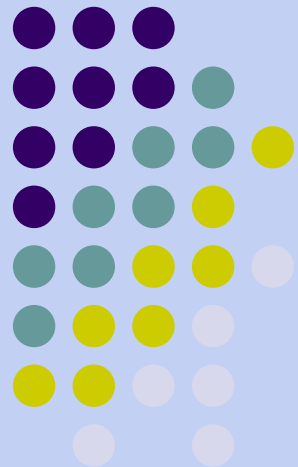
Guidelines and Tools for repository planning and assessment

Ann Green

Digital Life Cycle Research & Consulting

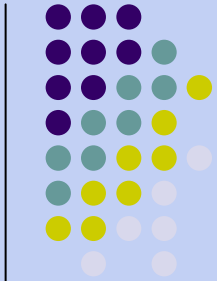
green.ann@gmail.com

dlifecycle.net



Presented at the DataShare project meeting
University of Edinburgh, Feb 5-6, 2008

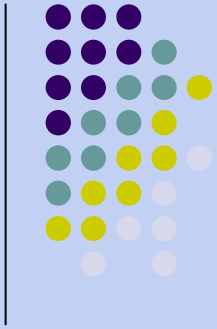
Guidelines and Tools for repository planning and assessment



Purposes:

- Define repository requirements, mandate, and scope
- Build policies
- Communicate scope, mission, policies, trustworthiness, and technological environment to other organizations and content providers
- Identify and compile ‘evidence’ of meeting requirements
- Undergo self assessment; identify and calculate risks
- Prepare for optional certification and audit
- Move policies into rules for implementation in repository software

OAIS Reference Model (Open Archival Information System)



An information model

- OAIS digital content “packages”
  - Submission Information Packages SIP
  - Archival Information Packages AIP
  - Dissemination Information Packages DIP
- Content Digital Object is described by
  - Representation Information (descriptive metadata --what you need to use and interpret the object) and
  - Preservation Descriptive Information ***PDI*** (what you need to preserve and access the object)

PREMIS:

PREservation Metadata Implementation Strategies

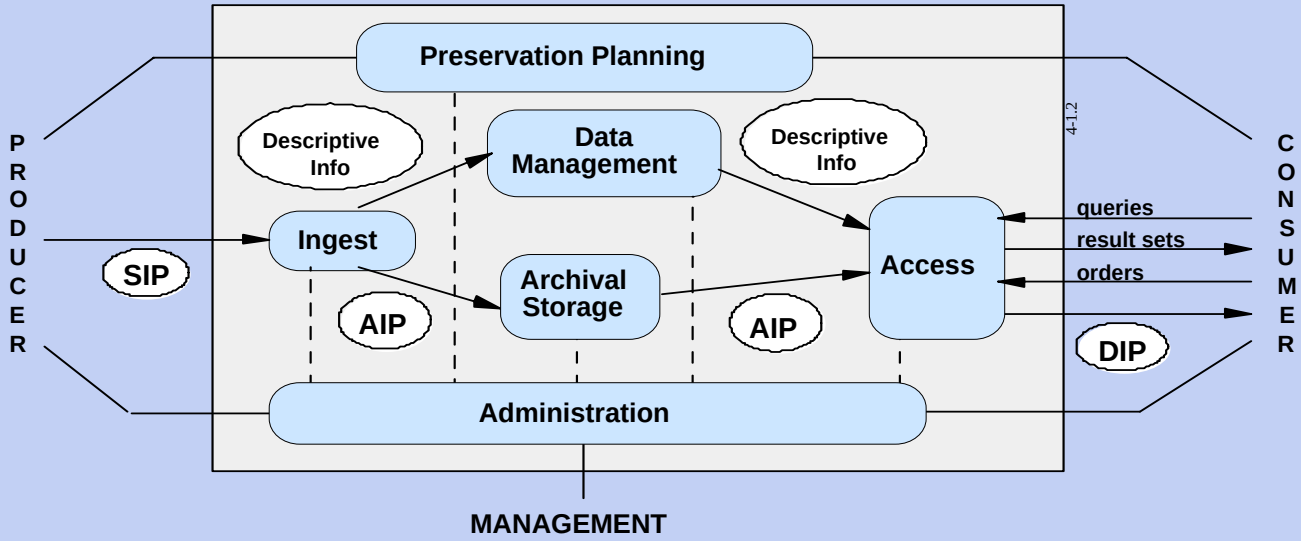
PREMIS took the PDI as a starting point; builds on OAIS concepts; offers an XML structure for preservation metadata

