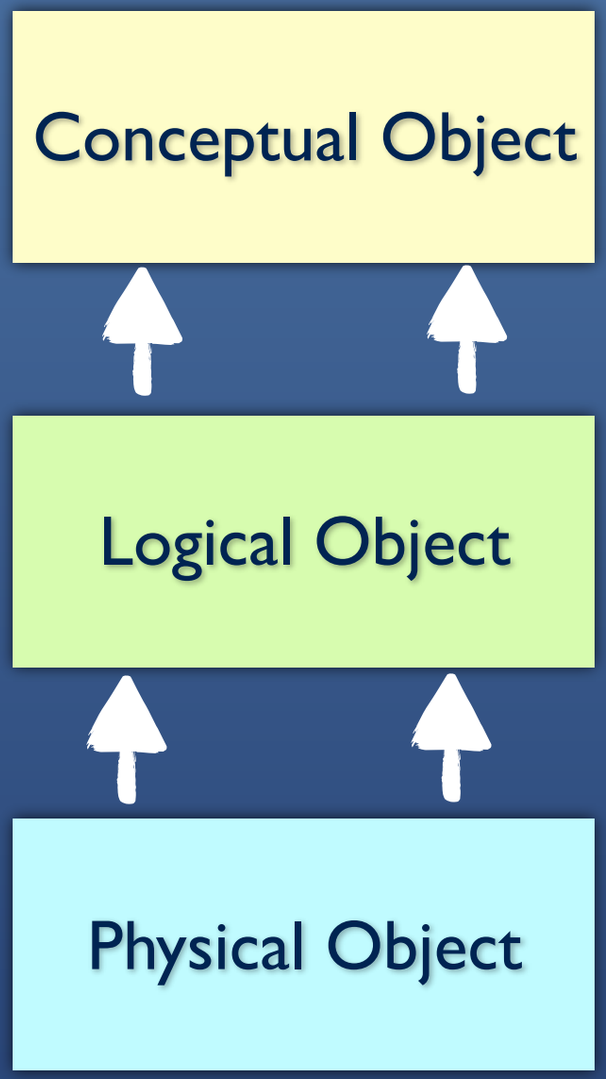
Francis Obikwelu recebeu a medalha de prata e diz-se cansado para os 200 metros

Edivaldo Monteiro sem ilusões para as meias-finais dos 400 metros barreiras

Como Angola defrontou os gigantes norte-americanos do basquetebol


FedEx | United States
FedEx
Save time. Ship online.
CNN
SEARCH
Home Page
World
U.S.
World Business
Technology
Science & Space
Entertainment
World Sport
Travel
Weather
Special Reports
ON TV
Back
Front

Anatomy of a digital object



Conceptual Level

What we deal with in the real world, i.e., what we recognize as a tangible unit of information,
e.g. journal, book, map, photo

Logical Level

Information codification.
Software generates a specific format for the artifact

Physical Level

Physical signs engraved upon some physical support
e.g. hard drive, CD, Pen-drive

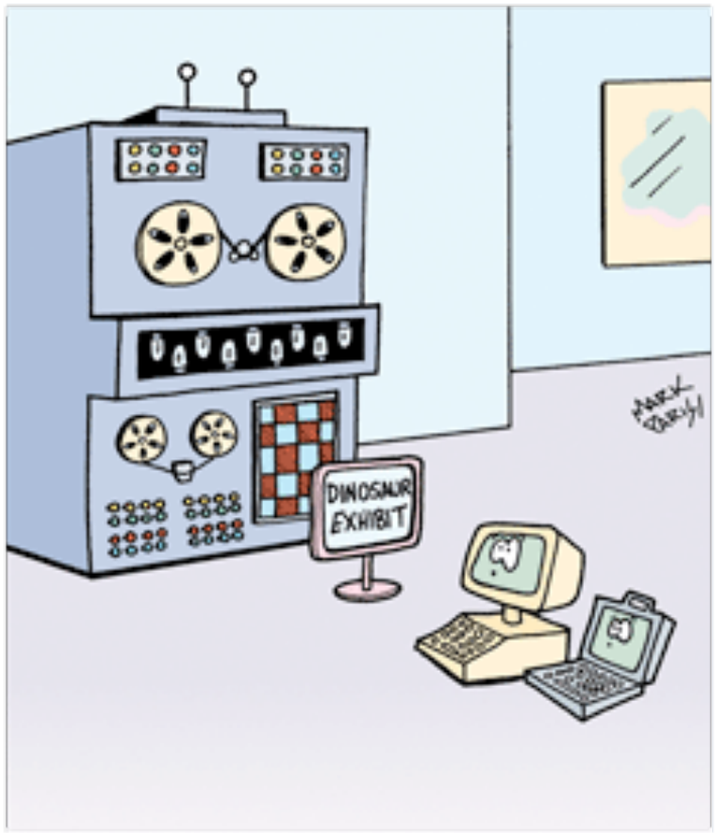
... if any of these levels becomes obsolete we lose access to the conceptual object ...

...in less then 5 years, at least one of these levels will be replaced due to technological obsolescence...

KEEPSOLUTIONS
University of Minho SPIN-OFF

Technological obsolescence. category

where a product falls when it ceases to be useful, even if it is in perfect working order, due to the emergence of a more technologically advanced product.



Software obsolescence

Object interpretation context

Hardware obsolescence

Computing systems that support software

Support or format obsolescence

Information trapped in the support or format

Threat examples

Hard disk malfunction

20 % of discs crash within the first 4 years of life

Threat examples

Hard disk malfunction

20 % of discs crash within the first 4 years of life

Computer has no diskette drive

Some laptops no longer have a CD drive

Threat examples

Hard disk malfunction

20 % of discs crash within the first 4 years of life

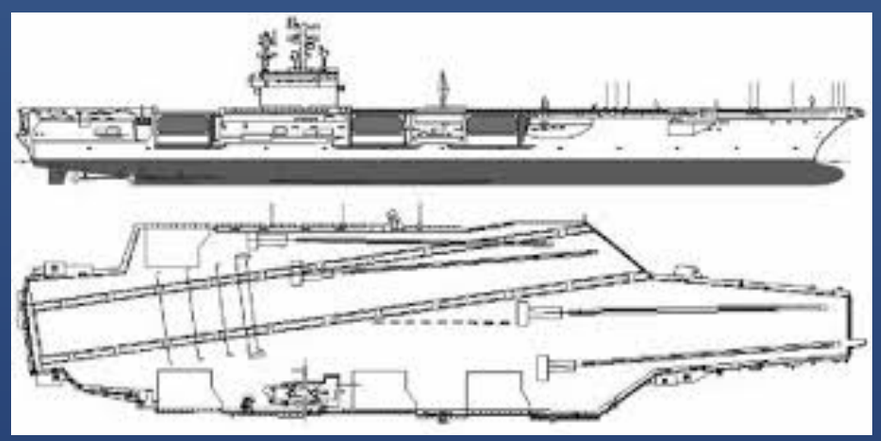
Computer has no diskette drive

Some laptops no longer have a CD drive

Email service ceases to exist

In 2011, Google was on the edge to loose all email from 35.000 users

Threat examples



Hard disk malfunction

20 % of discs crash within the first 4 years of life

Computer has no diskette drive

Some laptops no longer have a CD drive

Email service ceases to exist

In 2011, Google was on the edge to loose all email from 35.000 users

Formats become obsolete

500.000 diagrams of USS Carrier could not be read após after CAD software upgrade

Motivation: why should we preserve digital information?

KEEPSOLUTIONS
University of Minho SPIN-OFF

Public sector

Legal imposition
(e.g. legal deposit, compulsory incorporations)

Heritage and culture preservation

Transparency

Public activity evidence

Private sector

Legal imposition
(e.g. audits and inspections, finances, funded projects, etc.)

Commercial reasons / profit

Protection against possible disputes

Business continuity

What can we do?

KEEPSOLUTIONS
University of Minho SPIN-OFF

Ensure that access to information does not depend on any specific technology

KEEPSOLUTIONS
University of Minho SPIN-OFF

Ensure that the communication process is effective not only through space but also through time

KEEPSOLUTIONS
University of Minho SPIN-OFF

Ensure that the communication process is effective
not only through space but also through time

Time interoperability

KEEPSOLUTIONS
University of Minho SPIN-OFF

Ok... but what can we really do?

We can identify our information assets

What type/kind of information do we need to preserve?
What are the related formats?
What is the retention period of each asset?

We can determine what preservation level we want to reach

What is our goal or SLA?
What object properties do we really need to preserve?
What restrictions should we regard?

We can implement an effective risk management

What threats can we identify?
What are the odds of those to happen?
What would be the impact if they occurred?
What kind of controls have we implemented to mitigate these risks?

To Plan

Try several strategies
Select the best strategy to achieve intended goals
Define goals and milestones

To Act

Execute all operations needed to implement selected policies

To Control quality

Are the actions being executed having the expected outcomes?

To Monitor external environment

We need to identify internal and external risks so we can react in due time to threats
We also need to know/identify opportunities

To Audit

Using the services of external entities to verify our processes

…this set of activities is called digital preservation

KEEPSOLUTIONS
University of Minho SPIN-OFF

Digital Preservação. set of activities that aim to ensure continuous access to digital information

KEEPSOLUTIONS
University of Minho SPIN-OFF

…it involves technology, persons, financial resources...

Example

Digital Object Lifecycle inside a digital preservation environment

KEEPSOLUTIONS
University of Minho SPIN-OFF

Digital objects

KEEPSOLUTIONS
University of Minho SPIN-OFF

Digital objects

- Digital Repositories
- Web content
- Scientific data

→ **Characterization**
(format identification, property extraction)

KEEPSOLUTIONS
University of Minho SPIN-OFF

Environment Monitoring
(Internal and external)

Digital objects

Digital Repositories

Web content

Scientific data

Characterization
(format identification, property extraction)

KEEPSOLUTIONS
University of Minho SPIN-OFF

Planning
(decision on what action to take)

Environment Monitoring
(Internal and external)

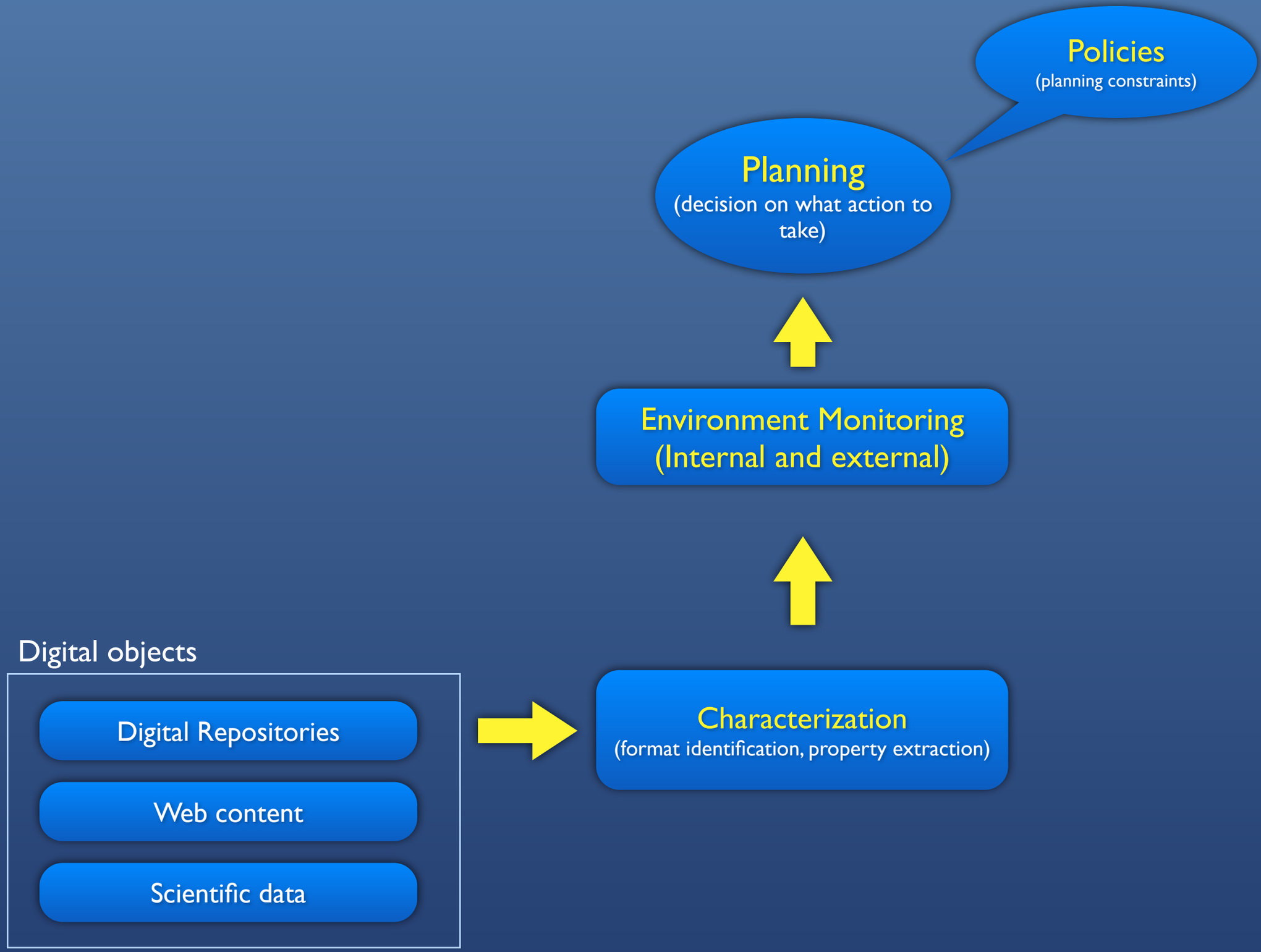
Digital objects

Digital Repositories

Web content

Scientific data

Characterization
(format identification, property extraction)


Policies
(planning constraints)
Planning
(decision on what action to take)
Environment Monitoring
(Internal and external)
Digital objects
Digital Repositories
Web content
Scientific data
Characterization
(format identification, property extraction)

Policies
(planning constraints)

Preservation actions
(Format migration)

Planning
(decision on what action to take)

Environment Monitoring
(Internal and external)

Digital objects

Digital Repositories

Web content

Scientific data

Characterization
(format identification, property extraction)

Policies
(planning constraints)

Preservation actions
(Format migration)

Planning
(decision on what action to take)

Quality control
(property comparison between properties extracted from original and converted object)

Environment Monitoring
(Internal and external)

Digital objects

Digital Repositories

Web content

Scientific data

Characterization
(format identification, property extraction)

KEEPSOLUTIONS
University of Minho SPIN-OFF

Policies
(planning constraints)

Planning
(decision on what action to take)

Preservation actions
(Format migration)

Quality control
(property comparison between properties extracted from original and converted object)

Environment Monitoring
(Internal and external)

Digital objects

Digital Repositories

Web content

Scientific data

Characterization
(format identification, property extraction)

KEEPSOLUTIONS
University of Minho SPIN-OFF

How can we define preservation policies?

Organizational level

General goals for the preservation of digital collections , ie organizational goals

- The authenticity of digital objects must be preserved.

Procedural level

Approaches that the organization will take to achieve top level goals.

- We must document all processes in the life cycle of a digital object with sufficient detail for current and future users to understand the context of the existence of the object itself

Control level

Rules and requirements that will allow to verify if defined approaches are being followed and goals achieved.

- All preservation events SHOULD be registered in PREMIS format
- We must capture all diplomatic elements to assess the reason for being of an object

Digital Preservation Strategies

Skip

KEEPSOLUTIONS
University of Minho SPIN-OFF

Hardware museums

- Original context preservation

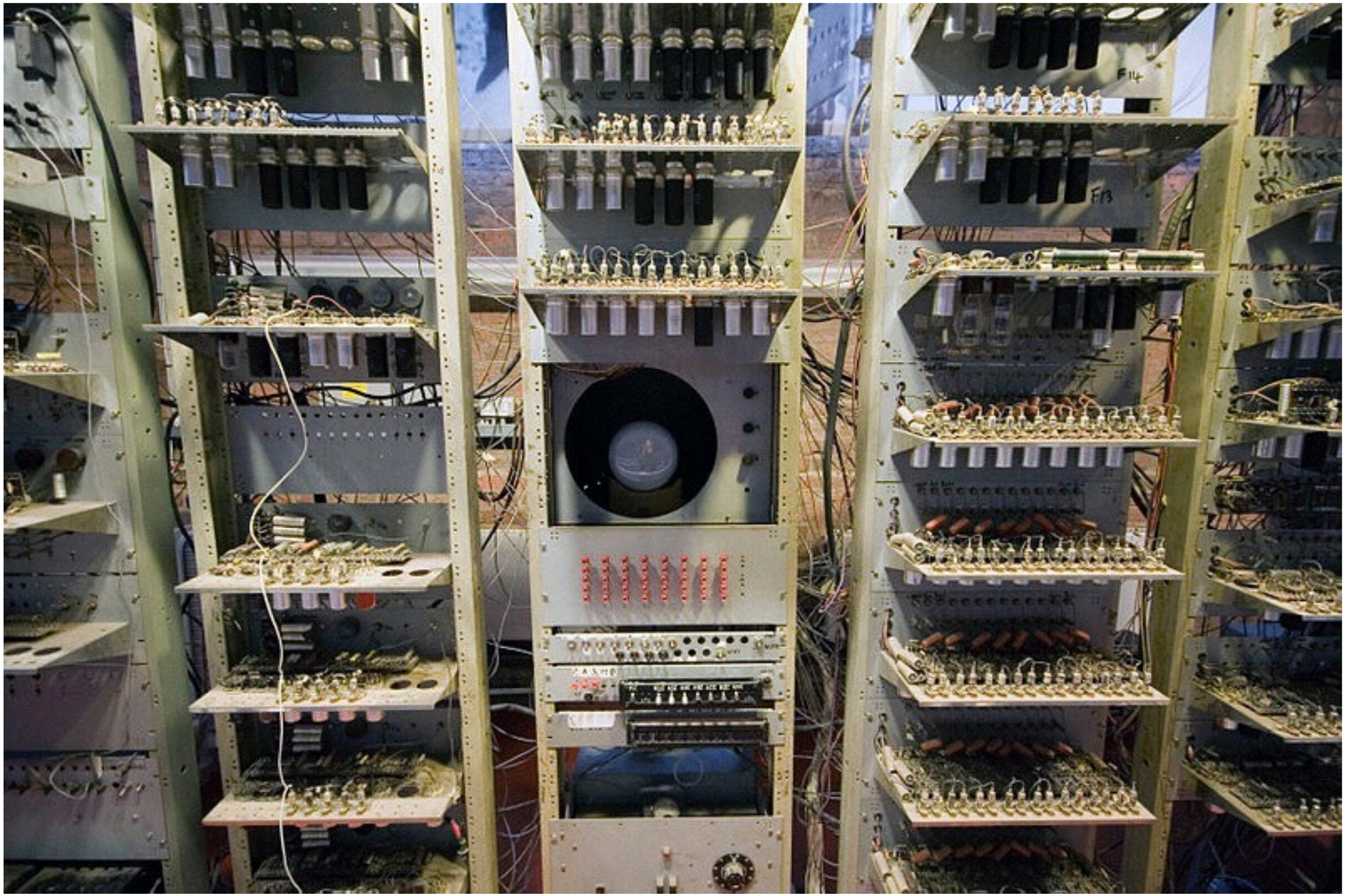
Advantages

- We can reproduce the object as it was originally created

Disadvantages

- Hard to maintain in the longterm (costs will increase with time)
- Access level restrictions (poor reuse rate)
- Someone has to know how to operate long gone systems

First computer to be able to store instructions electronically



Emulation

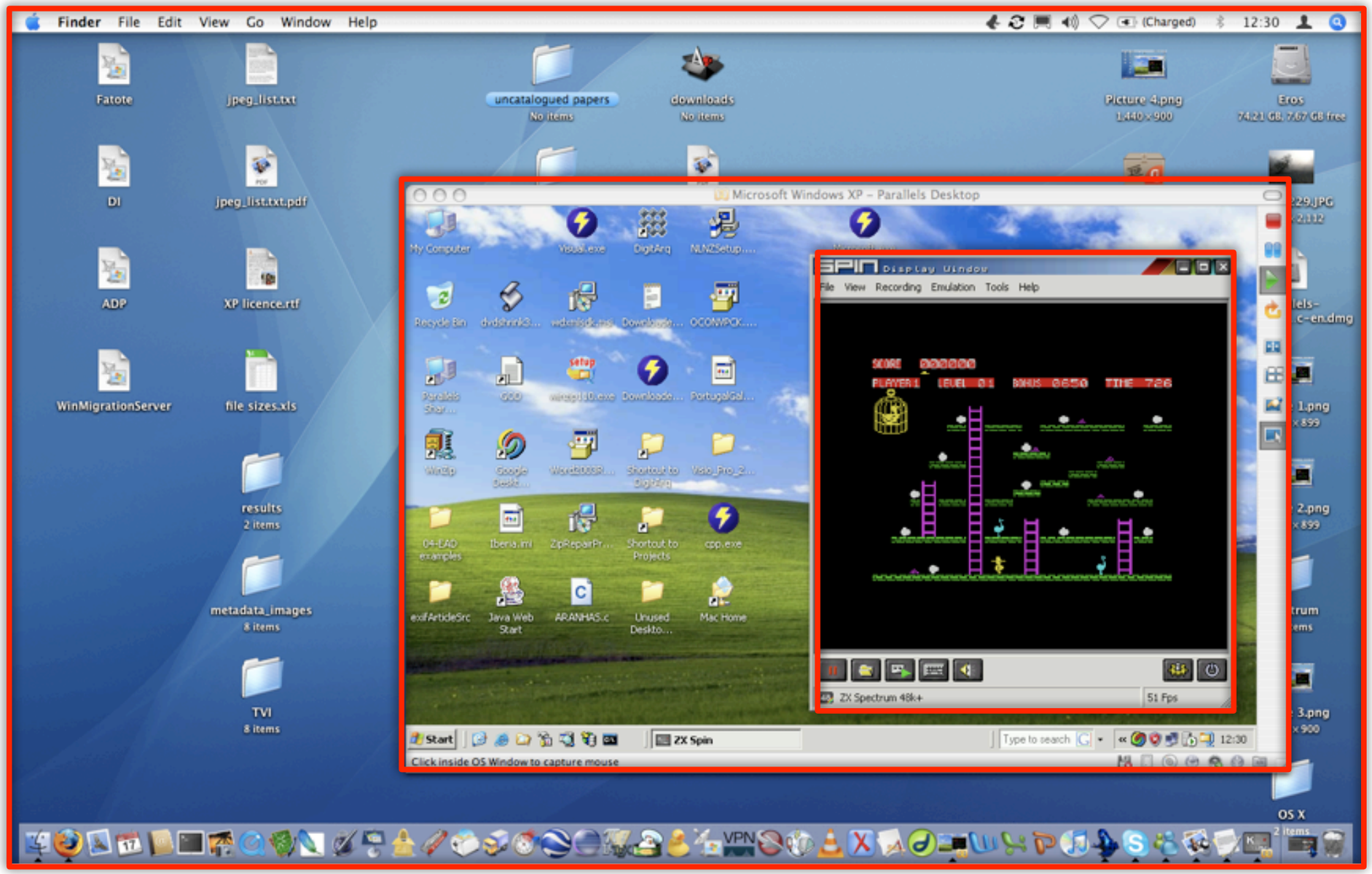
Use of software that is able to reproduce the behavior of an hardware platform and / or software on a different one

Advantages

- It does not need original hardware

Disavantages

- Emulator development is a complex process that depends on information from hardware/ software suppliers
- Information reuse in not guaranteed
- Knowledge on how to operate long gone technologies
- Emulator may become obsolescent


Finder File Edit View Go Window Help
(Charged) 12:30
Fatote
jpeg_list.txt
uncatalogued papers
No items
downloads
No items
Picture 4.png
1,440 × 900
Eros
74.21 GB, 7.67 GB free
DI
jpeg_list.txt.pdf
ADP
XP licence.rtf
WinMigrationServer
file sizes.xls
results
2 items
metadata_images
8 items
TVI
8 items
OS X
2 items
Microsoft Windows XP – Parallels Desktop
ZX Spin
Click inside OS Window to capture mouse
ZX Spectrum 48k+
51 Fps

Encapsulation

Original bit stream preservation along with documentation in order to enable future development of viewers , migrators or emulators

Advantages

- It allows to postpone preservation actions
- It is targeted to objects that will be accessed in a far future

Disadvantages

- Complex objects have complex specifications
- Incomplete specifications could have a disastrous effect

Format Migration

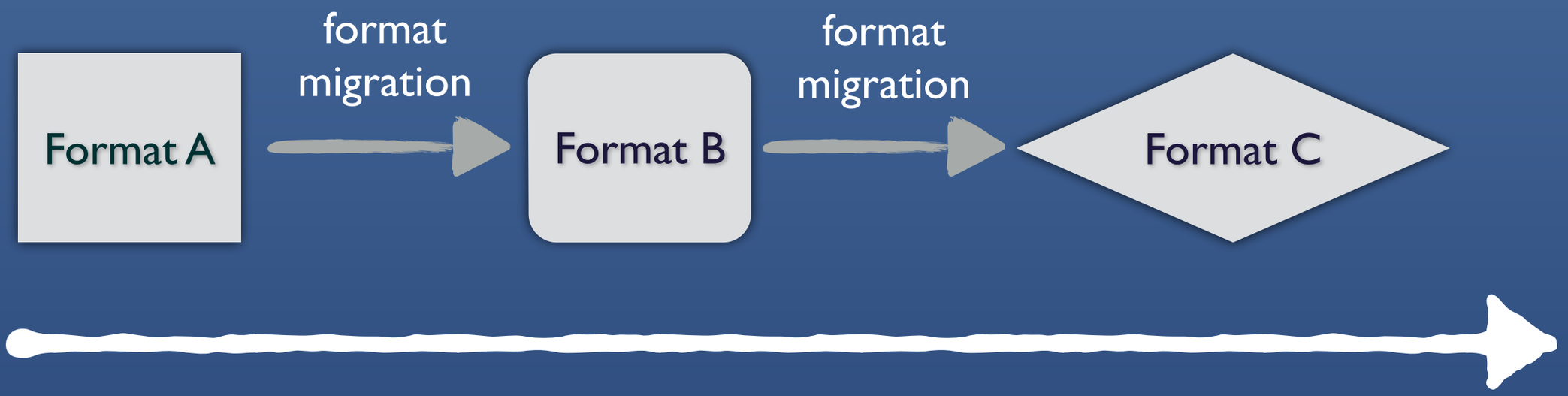
Digital material transfer from one hardware/software configuration to a more current one

Advantages

- Dissemination of materials on current formats that consumers can easily identify and process
- There is no need to preserve original software
- Most used strategy in all domains (home, organizational, etc.)

Disadvantages

- Possible loss of information during migration
- Continued diligence is needed (watch)
- Expensive in the longterm



However...

None of these strategies will work if we don't refresh media support in due time

How can we characterize our collections?

Characterization is to identify the attributes that make up a digital object

Some of these attributes are important to keep object integrity

We call these attributes significant properties

In this context...

integrity preservation **is not a synonym** of document binary sequence preservation!

It means to preserve significant properties what will allow us to say/prove that the document is as the original was

Format identification

To Identify the file format based on document's internal structure
e.g. application/pdf, version 1.4, fmt/18

Property extraction

Calculating object's properties values
e.g. number of pages: 14, image width: 300px, compression: LZW

Compliance validation

File compliance check against identified format specification.
e.g. according to PDF 1.4 file is valid.

FITS - File Information Tool Set

http://fitstool.org

```
<?xml version="1.0" encoding="UTF-8"?>
<fits xmlns="http://hul.harvard.edu/ois/xml/ns/fits/fits_output" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://hul.harvard.edu/ois/xml/ns/fits/fits_output
http://hul.harvard.edu/ois/xml/xsd/fits/fits_output.xsd" version="0.8" timestamp="10/15/14 11:35 AM">
  <identification>
    <identity format="Acrobat PDF 1.4 - Portable Document Format" mimetype="application/pdf" toolname="FITS"
toolversion="0.8">
      <tool toolname="fido" toolversion="v1.3.1" />
      <tool toolname="Droid" toolversion="6.1.3" />
      <tool toolname="Jhove" toolversion="1.5" />
      <tool toolname="NLNZ Metadata Extractor" toolversion="3.4GA" />
      <tool toolname="Tika" toolversion="1.3" />
      <tool toolname="file utility" toolversion="5.04" />
      <version toolname="Droid" toolversion="6.1.3">1.4</version>
      <externalIdentifier toolname="fido" toolversion="v1.3.1" type="puid">fmt/18</externalIdentifier>
    </identity>
  </identification>
  <fileinfo>
    <size toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">7186034</size>
    <creatingApplicationName toolname="Jhove" toolversion="1.5">Acrobat Distiller 7.0.5 for Macintosh/Acrobat:
pictwpstops filter 1.0</creatingApplicationName>
    <lastmodified toolname="Tika" toolversion="1.3" status="SINGLE_RESULT">2006-11-24T11:33:58Z</lastmodified>
  </fileinfo>
  <metadata>
    <document>
      <title toolname="Jhove" toolversion="1.5">Introducao à preservacao digital - Conceitos, estratégias e
actuais consensos</title>
      <author toolname="Jhove" toolversion="1.5">Miguel Ferreira</author>
      <pageCount toolname="Jhove" toolversion="1.5">88</pageCount>
      <isTagged toolname="Jhove" toolversion="1.5">no</isTagged>
      <hasOutline toolname="Jhove" toolversion="1.5">no</hasOutline>
      <hasAnnotations toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">no</hasAnnotations>
      <hasForms toolname="NLNZ Metadata Extractor" toolversion="3.4GA" status="SINGLE_RESULT">no</hasForms>
      <isProtected toolname="NLNZ Metadata Extractor" toolversion="3.4GA"
status="SINGLE_RESULT">no</isProtected>
      <subject toolname="Tika" toolversion="1.3" status="SINGLE_RESULT">Preservação Digital</subject>
    </document>
  </metadata>
  <filestatus>
    <well-formed toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">true</well-formed>
    <valid toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">true</valid>
  </filestatus>
</fits>
```

KEEPSOLUTIONS
University of Minho SPIN-OFF

FITS - File Information Tool Set

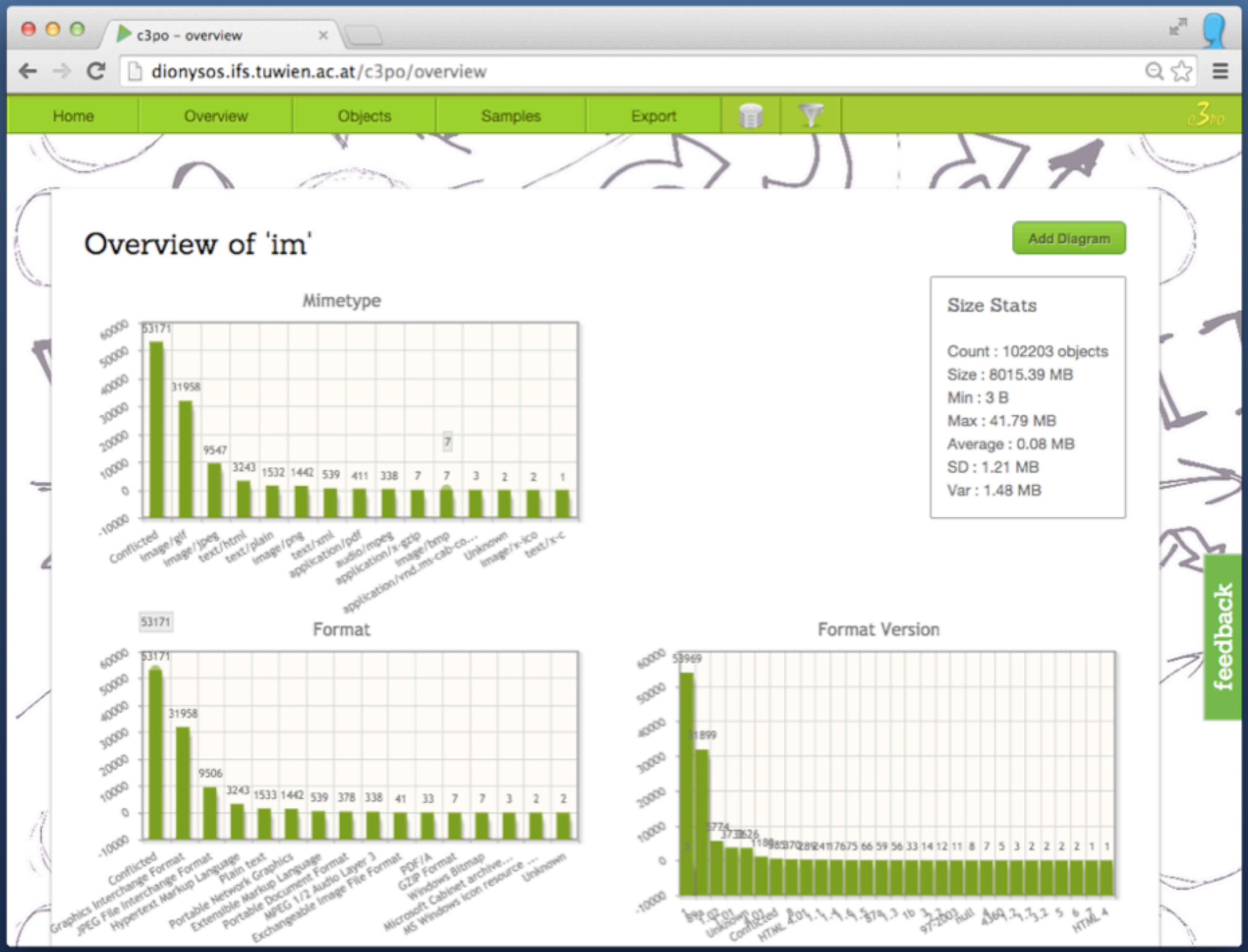
http://fitstool.org

```
<?xml version="1.0" encoding="UTF-8"?>
<fits xmlns="http://hul.harvard.edu/ois/xml/ns/fits/fits_output" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://hul.harvard.edu/ois/xml/ns/fits/fits_output
http://hul.harvard.edu/ois/xml/xsd/fits/fits_output.xsd" version="0.8" timestamp="10/15/14 11:35 AM">
  <identification>
    <identity format="Acrobat PDF 1.4 - Portable Document Format" mimetype="application/pdf" toolname="FITS"
toolversion="0.8">
      <tool toolname="fido" toolversion="v1.3.1" />
      <tool toolname="Droid" toolversion="6.1.3" />
      <tool toolname="Jhove" toolversion="1.5" />
      <tool toolname="NLNZ Metadata Extractor" toolversion="3.4GA" />
      <tool toolname="Tika" toolversion="1.3" />
      <tool toolname="file utility" toolversion="5.04" />
      <version toolname="Droid" toolversion="6.1.3">1.4</version>
      <externalIdentifier toolname="fido" toolversion="v1.3.1" type="puid">fmt/18</externalIdentifier>
    </identity>
  </identification>
  <fileinfo>
    <size toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">7186034</size>
    <creatingApplicationName toolname="Jhove" toolversion="1.5">Acrobat Distiller 7.0.5 for Macintosh/Acrobat:
pictwpstops filter 1.0</creatingApplicationName>
    <lastmodified toolname="Tika" toolversion="1.3" status="SINGLE_RESULT">2006-11-24T11:33:58Z</lastmodified>
  </fileinfo>
  <metadata>
    <document>
      <title toolname="Jhove" toolversion="1.5">Introducao à preservacao digital - Conceitos, estratégias e
actuais consensos</title>
      <author toolname="Jhove" toolversion="1.5">Miguel Ferreira</author>
      <pageCount toolname="Jhove" toolversion="1.5">88</pageCount>
      <isTagged toolname="Jhove" toolversion="1.5">no</isTagged>
      <hasOutline toolname="Jhove" toolversion="1.5">no</hasOutline>
      <hasAnnotations toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">no</hasAnnotations>
      <hasForms toolname="NLNZ Metadata Extractor" toolversion="3.4GA" status="SINGLE_RESULT">no</hasForms>
      <isProtected toolname="NLNZ Metadata Extractor" toolversion="3.4GA"
status="SINGLE_RESULT">no</isProtected>
      <subject toolname="Tika" toolversion="1.3" status="SINGLE_RESULT">Preservação Digital</subject>
    </document>
  </metadata>
  <filestatus>
    <well-formed toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">true</well-formed>
    <valid toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">true</valid>
  </filestatus>
</fits>
```

KEEPSOLUTIONS
University of Minho SPIN-OFF

C3PO - Clever, Crafty Content Profiling of Objects

http://ifs.tuwien.ac.at/imp/c3po



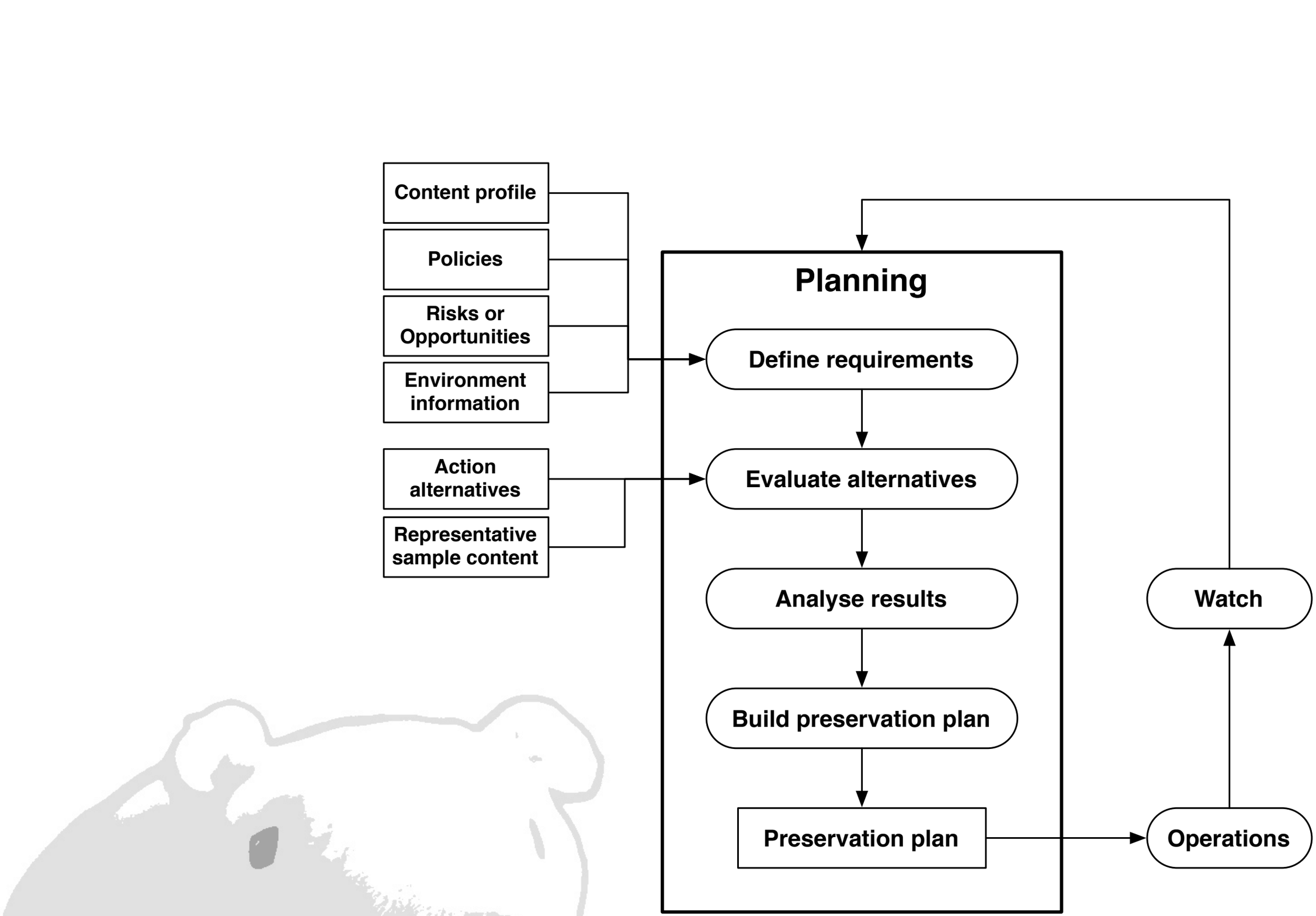
KEEPSOLUTIONS
University of Minho SPIN-OFF

How can we plan?

PLATO

Preservation Planning Tool

http://www.ifs.tuwien.ac.at/dp/plato


Content profile
Policies
Risks or Opportunities
Environment information
Action alternatives
Representative sample content
Planning
Define requirements
Evaluate alternatives
Analyse results
Build preservation plan
Preservation plan
Watch
Operations

How can we act?

Before ingestion

Manually imposing ingestion policies
Automatically using software

After ingestion

Using specialized software or some functionality available in the information management software

When accessing

Keeping information as it was ingested and only act when requested

README.md

Digital Preservation Toolkit

Installing Digital Preservation Tools was never so easy.

What does Digital Preservation Toolkit do?

The Digital Preservation Toolkit is a Debian metapackage, called digital-preservation-tools, and allows to easily install Digital Preservation tools. It does that by referencing three other metapackages:

- Migration (digital-preservation-tools-migration)
- Characterisation (digital-preservation-tools-characterization)
- Quality Assurance (digital-preservation-tools-quality-assurance)

Using these metapackages one may install all the tools (by installing digital-preservation-tools) or a specific set of related tools (by installing the one of the digital-preservation-tools-* packages). Each of the metapackages (Migration, Characterisation and Quality Assurance) references tools wrapped with SCAPE Toolwrapper, in the form of Debian packages.

Using it

Pre-requisites

Debian/Ubuntu system.

https://github.com/openplanets/digital-preservation-toolkit

How can we control quality?

Object A

A properties

Object A’

A’ properties

Quality control

e.g. xCorrSound, ffprobe, Matchbox, MarcAlizer, Jpylyzer

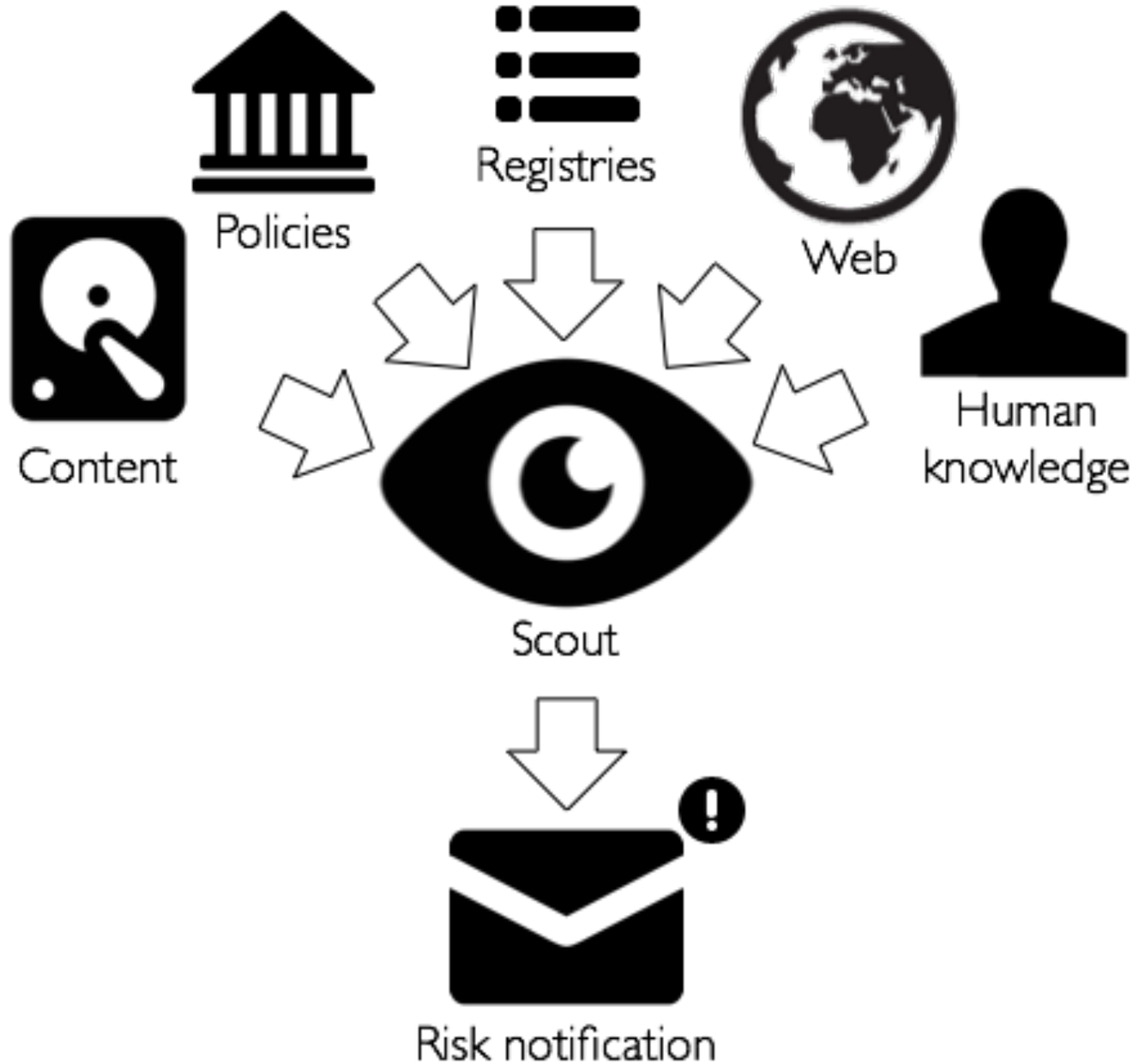
Similarity

http://www.scape-project.eu/tools

Media Type	Tool Name
Audio	xcorrSound
Documents	MS QA Tools
Document Collections	Matchbox
Web Pages	Pagelyzer
Images	Jpylyzer
Images	PhotoHawk
Images	Dissimilar
Images	JP2Check
PDFs / EPubs	DRMLint

Table 1 Overview of QA Tools

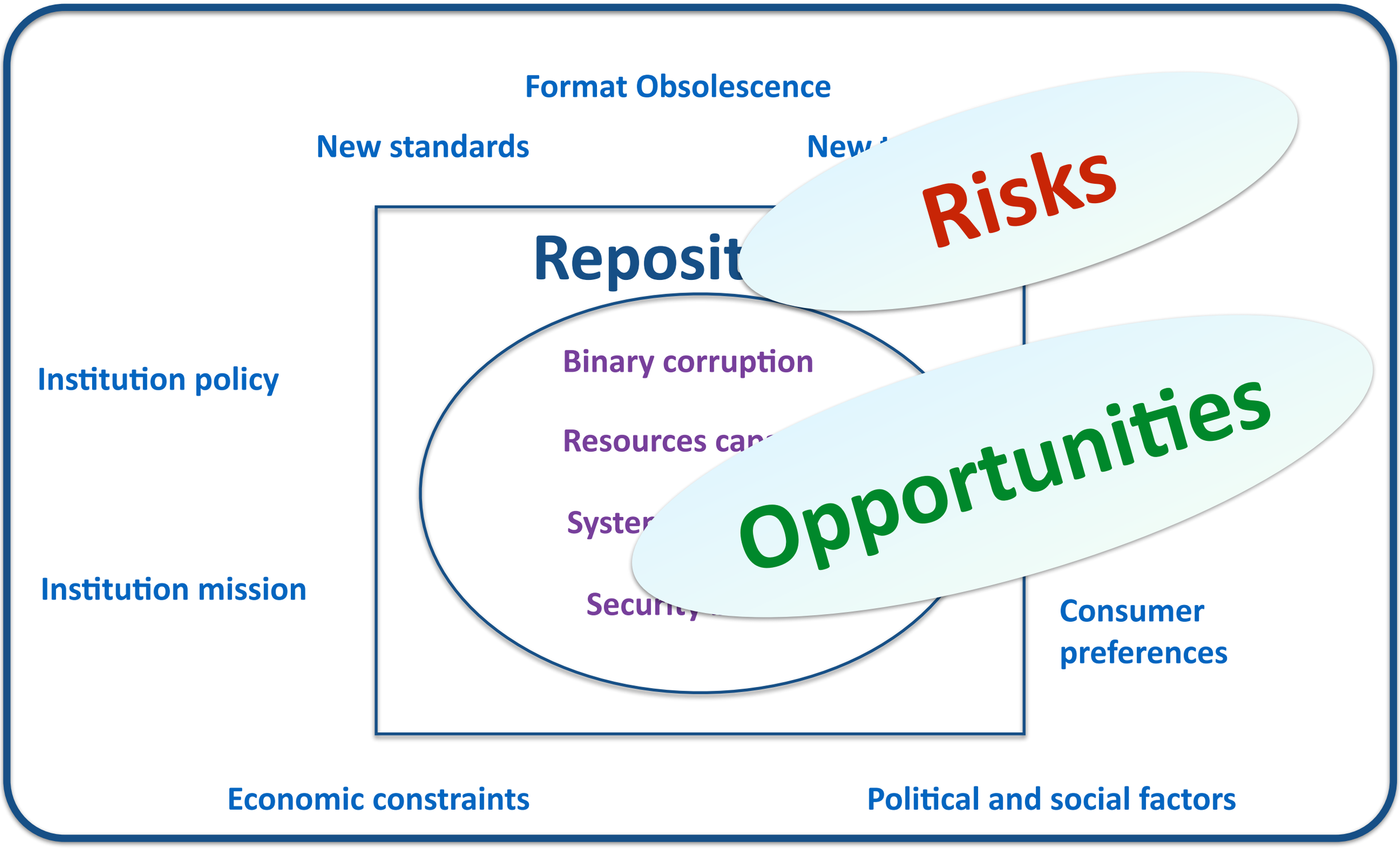
How can we monitor?



http://openplanets.github.io/scout/

Why should we monitor?

Why do we need to monitor?



How should we manage digital information?

KEEPSOLUTIONS
University of Minho SPIN-OFF

Digital Repository. information system able to store, preserve, organize and disseminate digital objects.