Provides:

- A core set of metadata elements
  - Data Dictionary for Preservation Metadata: May 2005.
  - Presents semantic units pertaining to Objects, Events, Rights and Agents.
- Guidelines and recommendations for management and use;
- Guidance for local implementations
- Standard for exchanging information packages between repositories
- ***Can be expressed in XML; METS is commonly used for packaging PREMIS metadata.***

SOURCE: DATA DICTIONARY FOR DIGITAL PRESERVATION: PREMIS TUTORIAL. Priscilla Caplan and Rebecca Guenther. University of Glasgow. July 17-19. Digital Curation Centre.

OAIS + PREMIS



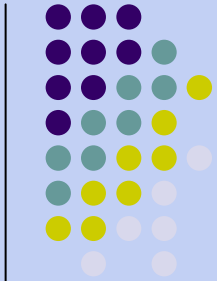
Assumes content arrives in SIPs and is stored in AIPs, and PREMIS is what the repository needs to know to ingest, store and preserve it for the future.

SOURCE: DATA DICTIONARY FOR DIGITAL PRESERVATION: PREMIS TUTORIAL. Priscilla Caplan and Rebecca Guenther. University of Glasgow. July 17-19. Digital Curation Centre.

	PREMIS metadata elements	Part of (if not main element)	PREMIS entity type	Comment/Description	Sources of this info?
1	objectIdentifier		Object	identifier of the eprint record (identifiers of the digital objects are their URLs)	
2	compositionLevel	objectCharacteristics	Object	e.g. compression, encryption, zip	
3	fixity	objectCharacteristics	Object	verifies if an object has been altered at the bit level by e.g. checksums	
4	size	objectCharacteristics	Object	size in bytes of the file or bitstream stored in the repository	
5	format	objectCharacteristics	Object	many preservation activities depend on detailed knowledge about the format of the digital object.	
6	significantProperties	objectCharacteristics	Object	may be objective technical characteristics subjectively considered important, e.g. pdf + links	
7	creatingApplication		Object	information about the creating application, including the version of the application	
8	originalName		Object	refers to files uploaded with the	

SOURCE: Applying preservation metadata to repositories. Steve Hitchcock. JISC Repositories Support Project. **21st January 2008**, British Library, London

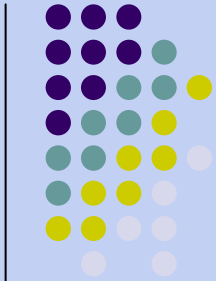
OAIS + PREMIS and digital repositories



1. OAIS provides a model for how content arrives in SIPs, is stored in AIPs, and is distributed in DIPs, described in Information Types
2. PREMIS provides a data dictionary about what the repository needs to know (metadata) to ingest, store and preserve it for the future
3. Guidelines for ***overall repository requirements***, (based upon OAIS and PREMIS) in regard to policies, organization, digital object management, security, and technical environment

   The essential attributes of digital repositories and the policies that guide them.