KEEPSOLUTIONS
University of Minho SPIN-OFF

OAIS - Open Archival Information Systems

ISO 14721:2012

It defines the functional components that should be part of an archival system aimed at digital preservation

It defines internal and external interfaces

It characterizes digital objects being manipulated

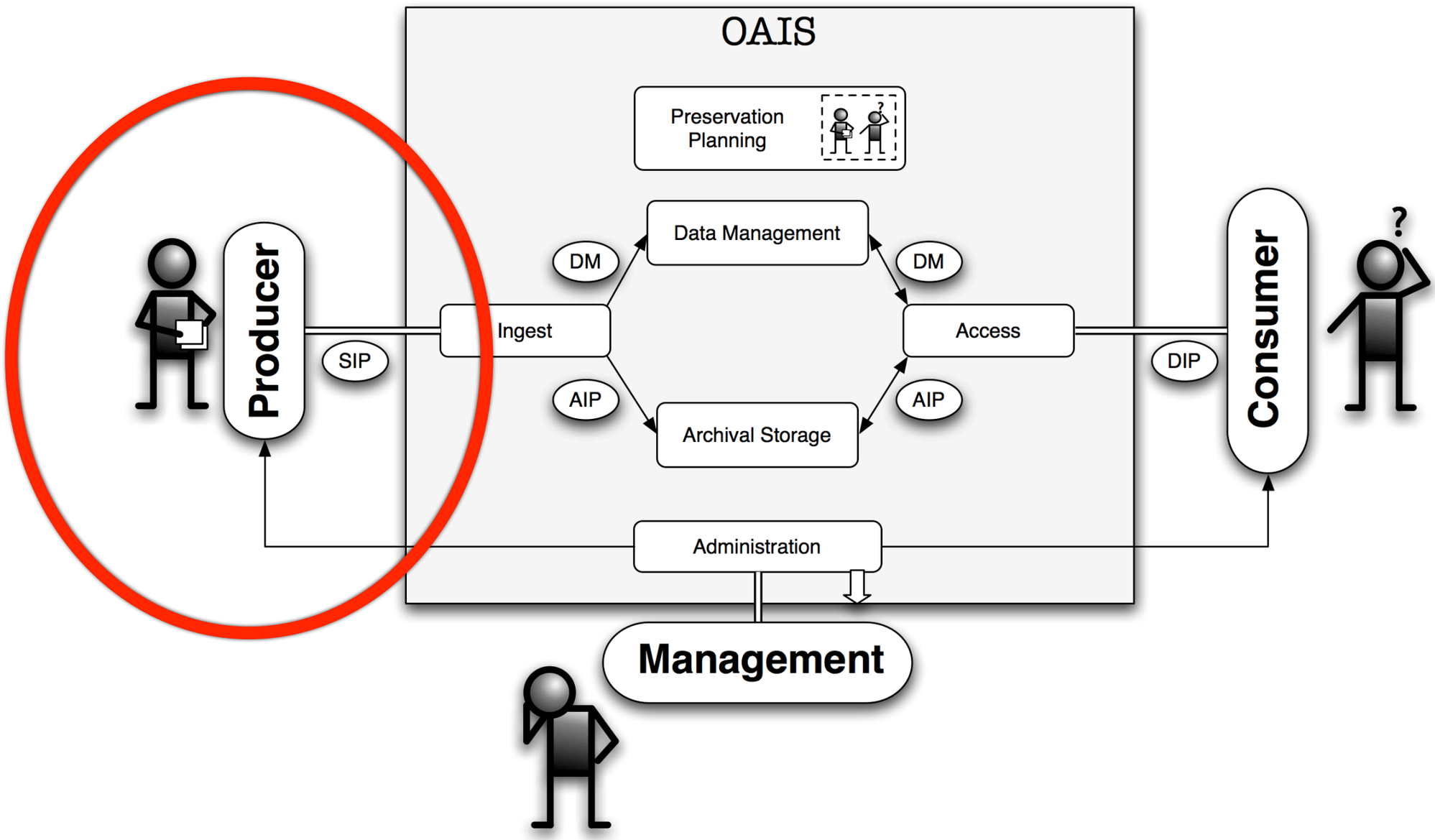
It determines the terminology to be used in a preservation context

Skip

Who are the stakeholders?

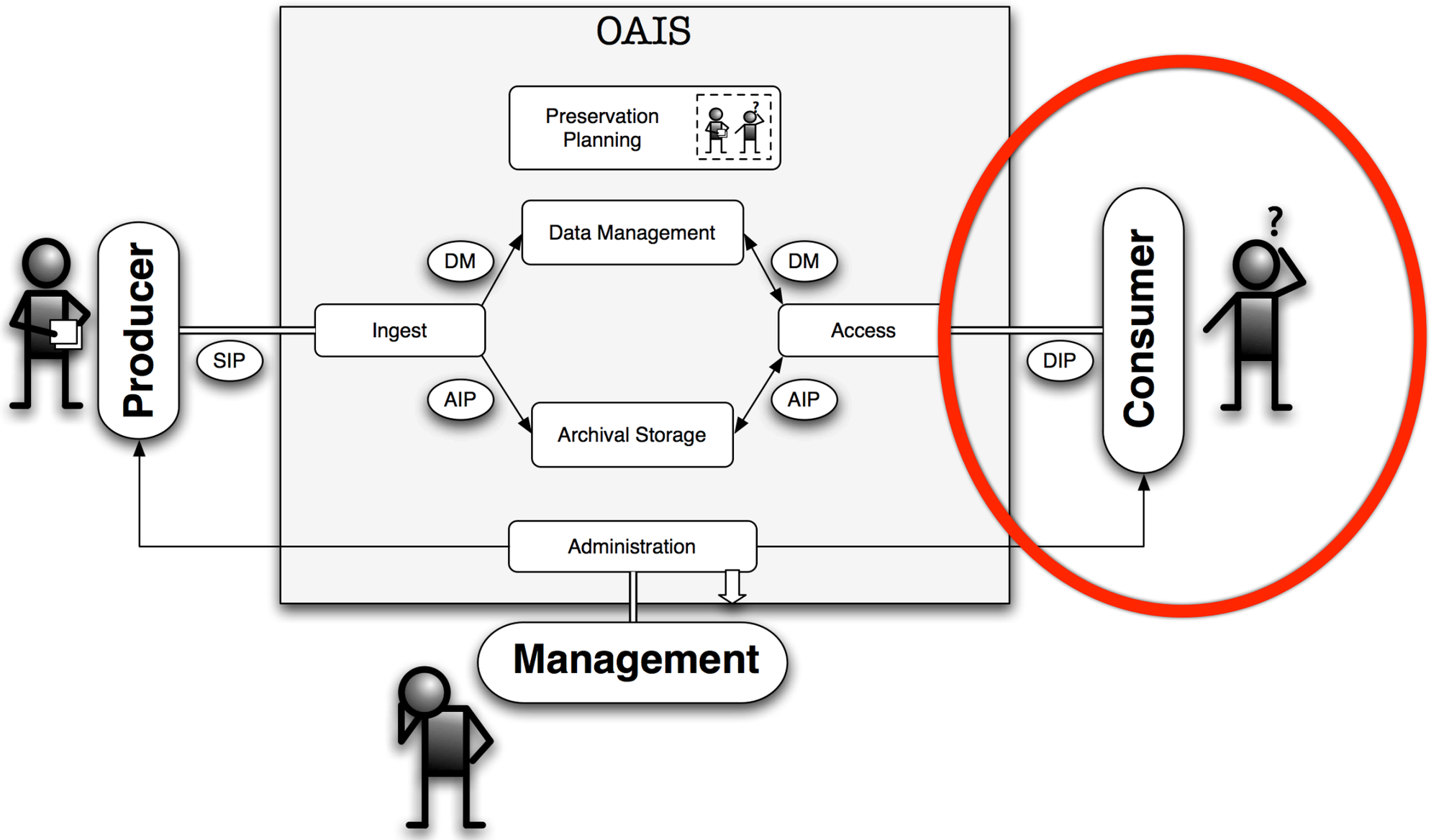
Producer

Responsible for material creation and deposit



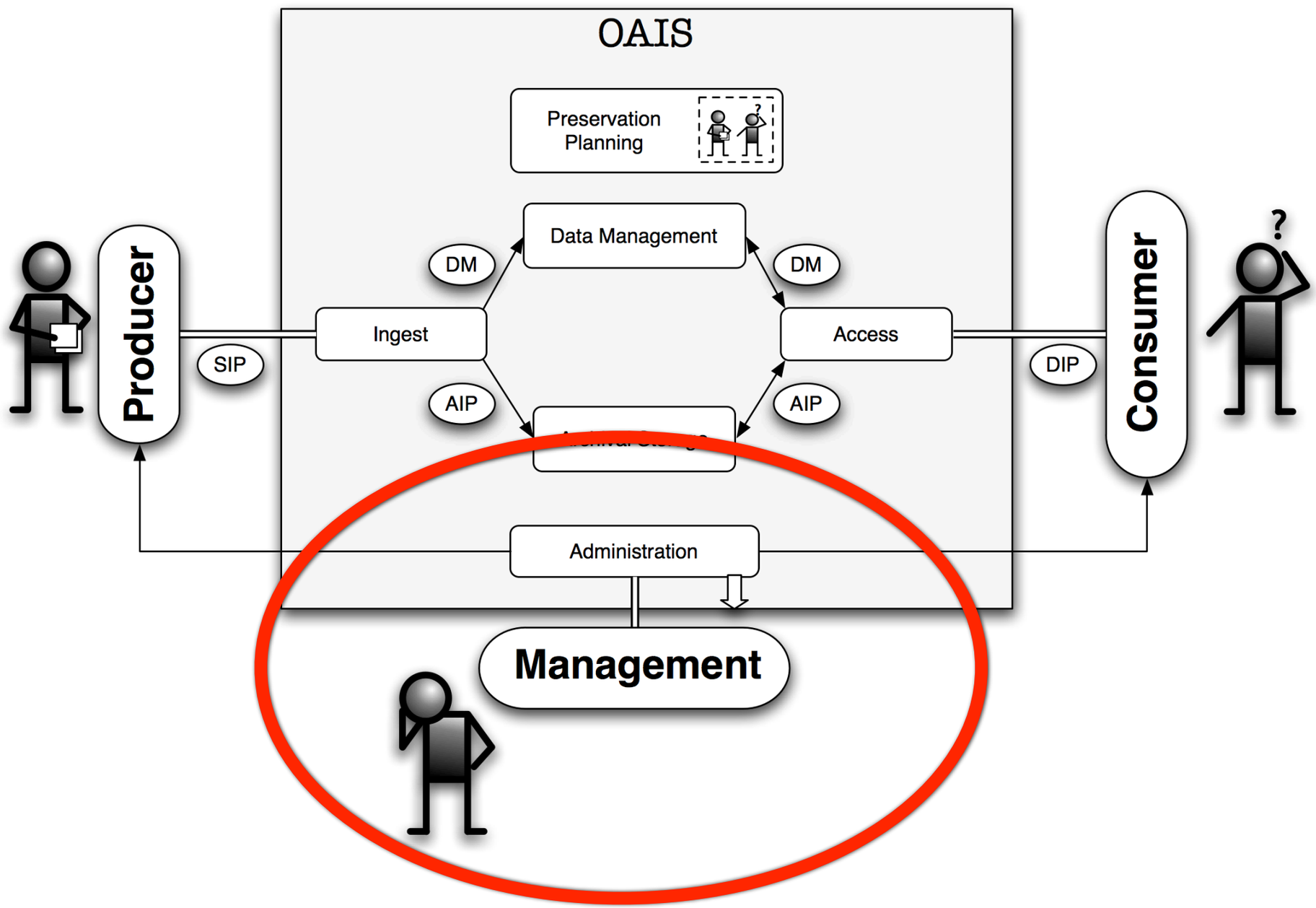
Consumer

The one who searches and accesses information



Administrator

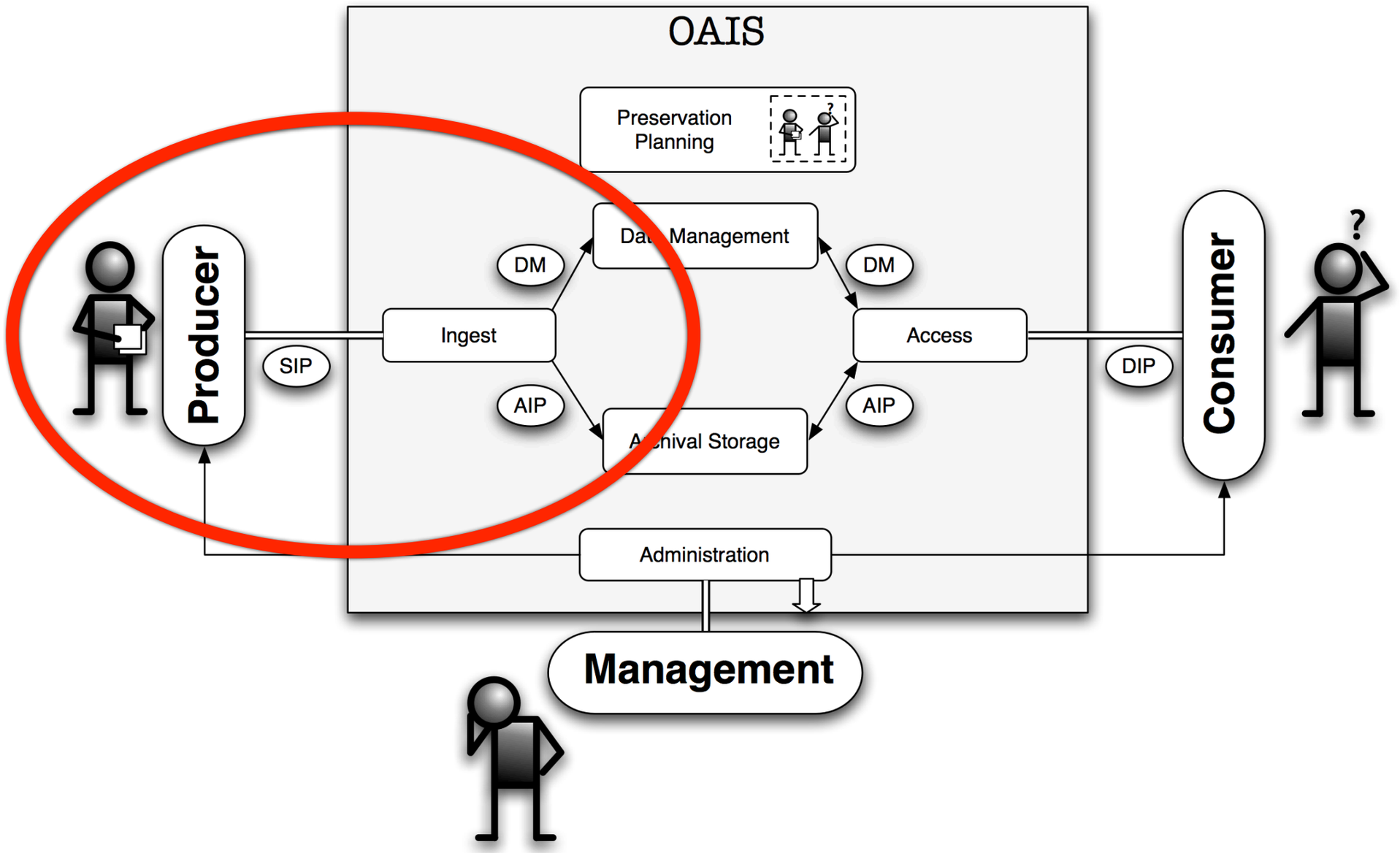
Takes care of system operation



What are the main processes?

Ingestion

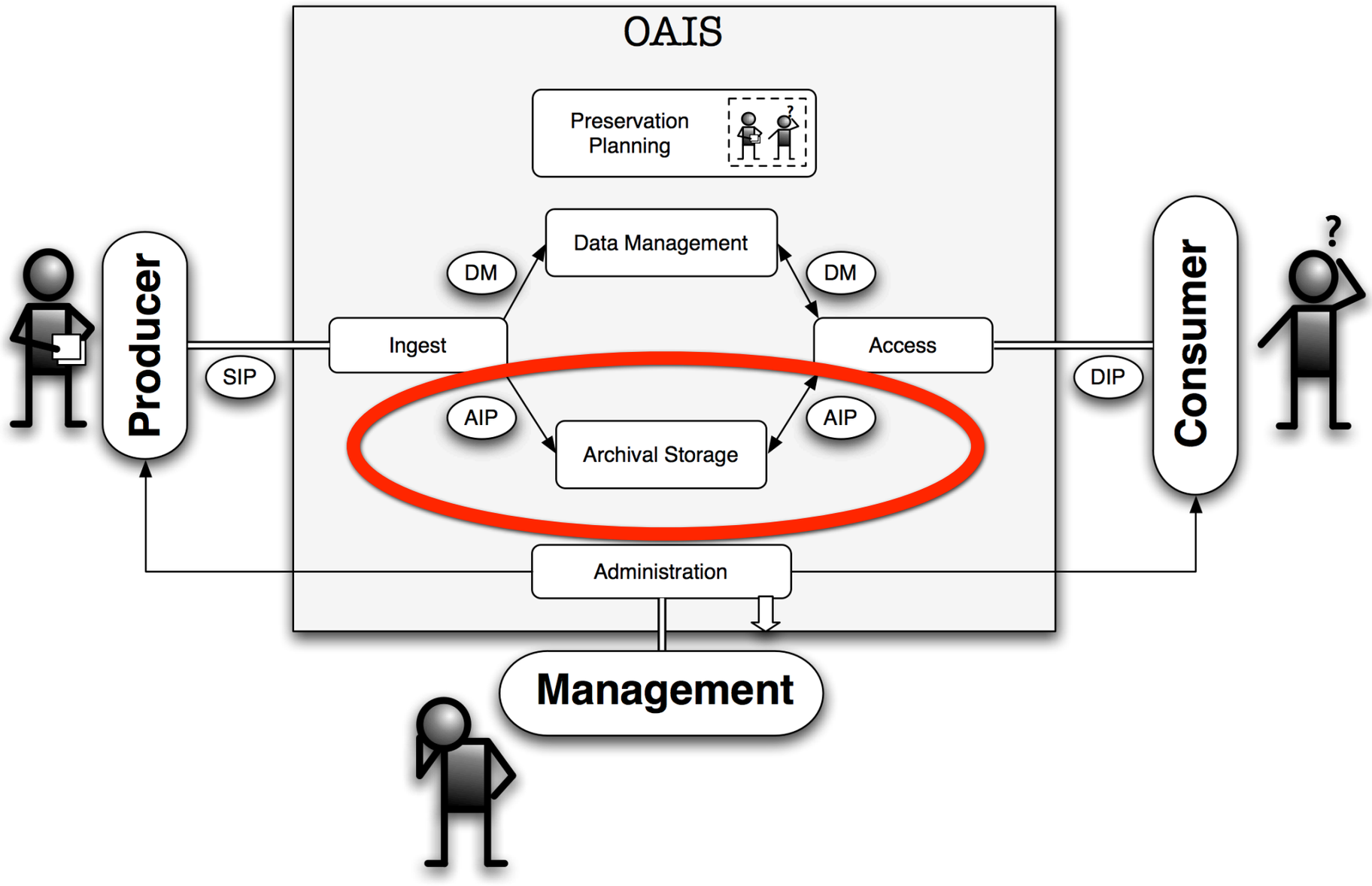
Reception, validation , transformation, description of the information submitted by producers



Archival/Storage

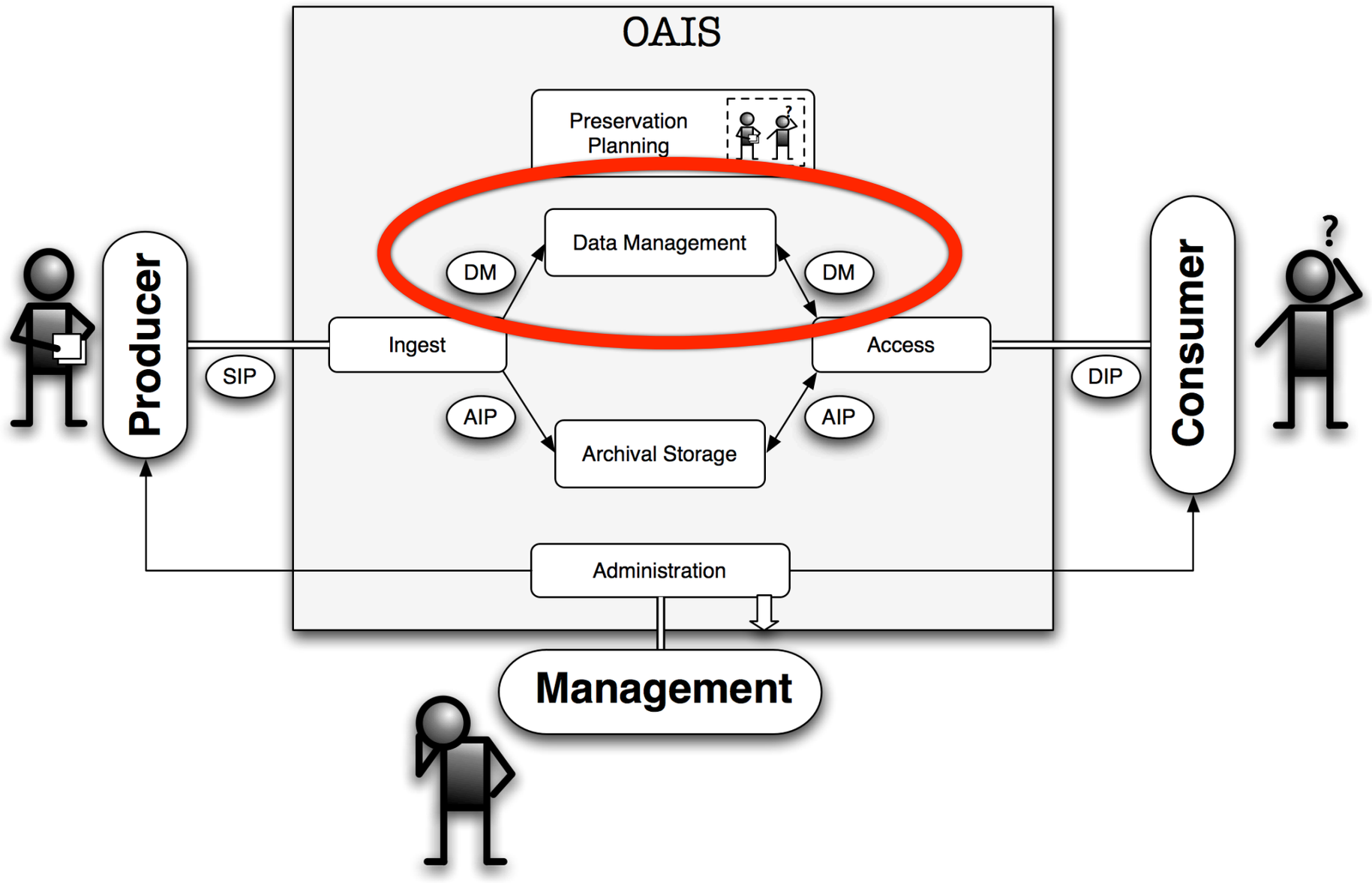
It ensures information preservation at physical and logical levels

- refreshment, format migration,
- integrity check,
- backups, disaster recovery



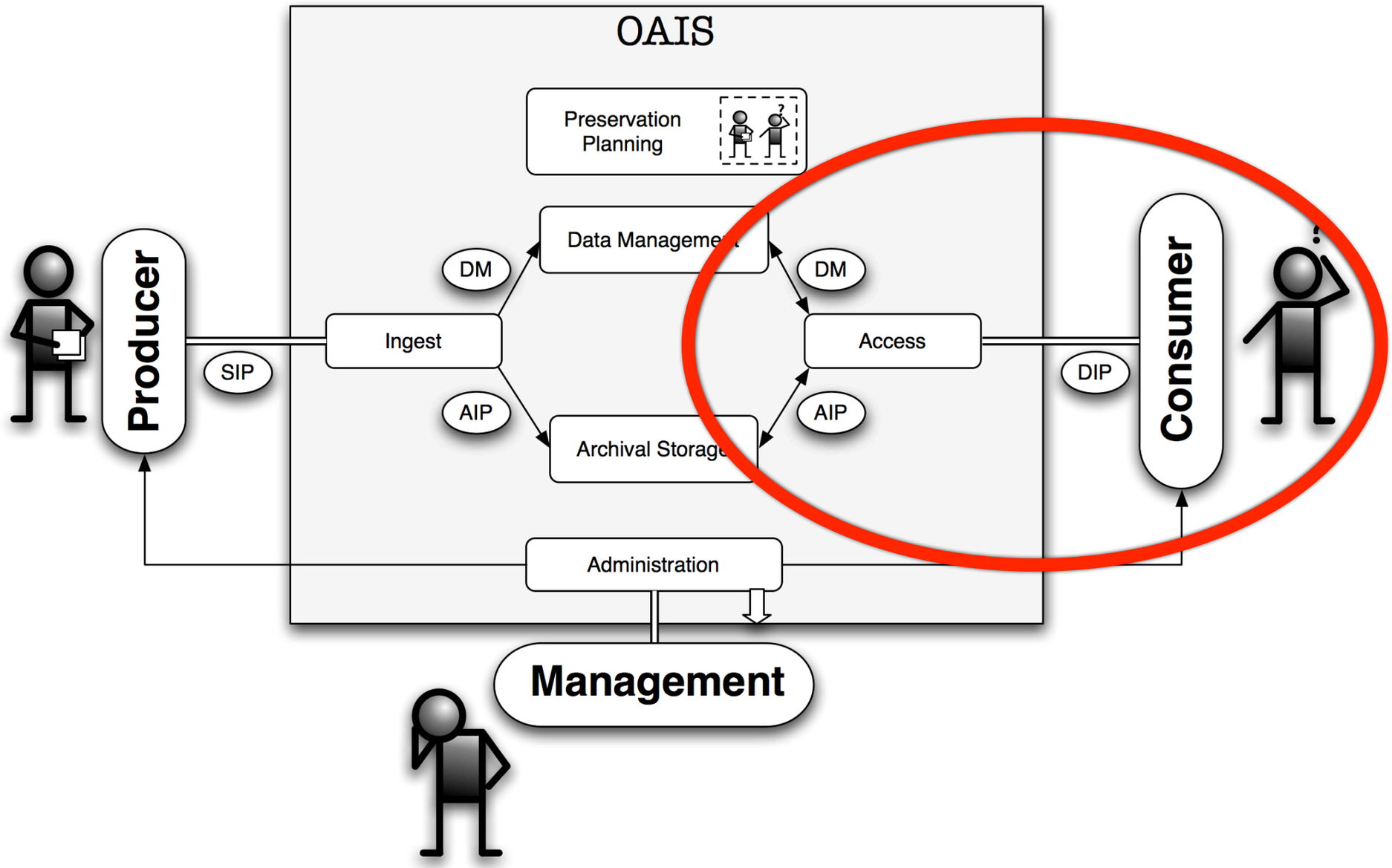
Metadata management

Responsible for digital object metadata management



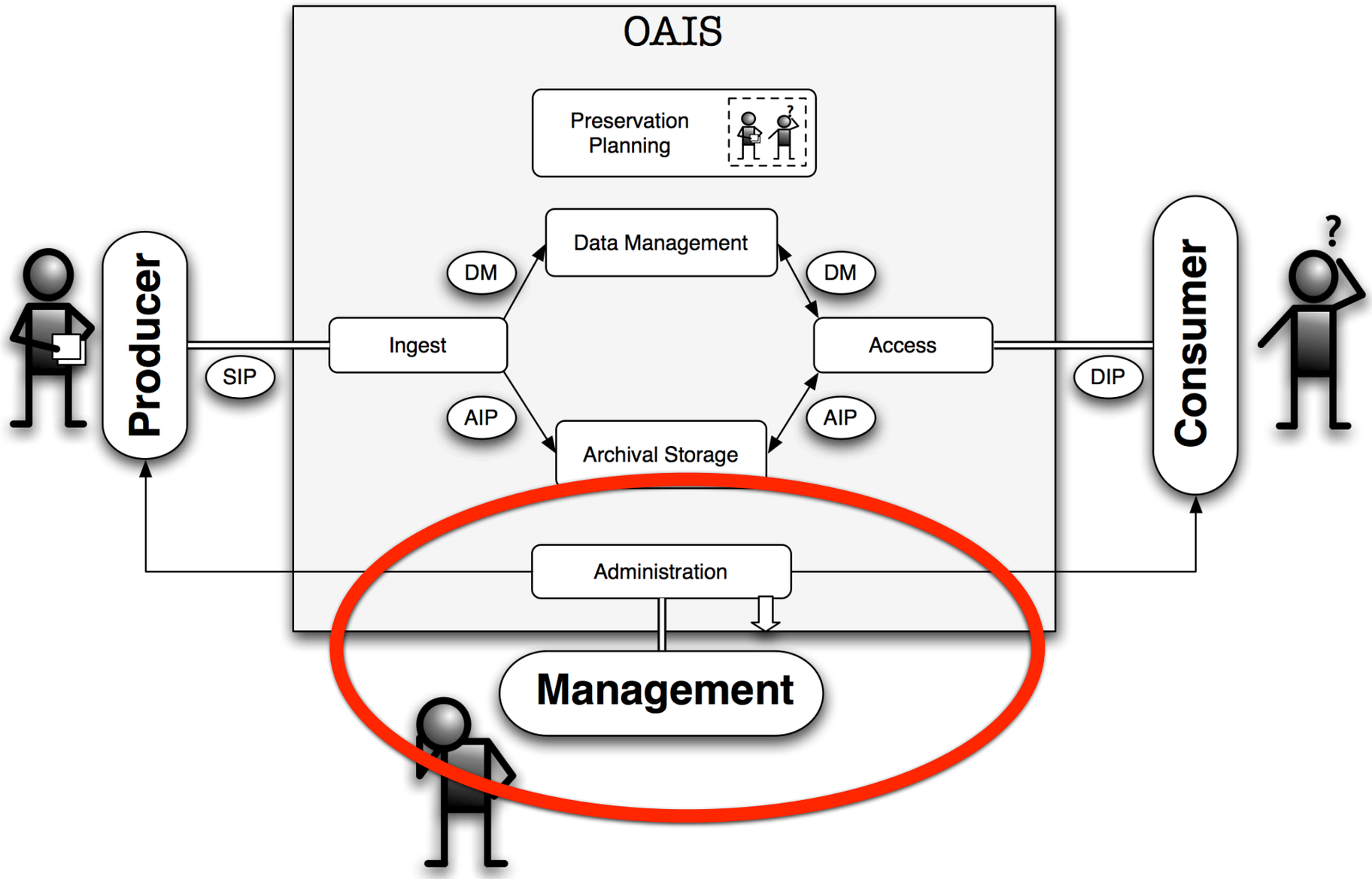
Access

Search, request, transformation and delivery
Access control



Administration

It coordinates everything

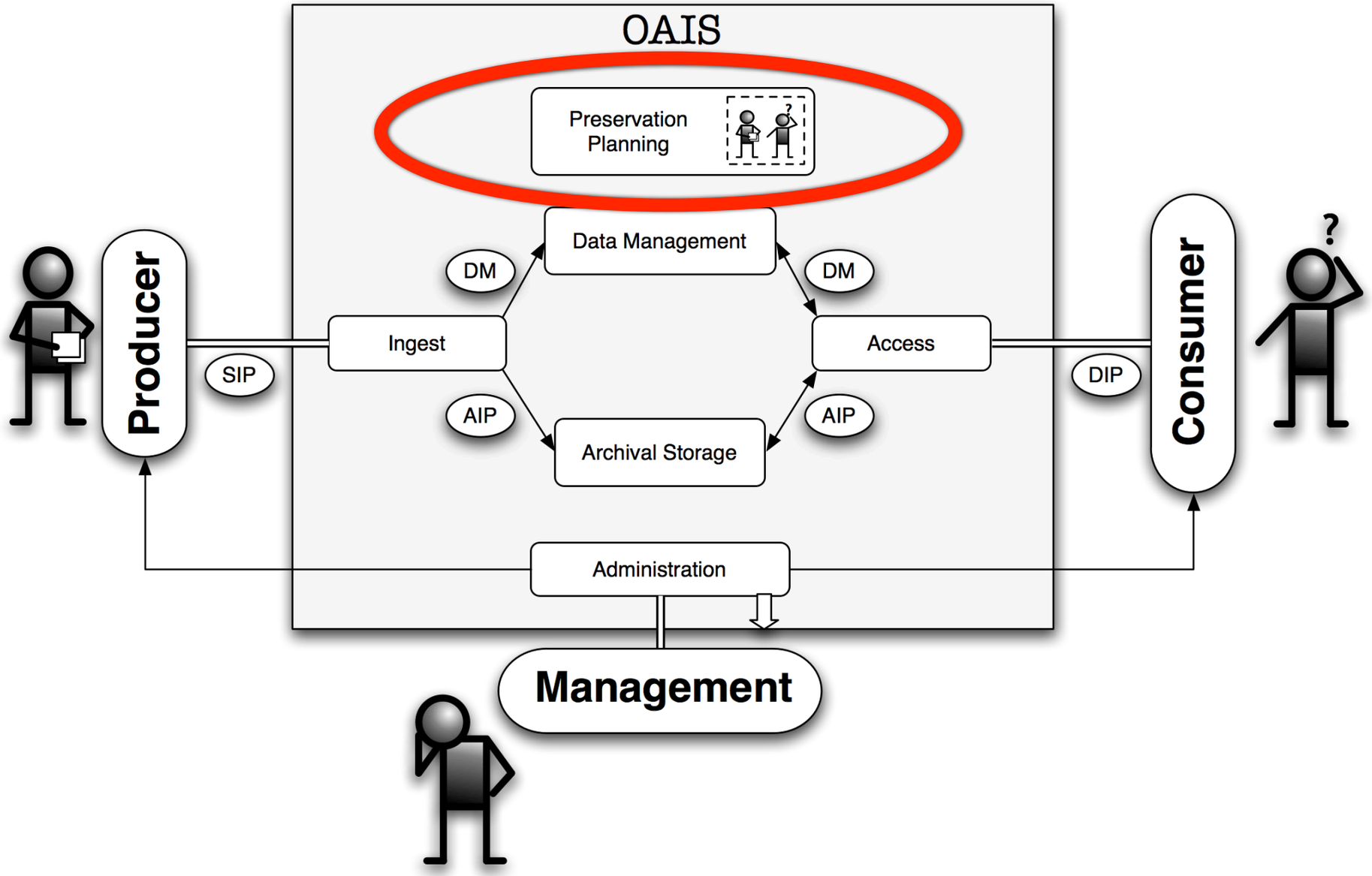


Preservation planning

Preservation policy definition

Environment monitoring (internal and external)

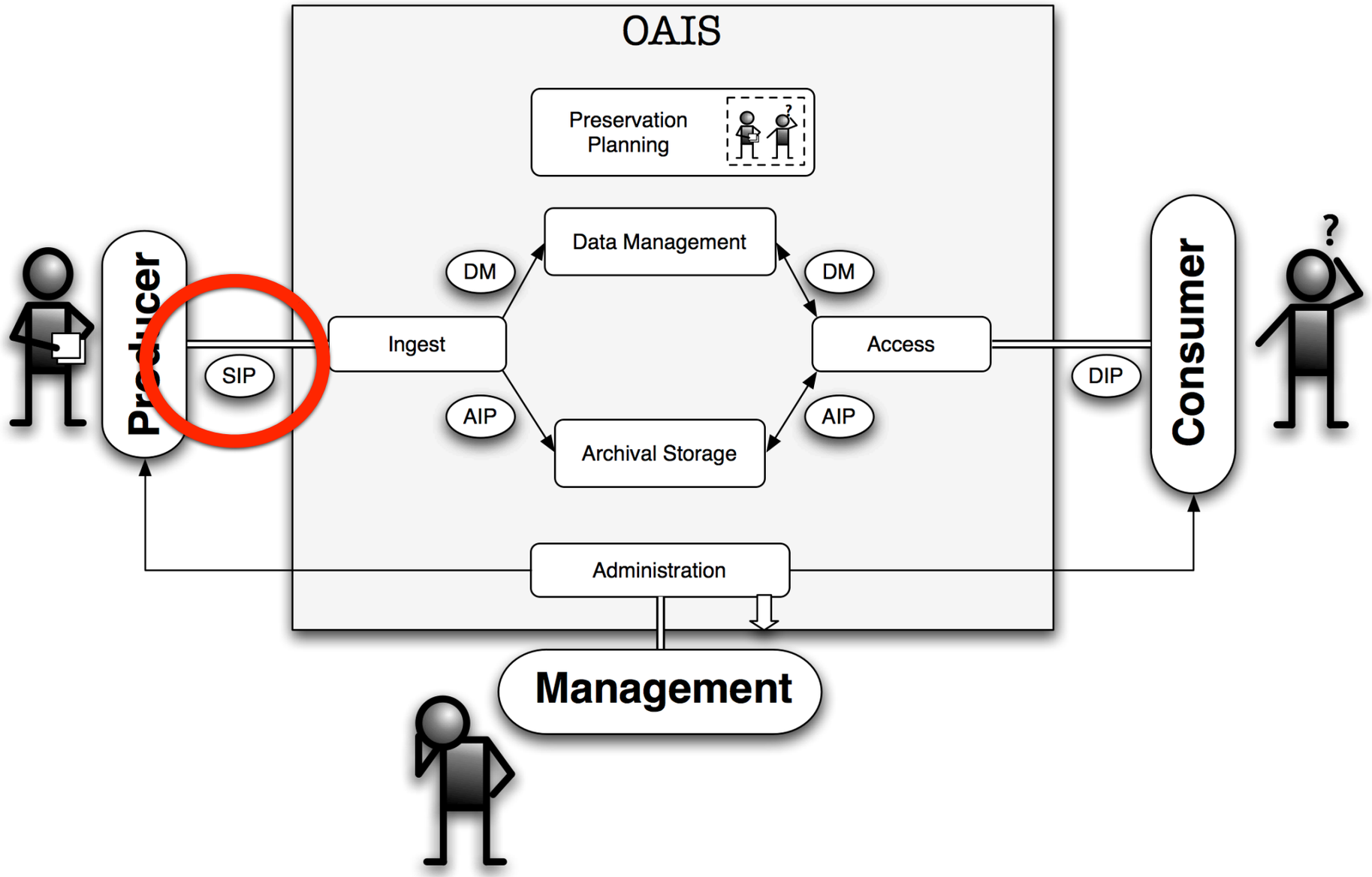
- technology
- community
- market
- etc



What information packages are exchanged?

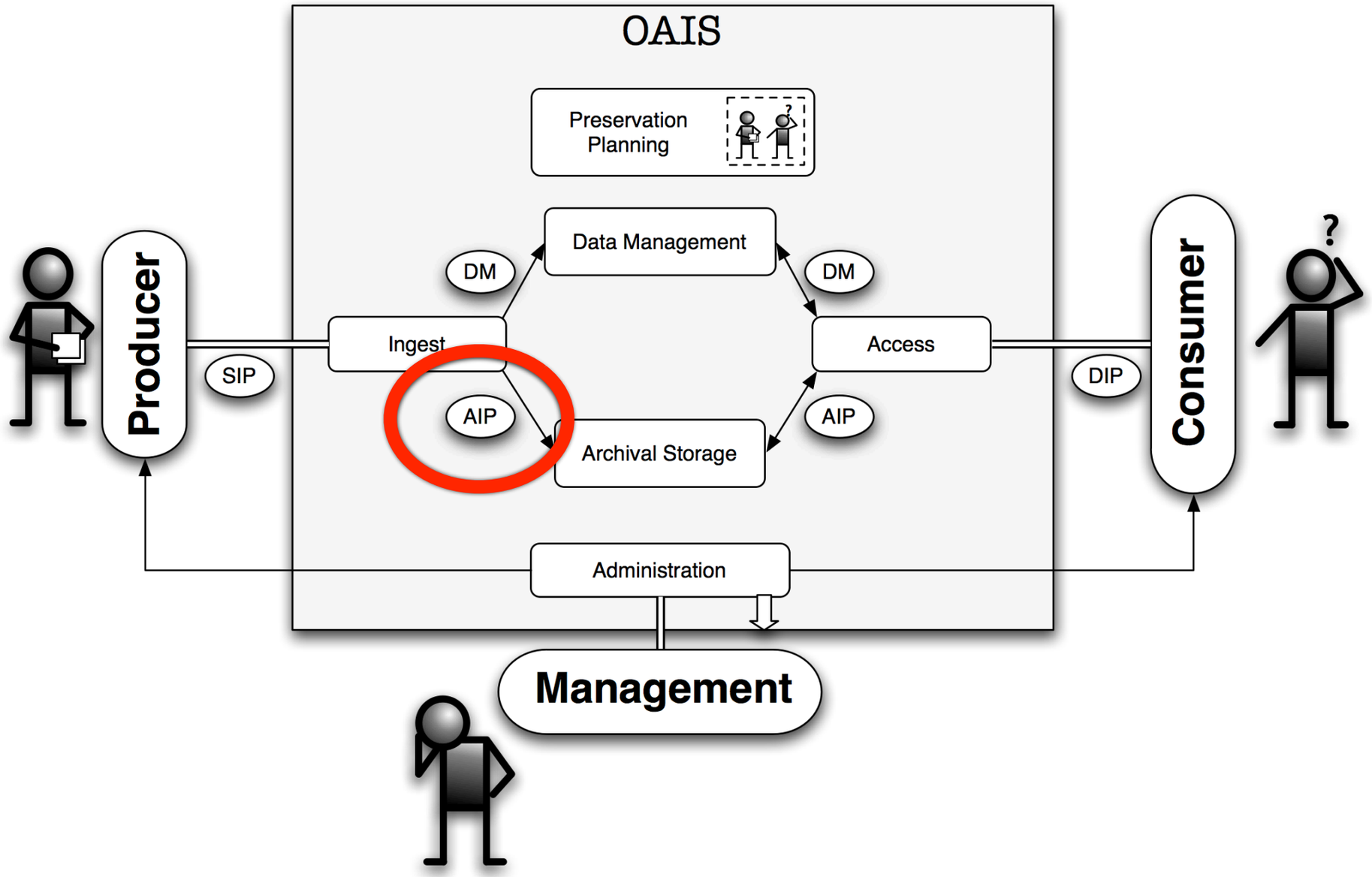
Submission Information Package(SIP)

Digital representation (i.e. content to be preserved)
Producer metadata



Archival Information Package (AIP)

Digital representation
Preservation metadata



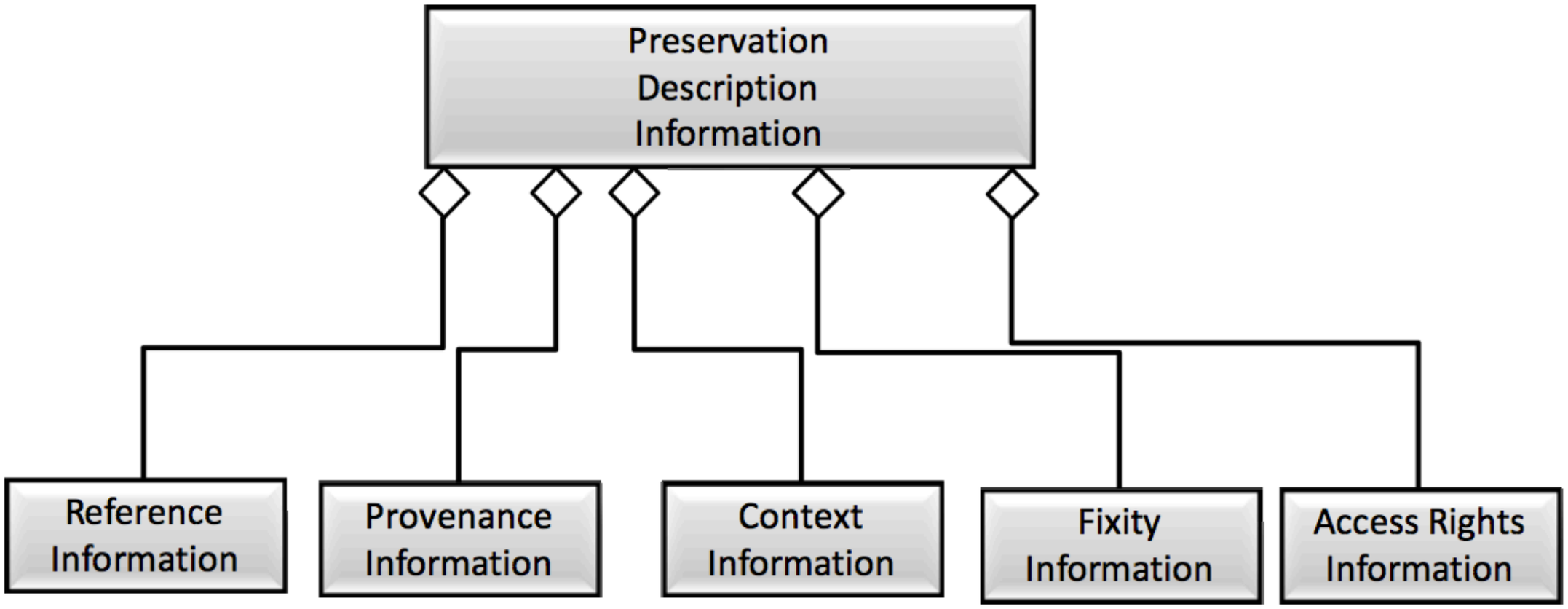
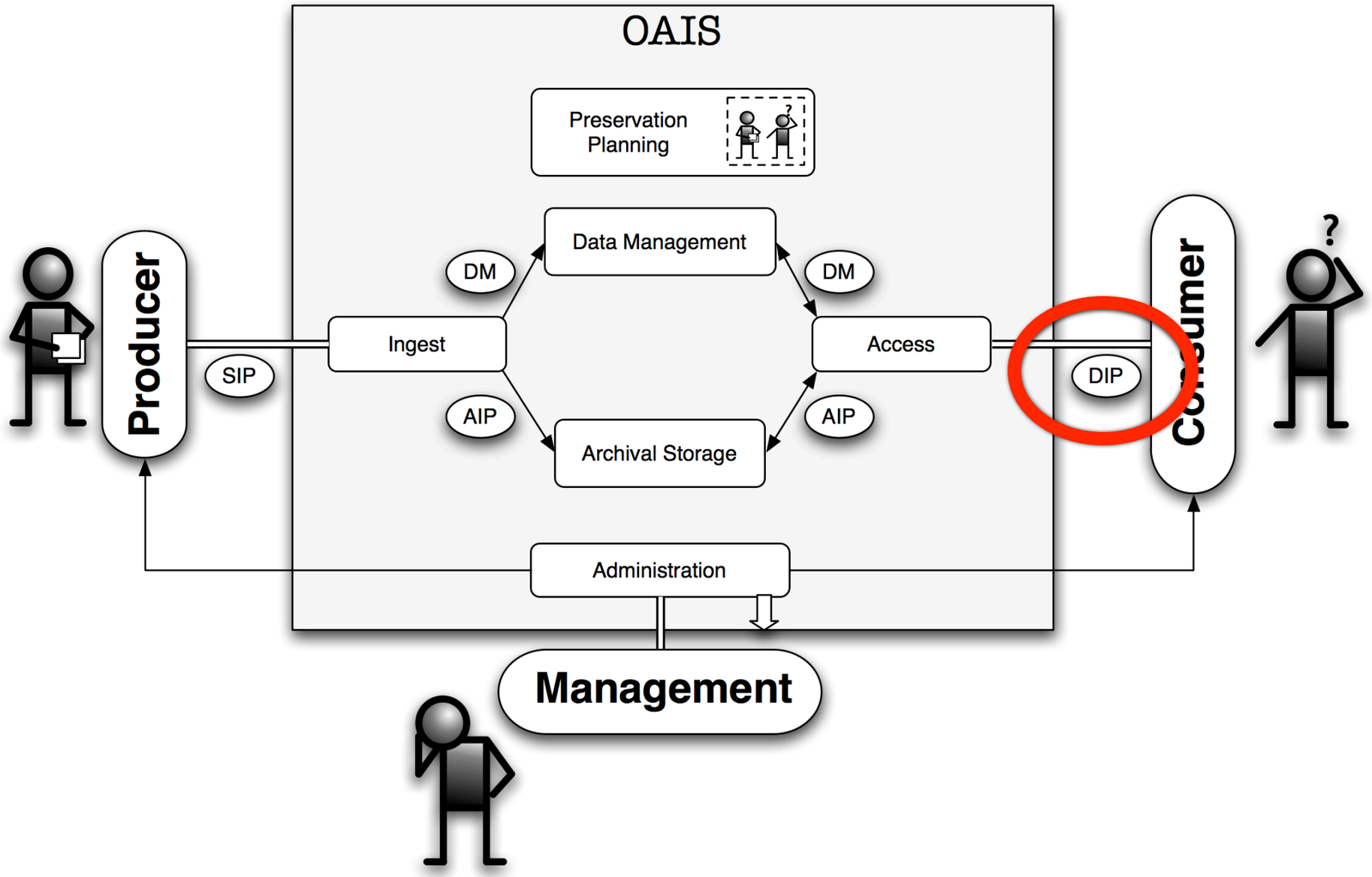


Figure 4-16: Preservation Description Information

Dissemination Information Package (DIP)

Object to be delivered to consumer
Relevant metadata



How detailed is OAIS (ISO 14721:2012)?

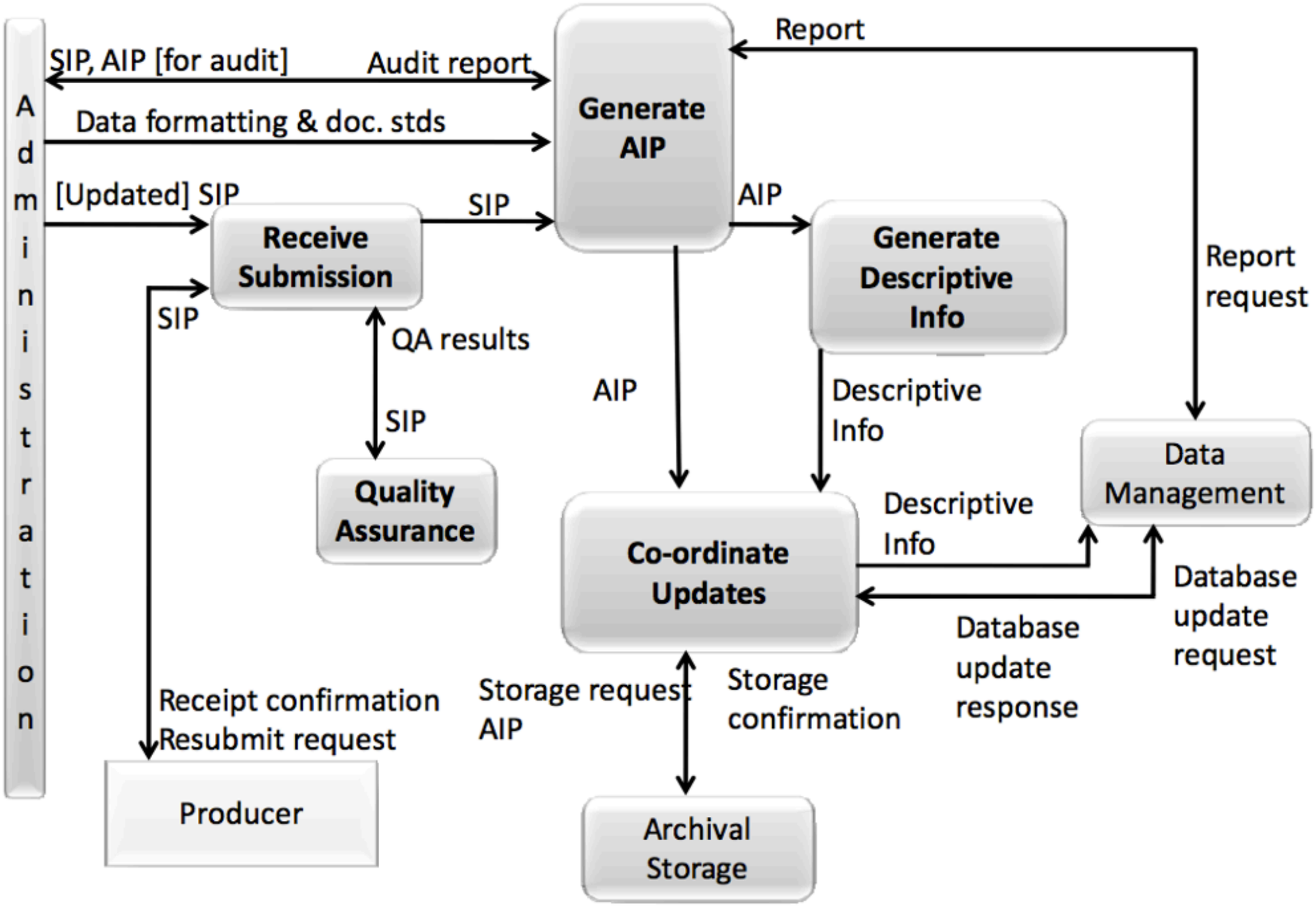


Figure 4-2: Functions of the Ingest Functional Entity

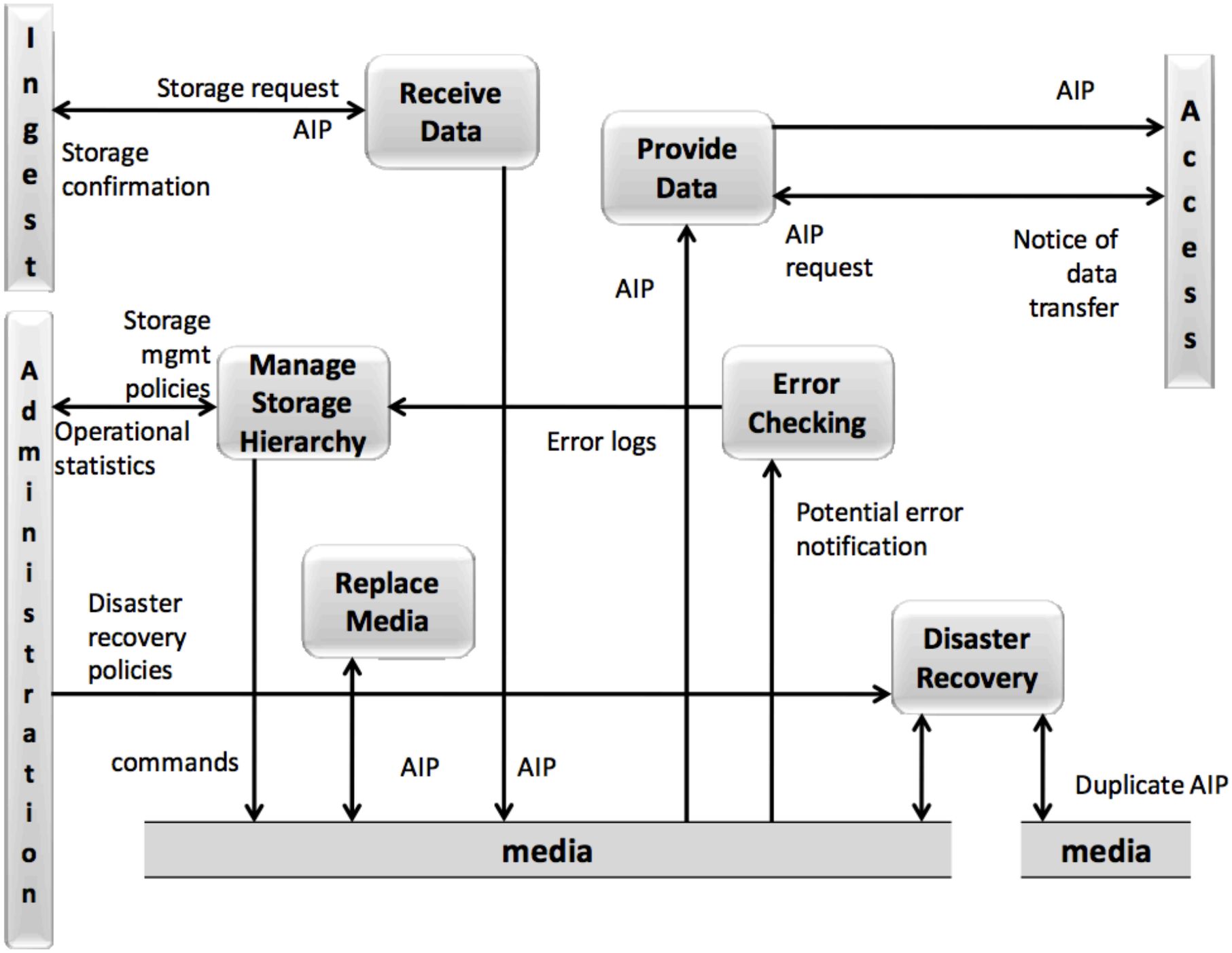


Figure 4-3: Functions of the Archival Storage Functional Entity

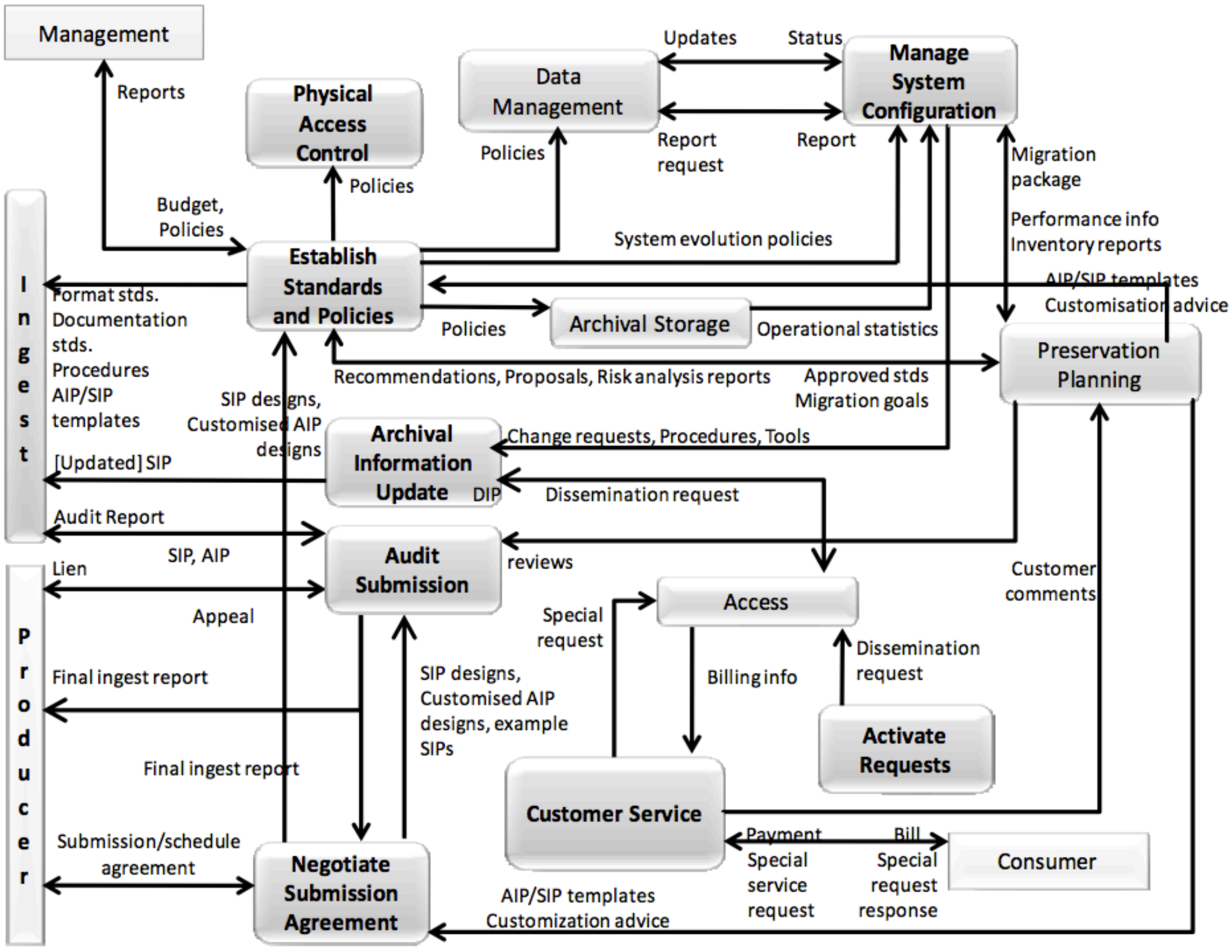


Figure 4-5: Functions of the Administration Functional Entity

It seems complex ... is there anything that can help us ?

KEEPSOLUTIONS
University of Minho SPIN-OFF

RODA

Repositório de Objetos Digitais Autênticos
(Authentic Digital Object Repository)

KEEPSOLUTIONS
University of Minho SPIN-OFF

Digital repository designed for Archives with the following features:

Open-source

Preservation actions and authenticity

Based on standards
ISADg, EAD, PREMIS, OAIS, METS, …

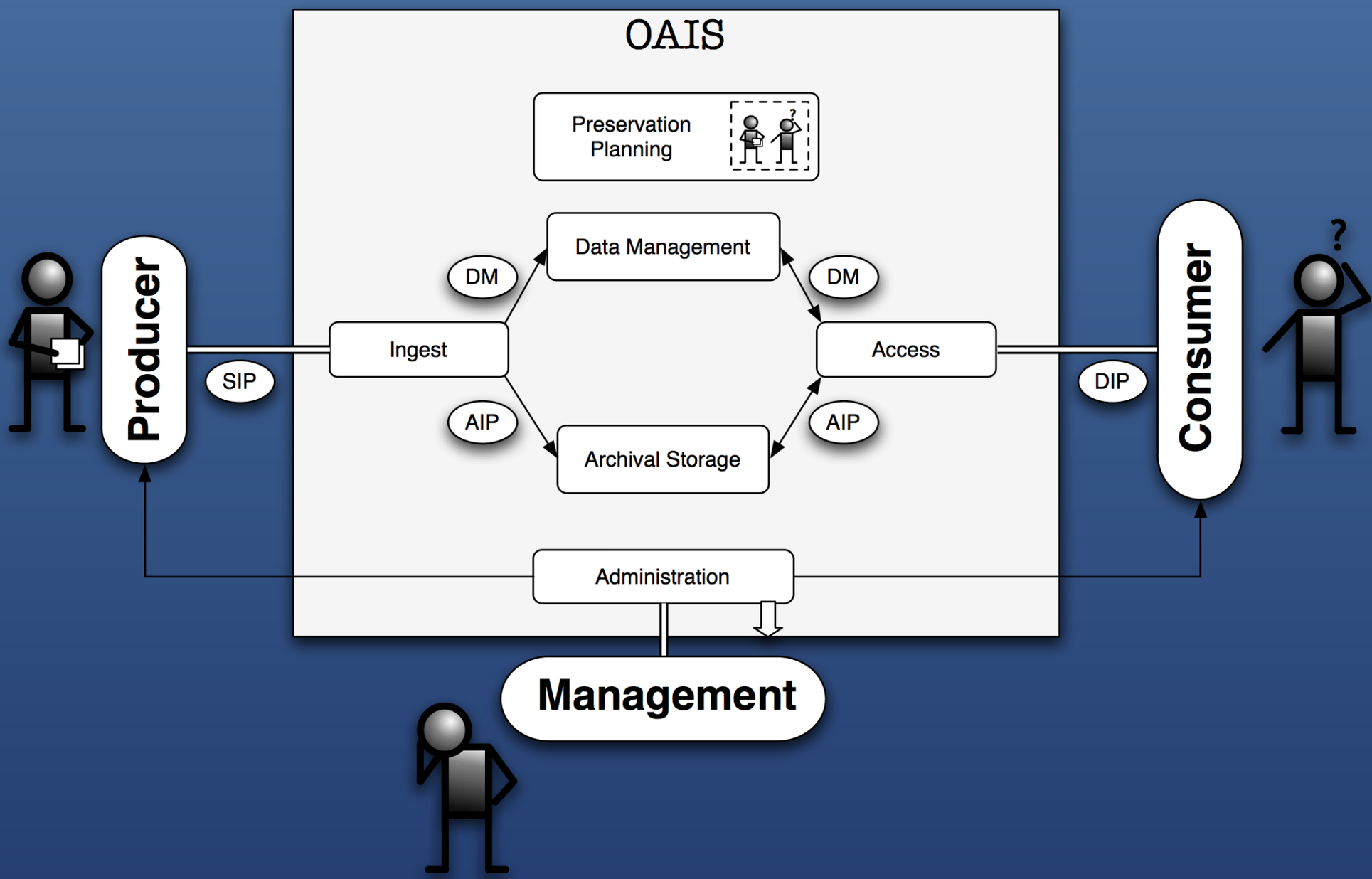
Safe
fine-grained permissions , user authentication, LDAP

Scalable service oriented architecture (SOA)

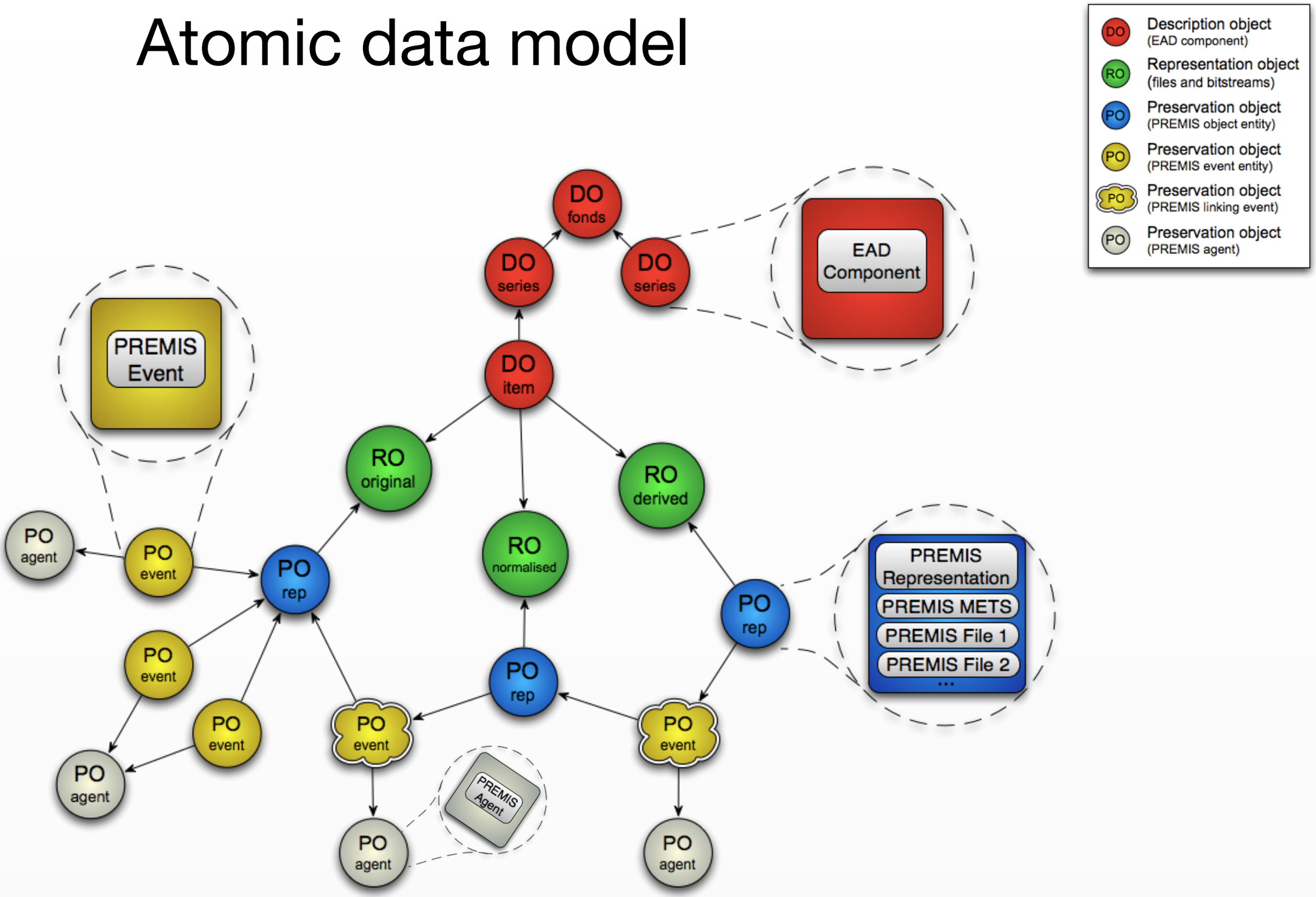
Simple user interface

Supplier and technology independent

OAIS compliant



Atomic data model



Supported object types

Text documents (PDF)

Relational databases (DBML/SIARD)

Video (MPEG-2)

Spreadsheets (CSV, PDF)

Audio (Wav)

Slides (PDF)

Scanned works (Tiff unc.)

Email

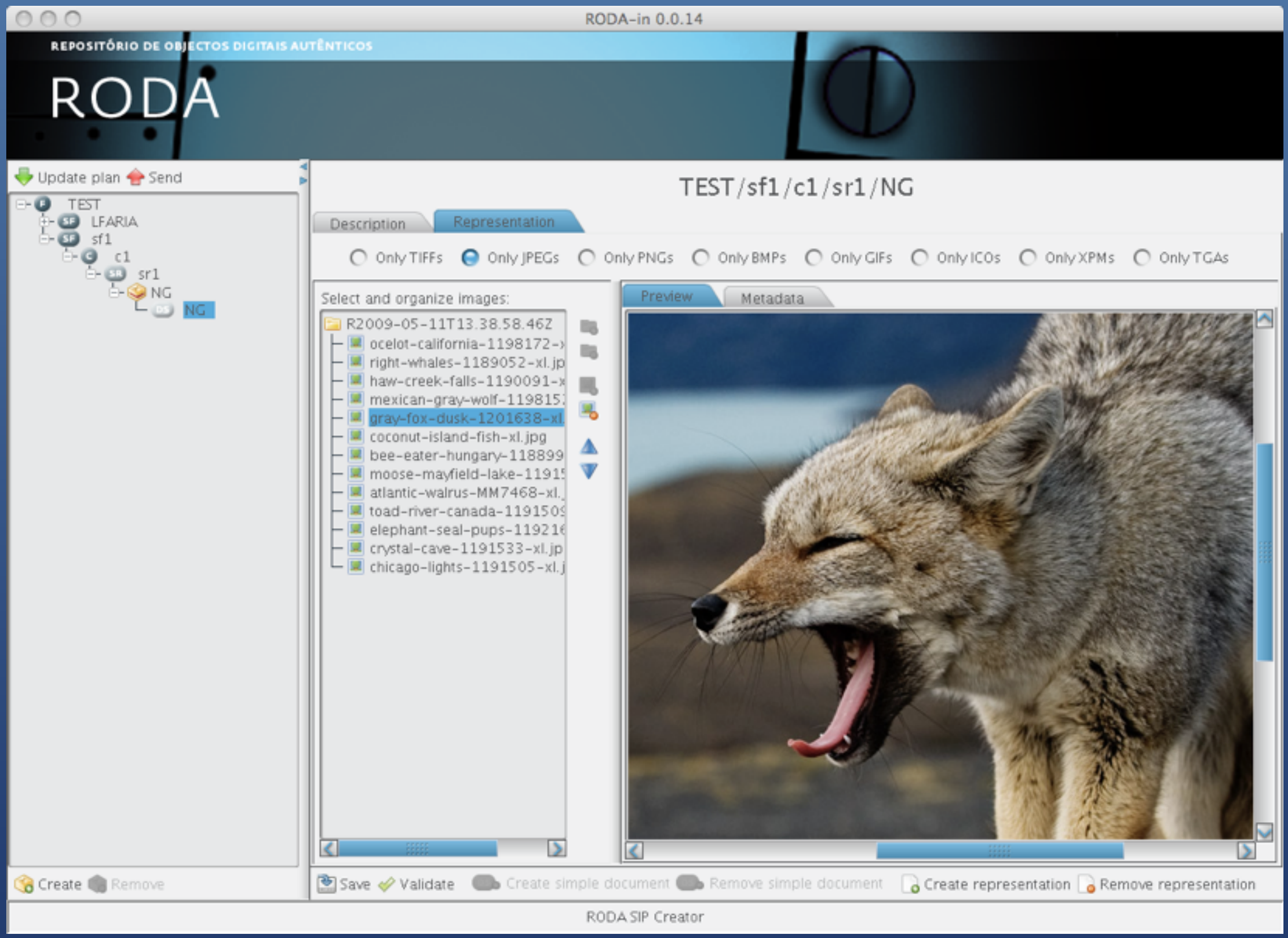
we can ingest other types of objects

But they will only be preserved at logic level

Ingestion workflow

#	Task	Description
1	Upload	Upload SIP to the ingest stage area by HTTP, FTP or local copy
2	Unpack	Open SIP package
3	Virus Check	Check all files for viruses with Clam Anti Virus
4	Syntax and fixity check	Check if metadata is well formed and integrity of all files against checksums in METS envelope
5	Authorization check	Check if user has authorization to ingest into the defined classification plan
6	File characterisation	File format identification (MIME and PRONOM), feature extraction and validation.
7	Ingest	Ingest original into Fedora. This ingestion is not final, can be undone in next steps.
8	Normalisation	Check if representation needs normalization, and enforce it as needed
9	Semantic check	Manual semantic check by an archivist, then object is marked active, indexed and published.

SIP desktop builder



RODA – STATUS

https://roda.dgarq.gov.pt/roda-wui-i18n/?locale=en#ingest.list

Most Visited · Getting Started · Latest Headlines

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS

RODA

Welcome lfaria

Logout

Preferences

ingest / status

STATUS

38 packets

File	Submission date	Current state	%	Producer
TEST_sf1_c1_sr1_NG.sip	2009-05-11 14:41:47	accepted	100%	lfaria
TEST_sf1_c1_sr1_fox.sip	2009-05-07 12:19:36	accepted	100%	lfaria
TEST_LFARIA_MP48298.sip	2009-03-20 17:50:39	accepted	100%	lfaria
TEST_LFARIA_PNG2.sip	2009-03-20 15:06:14	accepted	100%	lfaria
TEST_LFARIA_PNG492.sip	2009-03-20 15:05:28	accepted	100%	lfaria
TEST_LFARIA_XPM.sip	2009-03-20 11:40:25	accepted	100%	lfaria
TEST_LFARIA_WAV.sip	2009-03-20 11:40:24	accepted	100%	lfaria
TEST_LFARIA_TXT.sip	2009-03-20 11:40:16	accepted	100%	lfaria
TEST_LFARIA_TIFF.sip	2009-03-20 11:40:15	accepted	100%	lfaria
TEST_LFARIA_TGA.sip	2009-03-20 11:40:15	accepted	100%	lfaria
TEST_LFARIA_RTF.sip	2009-03-20 11:40:14	accepted	100%	lfaria
TEST_LFARIA_PDF.sip	2009-03-20 11:40:13	accepted	100%	lfaria
TEST_LFARIA_OGG.sip	2009-03-20 11:40:06	accepted	100%	lfaria
TEST_LFARIA_ODT.sip	2009-03-20 11:40:05	accepted	100%	lfaria
TEST_LFARIA_MP3.sip	2009-03-20 11:40:01	accepted	100%	lfaria
TEST_LFARIA_JPEG.sip	2009-03-20 11:39:28	accepted	100%	lfaria
TEST_LFARIA_ICO.sip	2009-03-20 11:39:25	accepted	100%	lfaria
TEST_LFARIA_GIF.sip	2009-03-20 11:39:24	accepted	100%	lfaria

Filter

processed
processing
accepted
rejected
All

Producer

lfaria

REPORT
VIEW
ACCEPT
REJECT

Done

roda.dgarq.gov.pt

RODA – STATUS

https://roda.dgarq.gov.pt/roda-wui-i18n/?locale=en#ingest.list

Most Visited · Getting Started · Latest Headlines

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS

RODA

Welcome lfaria

Logout

ABOUT RODA
CATALOG
INGEST
ADMINISTRATION

ingest / status

STATUS

Ingest report

Original file name: TEST_sf1_c1_sr1_NG.sip

- ▶ received (HTTP)
- ▶ unpacked
- ▶ virus-free
- ▶ well formed
- ▶ authorized
- ▶ ingestd
- ▶ normalized
- ▼ accepted

Date and time: 2009-05-11 15:10:44.000
Task: Semantic Check
Result: Success
Details: SIP aceite pelo utilizador lfaria: SIP valid for demoSIP objects marked active [rc

CLOSE

TEST_LFARIA_MP3.sip	2009-03-20 11:40:01	accepted	100%	lfaria
TEST_LFARIA_JPEG.sip	2009-03-20 11:39:28	accepted	100%	lfaria
TEST_LFARIA_ICO.sip	2009-03-20 11:39:25	accepted	100%	lfaria
TEST_LFARIA_GIF.sip	2009-03-20 11:39:24	accepted	100%	lfaria

Done

roda.dgarq.gov.pt

Access

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS

RODA

Welcome lfaria

Logout

Preferences

ABOUT RODA · SERVICES · POLICY AND TECHNICAL DOCUMENTS · R&D · CONTACTS · ABOUT THE REGISTER · HELP

CATALOG · BROWSE · SEARCH · ADVANCED SEARCH · HELP

INGEST · PRE-INGEST · SUBMIT · STATUS · HELP

ADMINISTRATION · USERS · SCHEDULER · STATISTICS · LOG · HELP

catalog / search / advanced search

ADVANCED SEARCH

Select the level of description you want to search

- ○ all levels
- ◉ choose levels

☐ F fonds ☐ C class ☐ SR series ☐ DC file

☐ SF subfonds ☐ SC subclass ☐ SSR subseries ☑ IT item

Select the time limits in which to search

- ○ without time restrictions
- ◉ specify interval

from 2008 January 01 to 2009 January 01

Fill out the words you want to find

- ○ all the fields
- ◉ choose fields

tests

◉ all words ○ exact phrase ○ at least one of the words ○ without the words

☐ Complete reference ☐ Quote ☐ Physical characteristics and technical requirements

☐ Title ☐ Administrative and biographical history ☐ Access restritions

☐ Origination ☐ Custodial history ☐ Reproduction restritions

☐ Acquisition number ☐ Acquisition information

☐ Material specification ☐ Other find aids

☐ Physical description ☑ Scope and content ☐ Related Materials

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS

RODA

Welcome lfaria

Logout

Preferences

ABOUT RODA · SERVICES · POLICY AND TECHNICAL DOCUMENTS · R&D · CONTACTS · ABOUT THE REGISTER · HELP

CATALOG · BROWSE · SEARCH · ADVANCED SEARCH · HELP

INGEST · PRE-INGEST · SUBMIT · STATUS · HELP

ADMINISTRATION · USERS · SCHEDULER · STATISTICS · LOG · HELP

catalog / browse / 50218

BROWSE

4 fonds

Reference

- AACC
- DGARQ
- TEST
  - 692
  - LFARIA
  - sf1
    - c1
      - sr1
        - 101
        - 815
        - 1030
        - 50192
        - 50218
- fbarbedo

Description | Visualization | Preservation | Permissions

Identification

Reference	50218
Complete reference	PT/DGARQ/TEST/sf1/c1/sr1/50218
Handle	http://hdl.handle.net/10384/50218
Title	National Geographic Photos
Level	item
Initial date	2008-01-01
Final date	2009-01-01
Country code	PT
Repository code	DGARQ
Origination	National Geographic
Acquisition number	NG
Acquisition date	2009-05-11

Content and structure

Scope and content	tests for or09

Notes

Note	Pacote de submissão validado semanticamente por lfaria

EDIT · SUBLEVEL · MOVE · EAD · CLOSE

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS

RODA

Welcome lfaria

Logout

Preferences

ABOUT RODA | SERVICES · POLICY AND TECHNICAL DOCUMENTS · R&D · CONTACTS · ABOUT THE REGISTER · HELP

CATALOG | BROWSE · SEARCH · ADVANCED SEARCH · HELP

INGEST | PRE-INGEST · SUBMIT · STATUS · HELP

ADMINISTRATION | USERS · SCHEDULER · STATISTICS · LOG · HELP

catalog / browse / 50218

BROWSE

4 fonds

Reference

- AACC
- DGARQ
- TEST
  - 692
  - LFARIA
  - sf1
    - c1
      - sr1
        - 101
        - 815
        - 1030
        - 50192
        - 50218
- fbarbedo

Description | Visualization | Preservation | Permissions

Disseminassions of 50218 - "National Geographic Photos"

JPEG
(original)

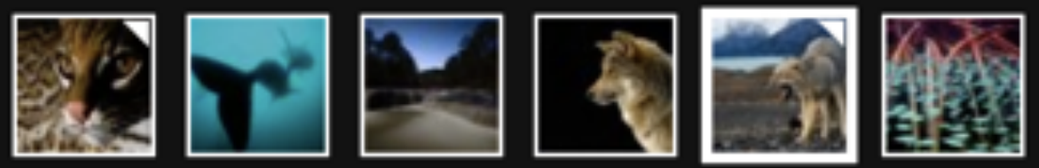
TIFF
(normalized)

Photo pre-viewer

Book pre-viewer

CLOSE

National Geographic Photos



(5/13) gray-fox-dusk-1201638-xl.jpg.tiff

REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS

RODA

Welcome lfaria

Logout

Preferences

ABOUT RODA | SERVICES · POLICY AND TECHNICAL DOCUMENTS · R&D · CONTACTS · ABOUT THE REGISTER · HELP

CATALOG | BROWSE · SEARCH · ADVANCED SEARCH · HELP

INGEST | PRE-INGEST · SUBMIT · STATUS · HELP

ADMINISTRATION | USERS · SCHEDULER · STATISTICS · LOG · HELP

catalog / browse / 50218

BROWSE

4 fonds

Reference

- AACC
- DGARQ
- TEST
  - 692
  - LFARIA
  - sf1
    - c1
      - sr1
        - 101
        - 815
        - 1030
        - 50192
        - 50218
- fbarbedo

Description | Visualization | Preservation | Permis

roda:50219 (original) roda:50222 (normalized)

ingestion

Description: The objects inside the SIP were added to the preservation repository
Result: success
Ingest details:

Day

ingestion

normalization

May 8 | May 9 | May 10 | May 11 | May 12 | May 13

Apr | May | Jun

PREMIS

CLOSE

How does RODA preserve information?

KEEPSOLUTIONS
University of Minho SPIN-OFF

Format normalization during ingestion

Configurable normalization rules

Preservation metadata automatic generation

PREMIS events, representation information, agents

Plato preservation planning compliant

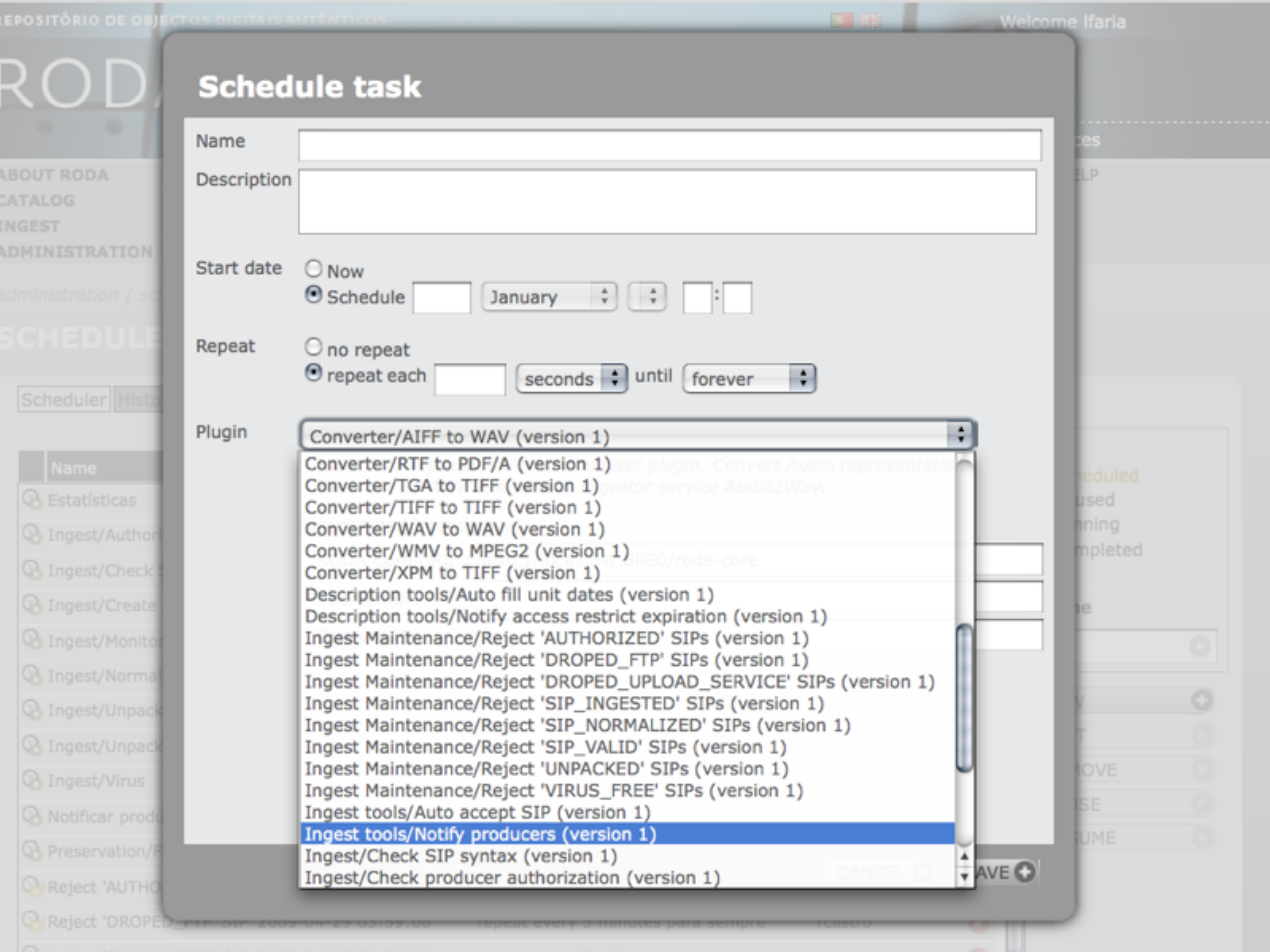
New preservation components can be easily added as plugins

Original representations are kept unchanged

Enabling possible future strategies

Built-in integrity check

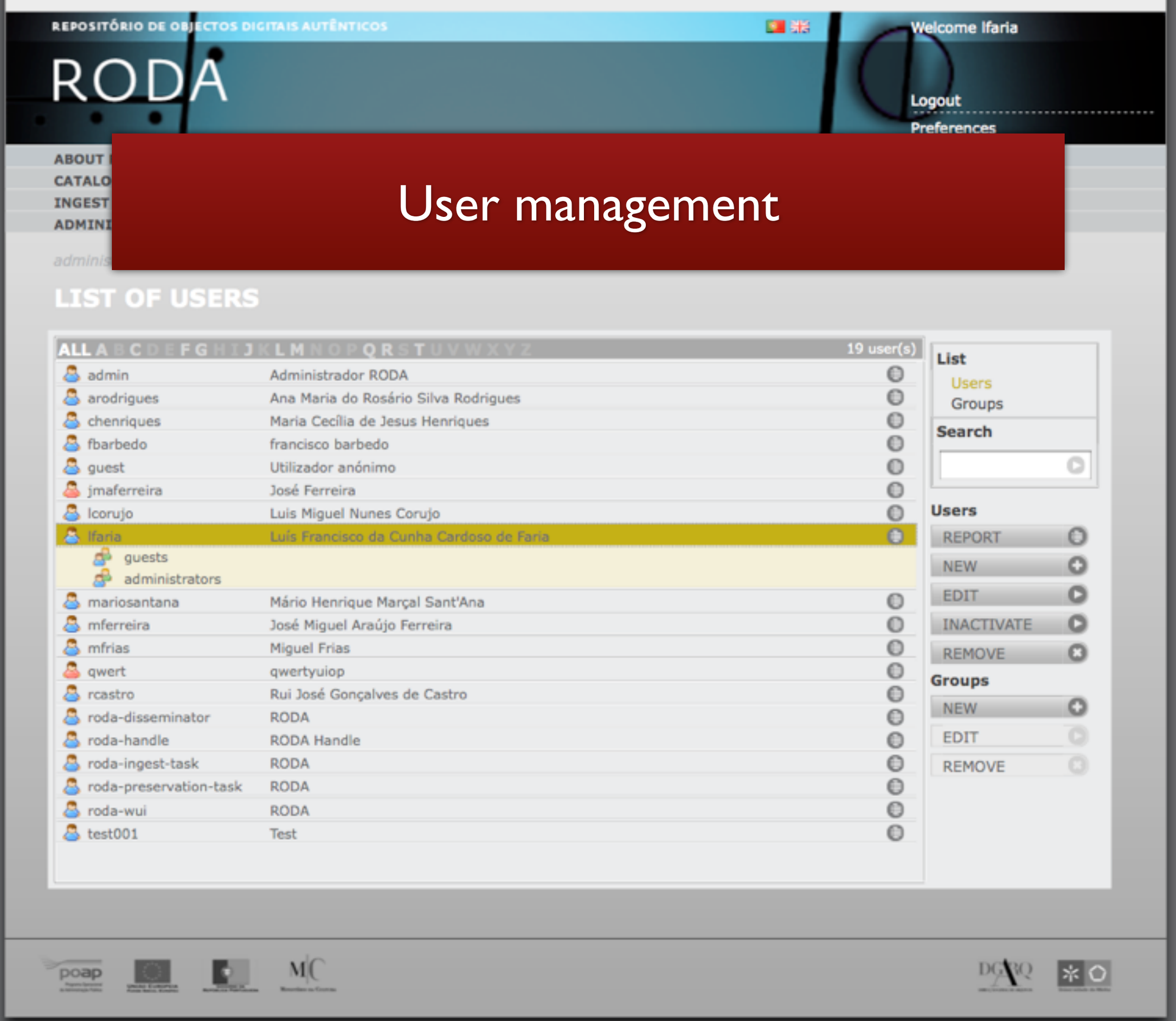
Administrators are notified on error occurrence


Schedule task
Name
Description
Start date
Now
Schedule
January
Repeat
no repeat
repeat each
seconds
until
forever
Plugin
Converter/AIFF to WAV (version 1)
Converter/RTF to PDF/A (version 1)
Converter/TGA to TIFF (version 1)
Converter/TIFF to TIFF (version 1)
Converter/WAV to WAV (version 1)
Converter/WMV to MPEG2 (version 1)
Converter/XPM to TIFF (version 1)
Description tools/Auto fill unit dates (version 1)
Description tools/Notify access restrict expiration (version 1)
Ingest Maintenance/Reject 'AUTHORIZED' SIPs (version 1)
Ingest Maintenance/Reject 'DROPED_FTP' SIPs (version 1)
Ingest Maintenance/Reject 'DROPED_UPLOAD_SERVICE' SIPs (version 1)
Ingest Maintenance/Reject 'SIP_INGESTED' SIPs (version 1)
Ingest Maintenance/Reject 'SIP_NORMALIZED' SIPs (version 1)
Ingest Maintenance/Reject 'SIP_VALID' SIPs (version 1)
Ingest Maintenance/Reject 'UNPACKED' SIPs (version 1)
Ingest Maintenance/Reject 'VIRUS_FREE' SIPs (version 1)
Ingest tools/Auto accept SIP (version 1)
Ingest tools/Notify producers (version 1)
Ingest/Check SIP syntax (version 1)
Ingest/Check producer authorization (version 1)
Welcome lfaria
ABOUT RODA
CATALOG
INGEST
ADMINISTRATION
Scheduler
Name
Estatísticas

Management features

Skip

KEEPSOLUTIONS
University of Minho SPIN-OFF


REPOSITÓRIO DE OBJECTOS DIGITAIS AUTÊNTICOS
RODA
Welcome lfaria
Logout
Preferences
ABOUT
CATALO
INGEST
ADMINI
User management
LIST OF USERS
ALL A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
19 user(s)
admin Administrador RODA
arodrigues Ana Maria do Rosário Silva Rodrigues
chenriques Maria Cecília de Jesus Henriques
fbarbedo francisco barbedo
guest Utilizador anónimo
jmaferreira José Ferreira
lcorujo Luis Miguel Nunes Corujo
lfaria Luís Francisco da Cunha Cardoso de Faria
guests
administrators
mariosantana Mário Henrique Marçal Sant'Ana
mferreira José Miguel Araújo Ferreira
mfrias Miguel Frias
qwert qwertyuiop
rcastro Rui José Gonçalves de Castro
roda-disseminator RODA
roda-handle RODA Handle
roda-ingest-task RODA
roda-preservation-task RODA
roda-wui RODA
test001 Test
List
Users
Groups
Search
Users
REPORT
NEW
EDIT
INACTIVATE
REMOVE
Groups
NEW
EDIT
REMOVE

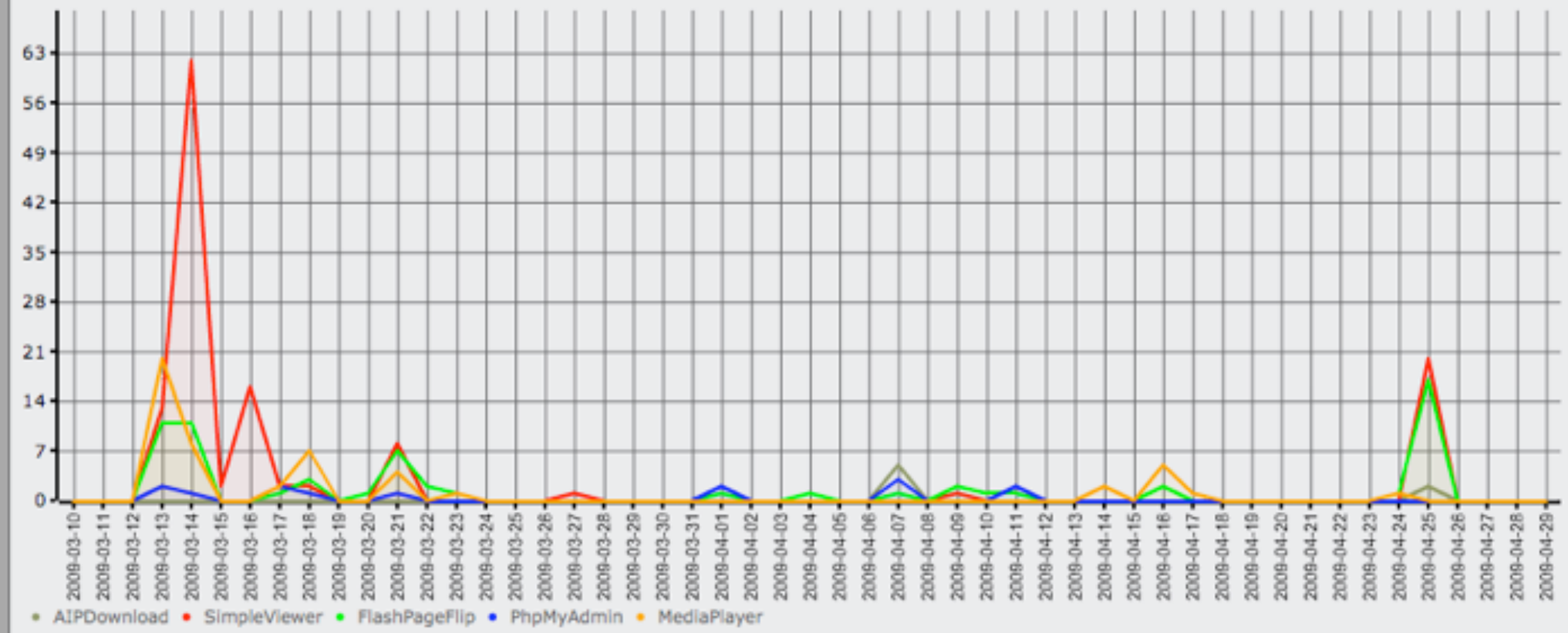
Reports & statistics

Distribution of disseminators accessed

Graph | Table

Unit of time: ○ years ○ months ◉ days

2009-03-10 at 2009-04-28



CLOSE

CATALOG
INGEST
ADMINISTRATION

Activity log
it enables system auditing

administration / log

LOG

Action

All actions

APPLY

Begin Date

January

End Date

January

327210 record entries

Date and Time	Action	Parameters	User
2009-05-12 12:07:35.0	RODAWUI pageHit	hostname: 193.136.19.138 address: 193.136.19.138 user: beta pagename: administration.log	lfaria
2009-05-12 12:05:10.0	Browser getProducers	doPID: roda:2	lfaria
2009-05-12 12:05:10.0	UserBrowser getUser	name: mferreira	lfaria
2009-05-12 12:05:10.0	UserBrowser getGroup	groupName: administrators	lfaria
2009-05-12 12:05:10.0	UserBrowser getGroup	groupName: producers	lfaria
2009-05-12 12:05:07.0	UserBrowser getUser	name: admin	lfaria
2009-05-12 12:05:07.0	UserBrowser getUser	name: fbarbedo	lfaria
2009-05-12 12:05:07.0	UserBrowser getUser	name: lfaria	lfaria
2009-05-12 12:05:07.0	UserBrowser getUser	name: mferreira	lfaria
2009-05-12 12:05:07.0	UserBrowser getGroup	groupName: administrators	lfaria
2009-05-12 12:05:06.0	Browser getDescriptionObject	pid: roda:2	lfaria
2009-05-12 12:05:06.0	Browser getRODAObjectUserPermissions	pid: roda:2	lfaria
2009-05-12 12:05:06.0	Browser getRODAObjectPermissions	pid: roda:2	lfaria
2009-05-12 12:05:05.0	RODAWUI pageHit	hostname: 193.136.19.138 address: 193.136.19.138 user: beta pagename: dissemination.browse.2	lfaria
2009-05-12 12:05:05.0	Logger addLogEntry	logEntry: LogEntry (datetime=null, address=192.168.111.154, username=lfaria, action=RODAWUI pageHit, parameters=	roda-wui

Do you want to know more about RODA?

KEEPSOLUTIONS
University of Minho SPIN-OFF

Documentation: available online

Help, development guide, scientific articles, mailing lists

Online demo

demo.roda-community.org

Commercial support

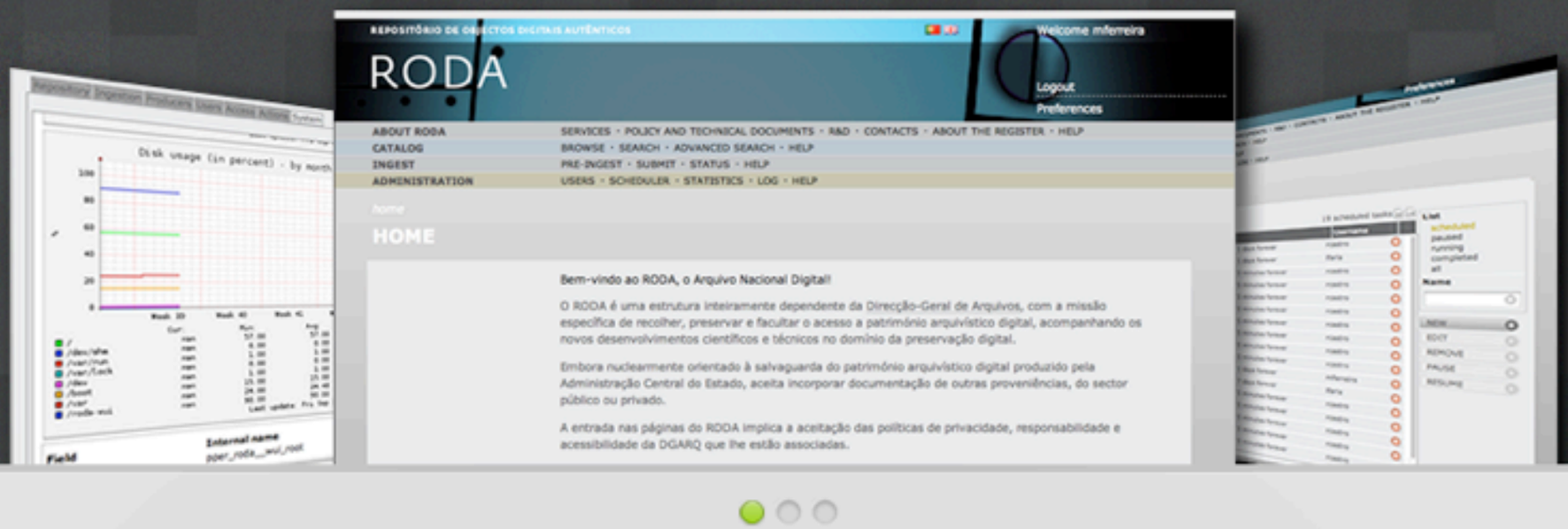
Download it and try it!

Source-code, binary, VM

Contributions: source code, translations, are all wellcome!

RODA

The world's most advanced **open-source** digital repository



An open-source **digital repository** designed for preservation

RODA is a complete digital repository that delivers functionality for all the main units of the OAIS reference model. RODA is capable of ingesting, managing and providing access to the various types of digital content produced by large corporations or public bodies. RODA is based on open-source technologies and is supported by existing standards such as the OAIS, METS, EAD and PREMIS.

Conforms to open standards

RODA follows open standards using EAD for description metadata, PREMIS for preservation

Vendor independent

RODA is 100% built on top of open-source technologies. The entire infrastructure required

www.roda-community.org

KEEPSOLUTIONS
University of Minho SPIN-OFF

How can we be sure we are doing things right?

Auditing and certification instruments

DRAMBORA
Digital Repository Audit Method Based on Risk Assessment

DSA
Data Seal of Approval

Nestor
Catalogue of Criteria for Trusted Digital Repositories

ISO 16363 (TRAC)
Audit and certification of trustworthy digital repositories

Auditing and certification instruments

DRAMBORA
Digital Repository Audit Method Based on Risk Assessment

DSA
Data Seal of Approval

Nestor
Catalogue of Criteria for Trusted Digital Repositories

ISO 16363 (TRAC)
Audit and certification of trustworthy digital repositories

Térmens, Miquel Las auditorías, una metodología para planificar la preservación digital. Experiencias en España. Ciencia da Informação, 2014, vol. 41, n. 1, pp. 140-142. [Journal article

Why should we certify a digital repository?

KEEPSOLUTIONS
University of Minho SPIN-OFF

To create a thrust and confidence climate around the repository and around the archived information
Producers, Consumers, Managers, Sponsors, etc.

To be able to demonstrate objectively how reliable is our repository

To be transparent before all players

To show that there are processes and procedures and that these are followed
based on evidence

How many of you thrust on your public administration information systems?

Will these systems be able to preserve your records for 20 , 30 or 40 years?

Can you imagine how many governments, policies and “crises” those information records will cross?

Can you imagine how many information systems, technologies, hardware architectures, etc., your records will face?

The last question is

How many people your government has thinking about these questions?

Auditing
national scientific system repositories

24 digital repositories were audited

Portuguese institutions belonging to the national scientific network

Shared infrastructure hosted by FCCN

26 repositories were consulted

2 repositories did not answer the call

Each one answered the first part of the standard (organizational structure)

FCCN answered parts 2 and 3 (Digital Object Management and Infrastructures and security)

Maturity average level was 2.0

Maturity lower level was 1.1

Maturity higher level was 2.8 (2 repositories)

4.0

is our goal

There was a set of requirements that no repository was able to demonstrate

3.1.2.1 - The repository shall have an appropriate succession plan…

3.1.2.2 - The repository shall monitor its organizational environment to determine when to execute its succession plan, contingency plans and/or escrow arrangements.

3.3.2 - The repository shall have Preservation Policies in place to ensure its Preservation Strategic Plan will be met.

3.3.2.1 - The repository shall have mechanisms for review, update and ongoing development of its Preservation Policies…

3.3.6 - The repository shall commit to a regular schedule of self-assessment and external certification.

3.5.2 - The repository shall track and manage intellectual property…

Currently there are no repository certification entities

One year later:

Maturity average level raised to 3.0

We have 6/26 repositories ready to be certified

First auditing raised awareness…

R&D

KEEPSOLUTIONS
University of Minho SPIN-OFF

SCAPE

Scalable Preservation Environments

INTERNET MEMORY FOUNDATION

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY

Universitatea de Vest din Timişoara

KB Koninklijke Bibliotheek National Library of the Netherlands

UPMC SORBONNE UNIVERSITÉS

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

KEEPSOLUTIONS

Microsoft Research

BRITISH LIBRARY

MANCHESTER 1824
The University of Manchester

ExLibris
The bridge to knowledge

Open Planets Foundation

Österreichische Nationalbibliothek

STATSBIBLIOTEKET

TU WIEN TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

BRNO UNIVERSITY OF TECHNOLOGY

Universität Berlin

Wielkopolskie Centrum Pulmonologii i Torakochirurgii
im. Eugenii i Janusza Zeylandów

Science & Technology Facilities Council

PSNC

Acronym: SCAPE

Duration: 3,5 years

Budget: € 12M €

Mission: To plan and execute preservation processes on large collections of data

Expected outcomes

- Open-source Software/tools giving digital preservation support
- Preservation workflows
- Physical infra-structure able to support scalable preservation solutions
- Preservation Good Practice Guides (repository migration and scientific data preservation)

http://www.scape-project.eu

Preservation Watch life-cycle

RODA, C3PO, SCOUT, PLATO, TAVERNA

Digital Preservation Toolkit

Um kit com mais de 80 ferramentas de migração, caracterização e controlo de qualidade

Scalable workflows

Ferramentas que permitem executar workflows sobre a tecnologia Hadoop para processamento em grande escala

4C

Collaboration to Clarify the Costs of Curation

DANS

DCC

EESTI RAHVUSRAAMATUKOGU
NATIONAL LIBRARY OF ESTONIA

THE ROYAL LIBRARY
NATIONAL LIBRARY OF DENMARK AND COPENHAGEN UNIVERSITY LIBRARY

KEEPSOLUTIONS
University of Minho SPIN-OFF

DEUTSCHE NATIONAL BIBLIOTHEK

Jisc

STATENS ARKIVER

University of Glasgow

inesc id lisboa

UK · DATA ARCHIVE

DigitalPreservationCoalition

secure
sba-research.org

Acronym: 4C

Duration: 2 years

Budget: 1.6M €

Mission: Answer the question - How does it cost to preserve?

Expected outcomes

- Mentality transformation - cost > investment, risk > benefit
- Methods to quantify preservation costs
- Platform for cost information exchange - CCEX

http://www.4cproject.eu

Investing in Curation

A Shared Path to Sustainability

(Draft)

Investing in Curation

http://www.4cproject.eu/roadmap

A Shared Path to Sustainability

(Draft)

www.curationexchange.org

Get started

Get acquainted with the Curation Costs Exchange, what it does and why

Compare costs

Add your curation costs and see how they compare with others

Understand your costs

Understand how to assess your curation costs and how to make use of cost models to help you invest

Read more

Browse books, papers and articles to help you get started in curation costing

Discuss and share

Share your experiences and read about challenges in other organisations like yours

Find services

See a list of digital curation tools and service providers, and find out more about what they offer

E-ARK

AUSTRIAN INSTITUTE OF TECHNOLOGY

ama AGÊNCIA PARA A MODERNIZAÇÃO ADMINISTRATIVA

University of Portsmouth

REPUBLIC OF SLOVENIA
MINISTRY OF CULTURE
ARCHIVES OF THE REPUBLIC OF SLOVENIA

STATENS ARKIVER
THE DANISH NATIONAL ARCHIVES

DLM FORUM FOUNDATION

DigitalPreservationCoalition

ES Solutions

KEEPSOLUTIONS

MAGENTA aps

TÉCNICO LISBOA

RIKSARKIVET

archívum

GOBIERNO DE ESPAÑA
MINISTERIO DE HACIENDA Y ADMINISTRACIONES PÚBLICAS

RAHVUSARHIIV
THE NATIONAL ARCHIVES OF ESTONIA

Acronym: E-ARK

Duration: 3 years

Budget: 6M €

Mission: To facilitate incorporation processes , preservation and access to archival information in digital format.

Expected outcomes

- To integrate electronic archiving processes at European level
- To create methodologies for document and database authenticity preservation
- To develop pilot applications to demonstrate results

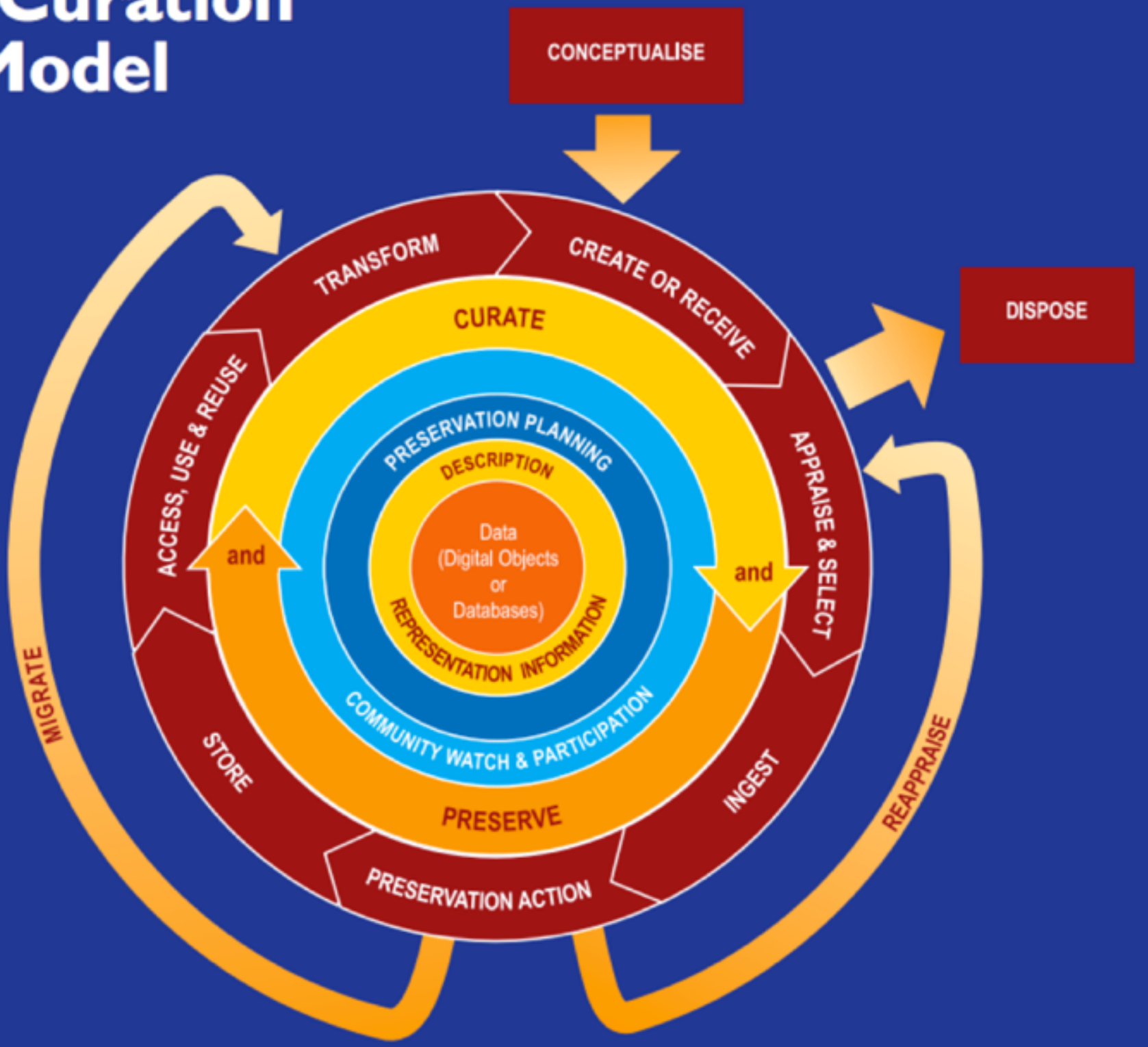
http://www.eark-project.com

SIP, AIP and DIP specifications (pan-european)

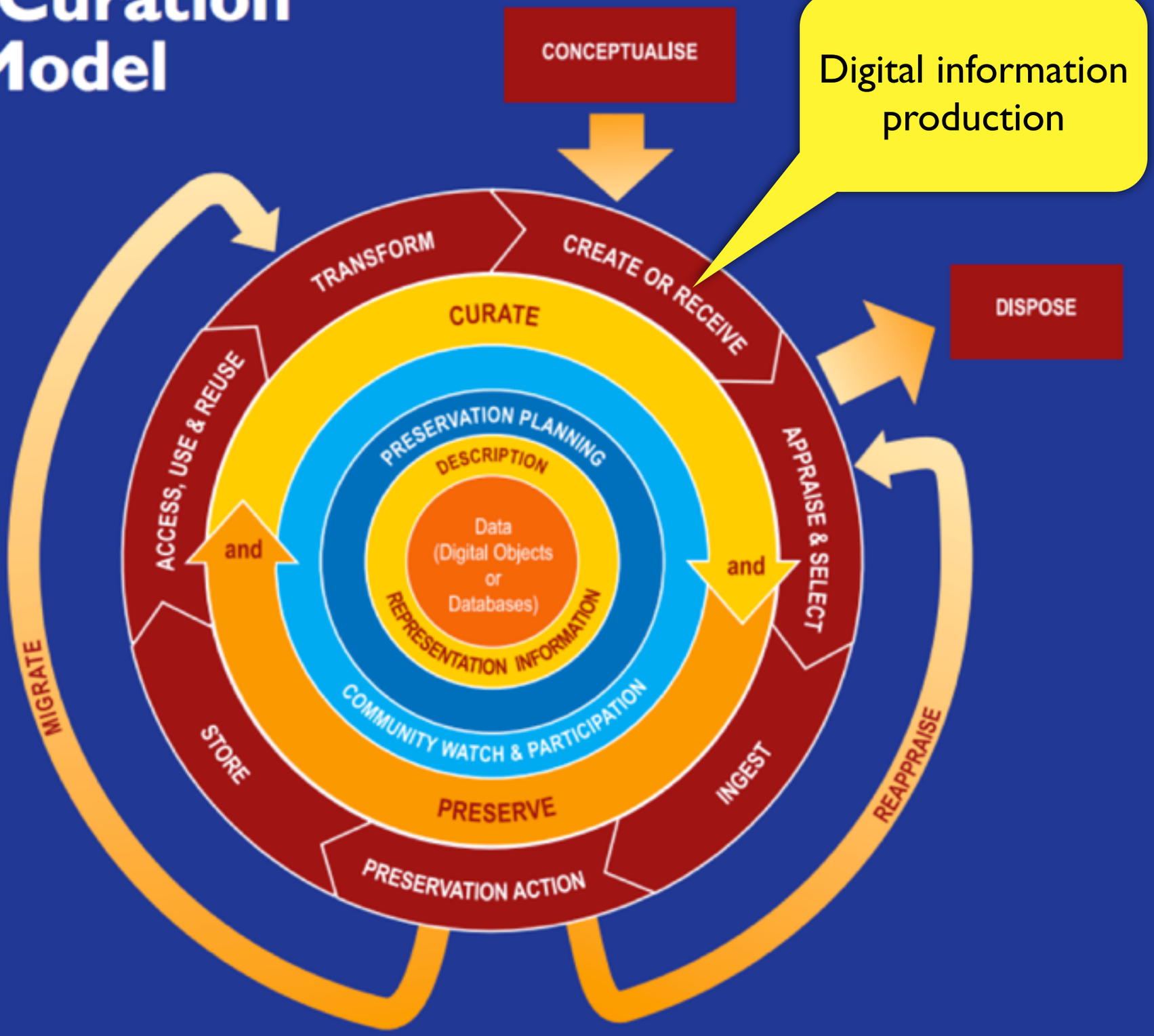
Databases preservation

still under development ...

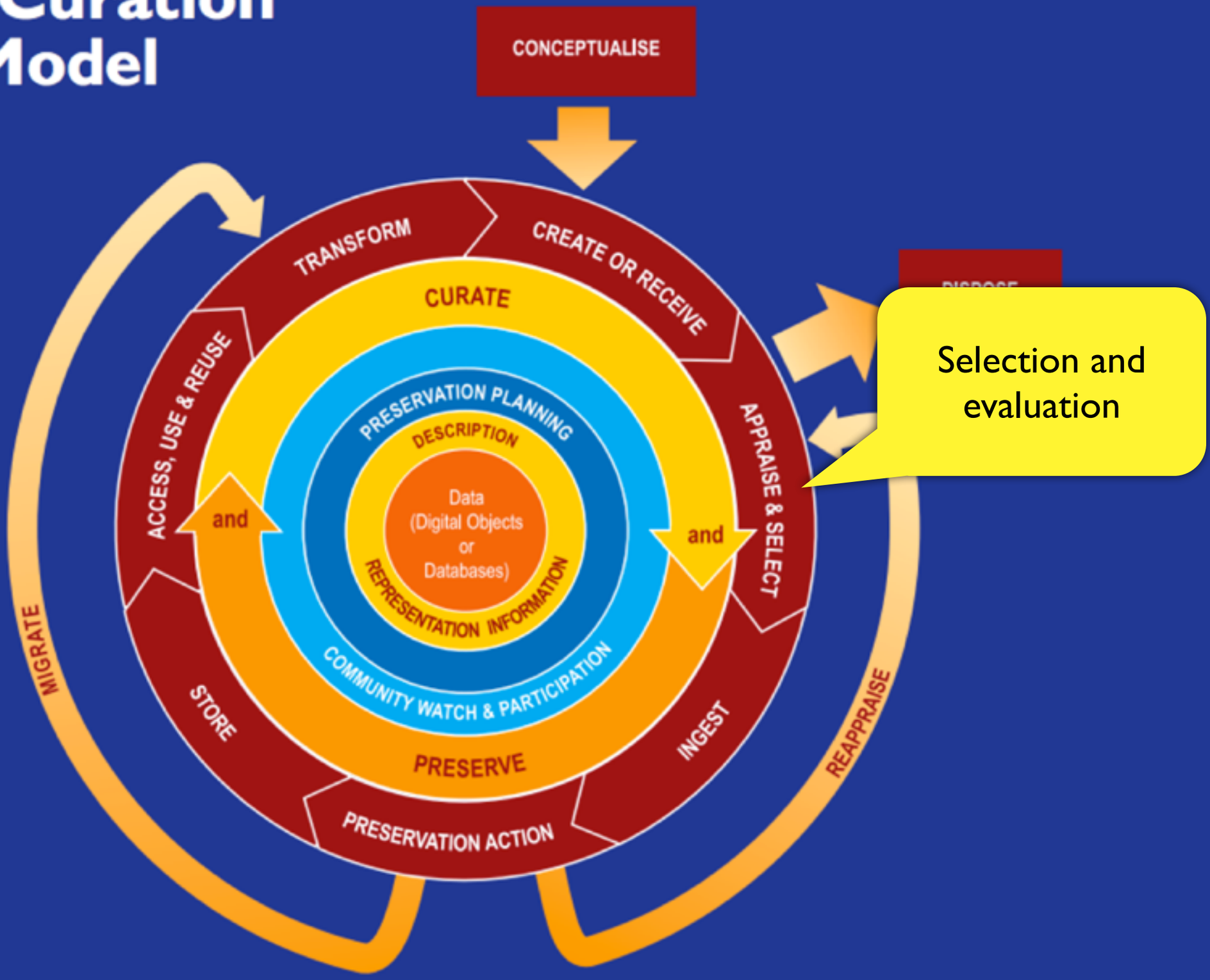
The DCC Curation Lifecycle Model



The DCC Curation Lifecycle Model

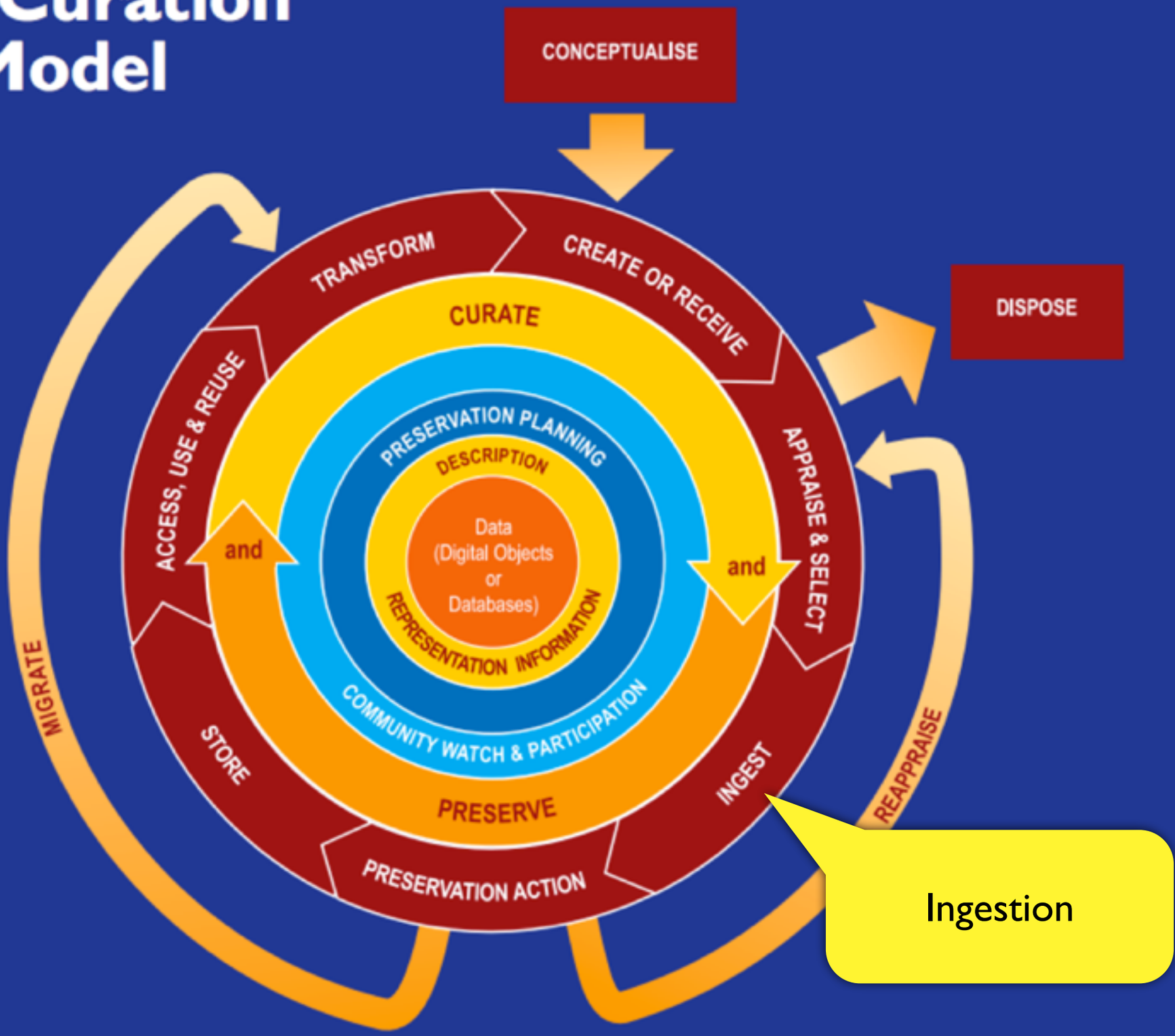


The DCC Curation Lifecycle Model



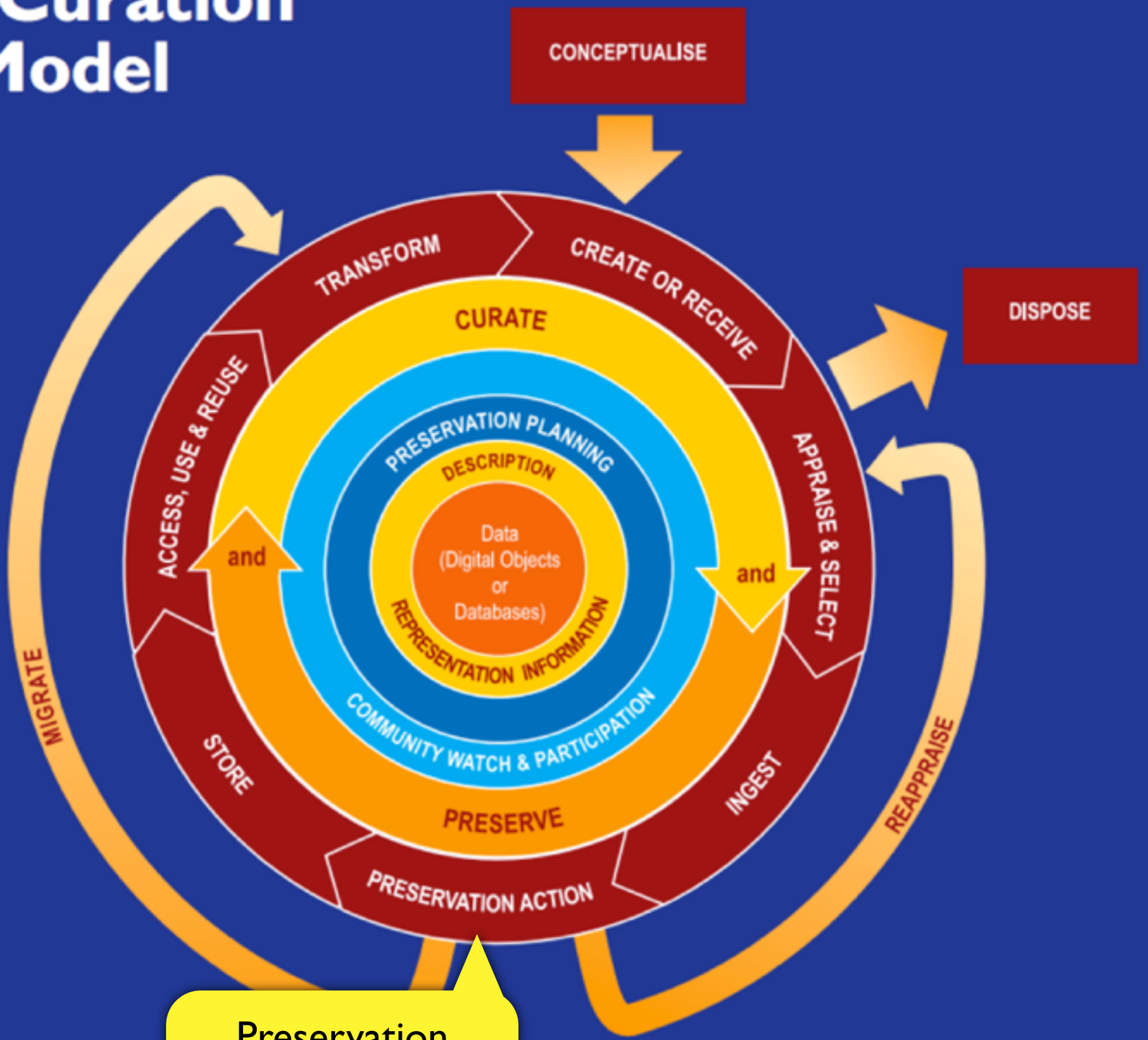
Selection and evaluation

The DCC Curation Lifecycle Model



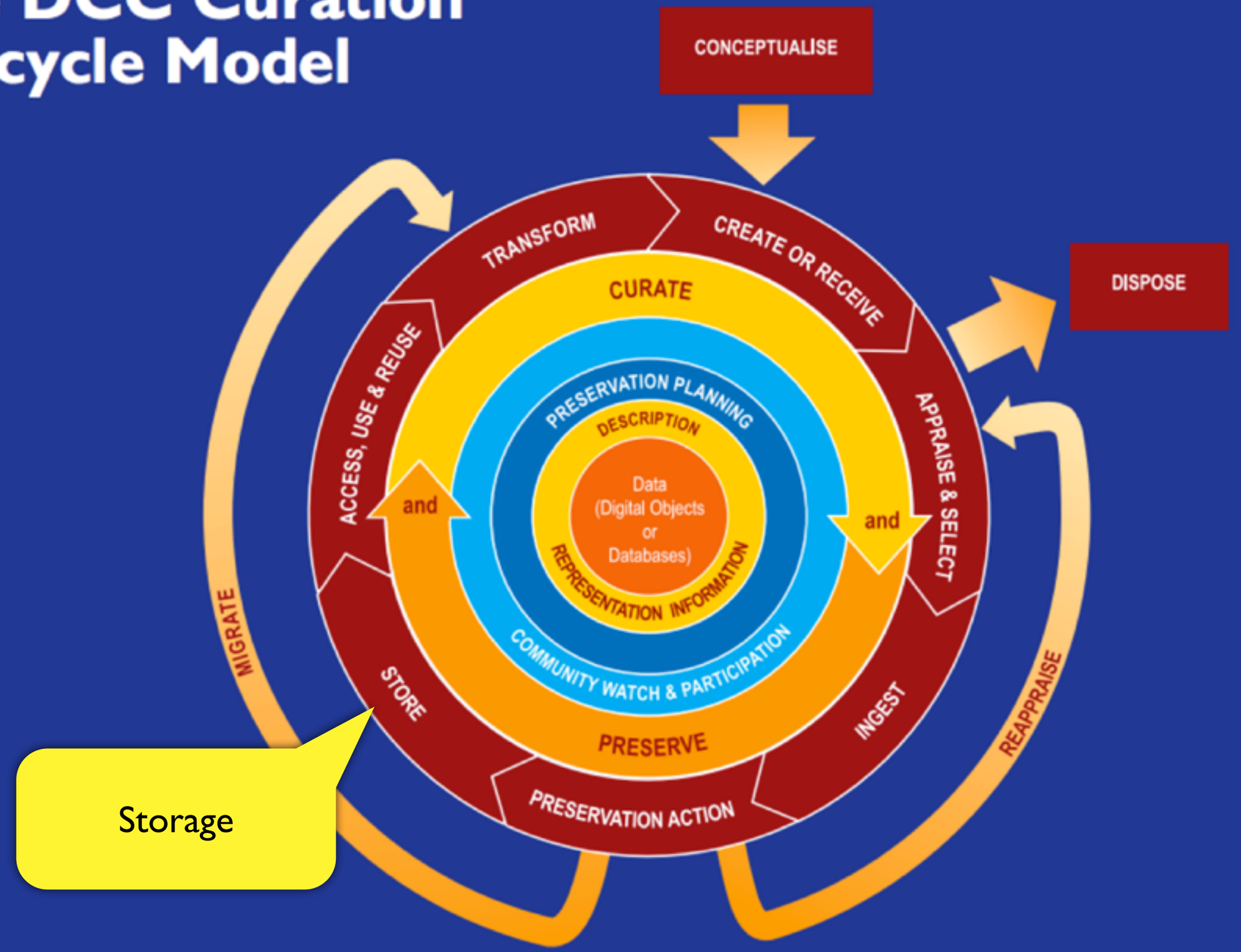
Ingestion

The DCC Curation Lifecycle Model



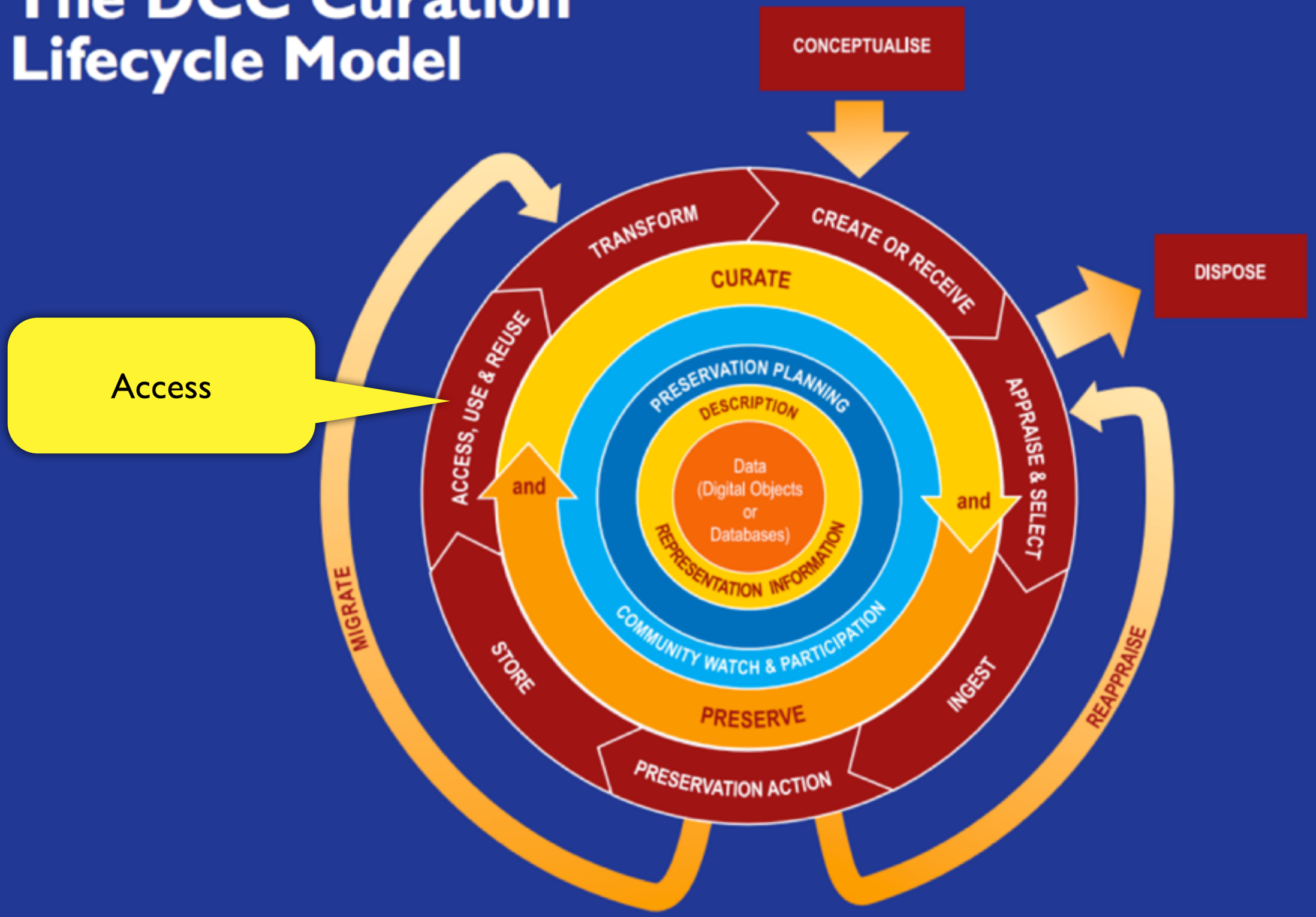
Preservation actions

The DCC Curation Lifecycle Model



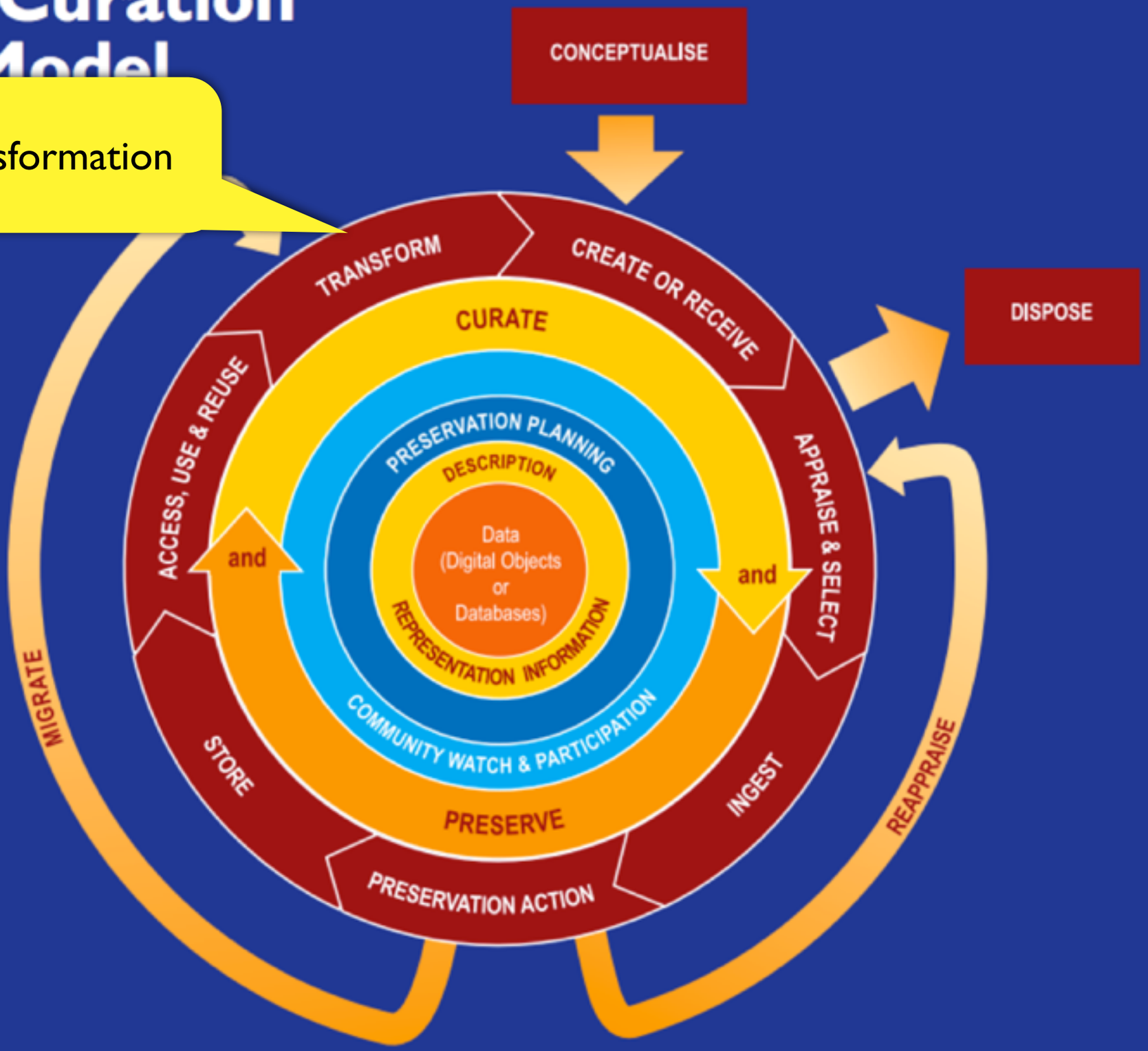
Storage

The DCC Curation Lifecycle Model



The DCC Curation Lifecycle Model

Transformation

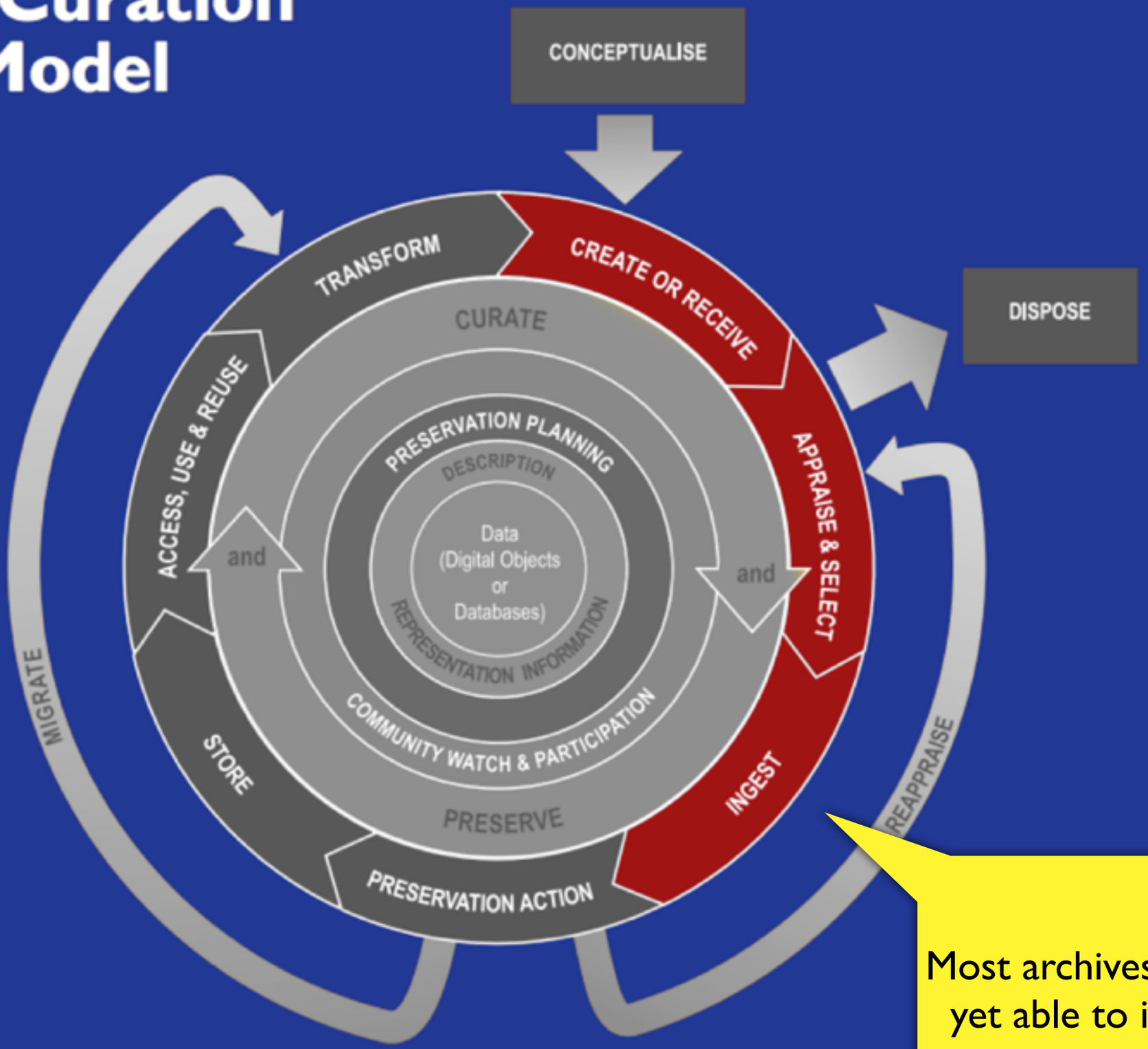


Preervation

Metadata,
planning,
technology watch,
etc...

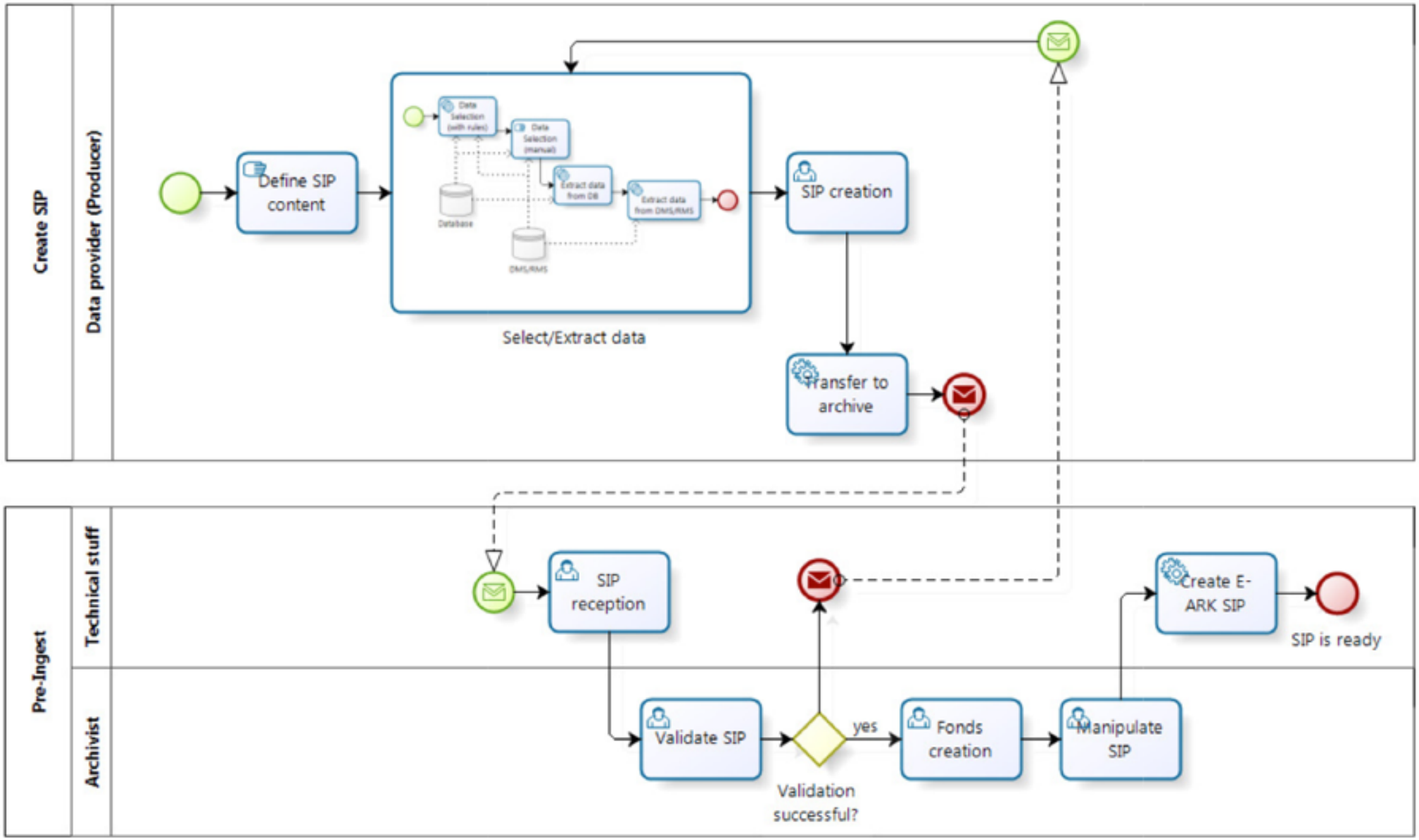
However...

The DCC Curation Lifecycle Model

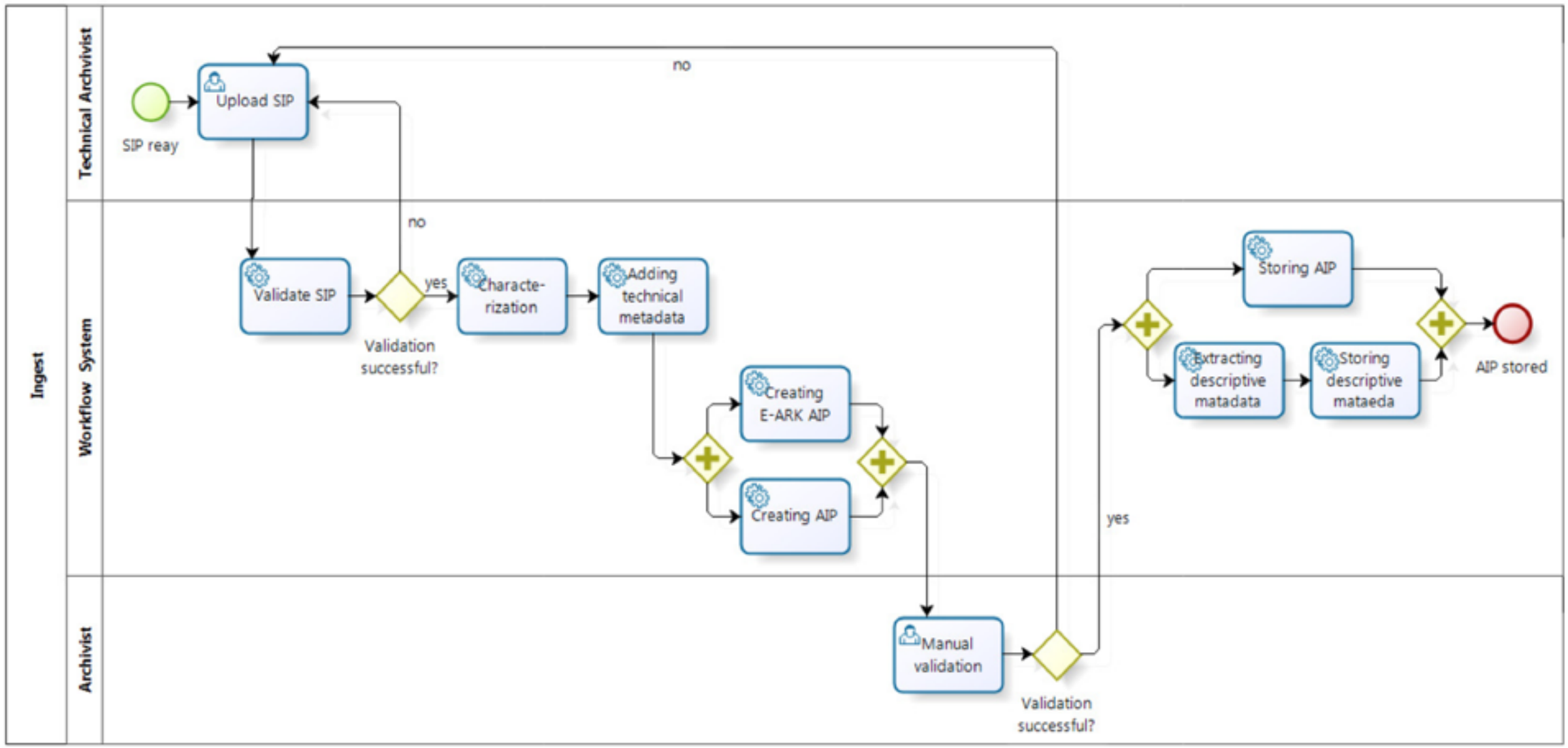


Most archives is not yet able to ingest

PRE-INGEST


Create SIP
Data provider (Producer)
Define SIP content
Data Selection (with rules)
Data Selection (manual)
Extract data from DB
Extract data from DMS/RMS
Database
DMS/RMS
Select/Extract data
SIP creation
Transfer to archive
Pre-Ingest
Technical stuff
SIP reception
Create E-ARK SIP
SIP is ready
Archivist
Validate SIP
Validation successful?
yes
Fonds creation
Manipulate SIP

INGEST


Ingest
Technical Archivist
Workflow System
Archivist
SIP reay
Upload SIP
no
Validate SIP
yes
Validation successful?
no
Characte-rization
Adding technical metadata
Creating E-ARK AIP
Creating AIP
Storing AIP
Extracting descriptive matadata
Storing descriptive mataeda
AIP stored
yes
Manual validation
Validation successful?

Some deliverables already available

DELIVERA

Project Acronym: E

Grant Agreement Number: 6

Project Title: E
Preservation

DELIVERABLE D

DELIVERABLE REFERENCE NO.	**D3.1**
DELIVERABLE TITLE	Report on availab
REVISION	**1.0**

Project 620998: European Archival Records and Knowledge Preservation - E-ARK

DELIVERABLE

Project Acronym: E-ARK

Grant Agreement Number: 620998

Project Title: European Archival Records and Knowledge
Preservation

DELIVERABLE DETAILS

DELIVERABLE REFERENCE NO.	**D3.2**
DELIVERABLE TITLE	E-ARK SIP Draft Specification
REVISION	**1.0**

Good-practice identification: pre-ingest, ingest and access

SIP, AIP e DIP specifications (DRAFT)

SIP specification will be available the end of this month

Legal restrictions study already started

…not over yet

Work on database ingestion is ongoing

Collaboration with the Federal Archives of Switzerland to launch SIARD 2.0

www.eark-project.com

Towards the future...

KEEPSOLUTIONS
University of Minho SPIN-OFF

Cloud data storge

- Remote data replication, without user awareness
- Access protocols to open data throughout Europe

“Elastic” repositories, i.e. they can adapt to peaks

International authentication and authorization services

Directories: services, tools, formats, technologies, suppliers, etc.

Cloud format migration services

Research data Preservation

Reuse, new value creation

Big data and linked data

Automatic appraisal and selection

Digitally signed document preservation

More and better preservation components

Put digital preservation at the beginning of digital object lifecycle

KEEPSOLUTIONS

University of Minho SPIN-OFF

www.keep.pt

José Carlos Ramalho
jcr@keep.pt

KEEPSOLUTIONS
University of Minho SPIN-OFF