Tools and guidelines for digital repositories



Summary criteria:

- 10 Core Requirements for Digital Archives, Jan 2007
- DPC Handbook, in development since 2002

Audit and Certification Criteria:

- TRAC (checklist) March 2007
- Nestor (catalog) June 2006

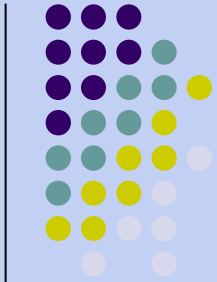
Self Audit and Risk Assessment:

- DRAMBORA (toolkit) March 2007

Policy development:

- DPC Tool for Selection of Materials March 2006
- OpenDoar: Open Access policy development 2006

Core Requirements for Digital Archives



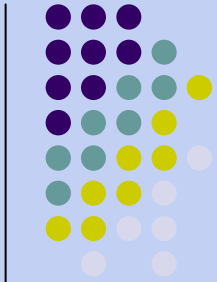
Top Ten Requirements

"preservation activities must be scaled to the needs and means of the defined community or communities"

Commitment to digital object maintenance
Organisational fitness
Legal & contractual rights
Effective & efficient policies
Acquisition & ingest criteria
Integrity, authenticity & usability
Provenance
Dissemination
Preservation planning & action
Adequate technical infrastructure

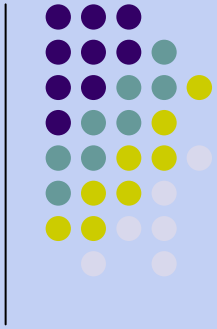
The Digital Curation Center, DigitalPreservationEurope, NESTOR, Center for Research Libraries, January 2007.

Core Requirements for Digital Archives



1. **The repository commits to continuing maintenance of digital objects for identified community/communities.**
2. **Demonstrates organizational fitness (including financial, staffing structure, and processes) to fulfill its commitment.**
3. **Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.**
4. **Has an effective and efficient policy framework.**
5. **Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.**
6. **Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.**
7. **Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.**
8. **Fulfills requisite dissemination requirements.**
9. **Has a strategic program for preservation planning and action.**
10. **Has technical infrastructure adequate to continuing maintenance and security of its digital objects.**

Policy development: DPC Handbook



Digital Preservation Coalition

http://www.dpconline.org/graphics/handbook/

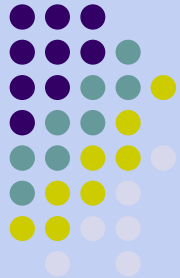
- “a bridge between broad, high level overviews and explicit, detailed guidelines applicable to the needs of a specific institution.
- a mechanism to help focus thoughts, increase overall understanding, promote training, and act as a catalyst for further action.”

Sections:

- Digital preservation
- Institutional strategies
- Organizational Activities
- Media and Formats

TRAC:

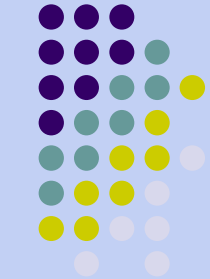
Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist



- **Organizational infrastructure**
- **Digital object management**
  - **Ingest: acquisition of content**
  - **Ingest: creation of the archivable package**
  - **Preservation planning**
  - **Archival storage & preservation/maintenance of AIPs**
  - **Information management**
  - **Access management**
- **Technologies, technical infrastructure, and security**

Concepts: **Designated Communities**
Trustworthiness
Evidence: policies, logs, business plans, etc.
Understandability

nestor:
Catalogue of Criteria for Trusted Repositories



- Importance of national conditions

  "…relates to the German context, manifested in terms of judicial constraints, the establishment of public institutions (in financial and human resource terms), national organizational decisions..."

  DCC: http://www.dcc.ac.uk/tools/nestor/

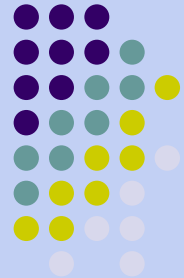
- Formal certification process
- From the nestor Working Group on Trusted Repositories Certification

http://www.wepreserve.eu/events/dr-2007/programme/presentations/wednesday/DCCDPEnestor_2007_3.pdf

10 Core Requirements For Trustworthy Digital Archives. Susanne Dobratz and Astrid Schoger. on behalf of the nestor-WG. DCC/DPE/DRIVER/nestor Joint Workshop. Berlin, 27-28 November 2007

nestor:

Catalogue of Criteria for Trusted Repositories



Content also in 3 parts:

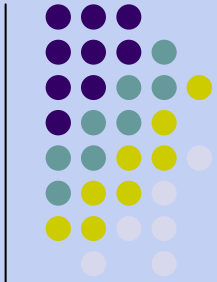
- Organisational Framework
- Object Management
- Infrastructure and Security

Conforms to OAIS model

Criteria include detailed explanations and concrete examples

Developed with the collaboration of DCC

nestor: 1.3 and TRAC, A3.1



1.3 The digital repository has defined its designated community(ies).

The general definition of the framework for a DR involves defining the designated community(ies) / designated community. This includes knowledge of the specific requirements of the designated community(ies) influencing the selection of the services to be provided. If the designated community or its requirements change over time, the DR should respond by adapting its services.

Possible designated communities include:

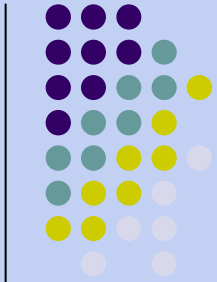
- Employees of an official body, a research institute etc.
- Scientists working in a particular discipline
- The general public

A3.1 Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met.

Evidence: Mission statement; written definitions of the designated community(ies); documented policies; service-level agreements.

DRAMBORA:

Digital Repository Audit Method Based on Risk Assessment

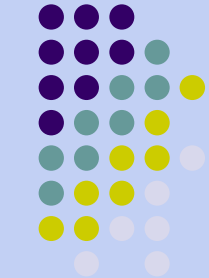


developed jointly by the DCC and Digital Preservation Europe

- ***Quantitative risk assessment***
- ***Toolkit*** for providing repository administrators with a self check (internal audit) of their digital repository.
- ***Each organization decides the scope*** of the audit and the benchmarks against which it will be assessed
- ***Methodology for self-assessment***
- "Fundamental aim is to help institutions to create, aggregate and present ***evidence*** that demonstrates that they are equipped to act as stewards for our digital resources."
- 24 pages of worksheets in Word, 10 worksheets in Excel
- Risk probability scale and risk impact scale: 1-6
- Suggests 14 characteristics for each risk description (name, date, stakeholders, nature of risk, etc.)
- Estimate of time required to complete the self audit: 24 – 40 hours

Source: http://www.dcc.ac.uk/resource/briefing-papers/trust-through-self-audit/#7

DRAMBORA:
6 stages and 2 groups of functions



STAGES of an assessment:

- **Identify organizational context**
- **Document policy and regulatory framework**
- **Identify activities, assets and their owners**
- **Identify risks associated with activities and assets**
- **Assess risks**
- **Manage risks**

Operational functions

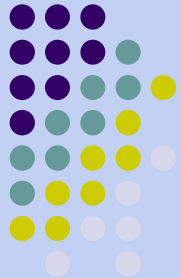
Acquisition & Ingest
Preservation & Storage
Metadata management
Access & Dissemination

Support functions

Organisation & Management; Staffing; Financial management; Technical infrastructure & Security

Source: Planning a digital preservation assessment using TRAC:CC and DRAMBORA. Blog entry by Disruptive Library Technology Jester, May 23rd, 2007. http://dltj.org/2007/05/trac-and-drambora/

Comparison of TRAC, DRAMBORA, and nestor



TRAC is an audit checklist; external certification driven. Can be used for goal setting, planning, policy building, assessment. Focus on preservation repositories.

DRAMBORA is a risk assessment toolkit; quantitative self assessment with risks ranked on scales; coverage similar to TRAC; focuses on broader range of repositories. Best for self review of existing repository, assess risks and measures to prevent them.

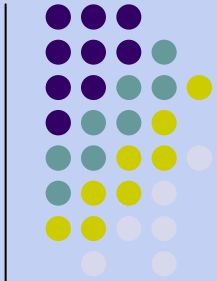
nestor is a catalog with references and notes; developed from a specific perspective and agenda. Can be used for planning, set up and evaluation.

They line up in coverage and level of detail.

Do not provide implementation specifics.

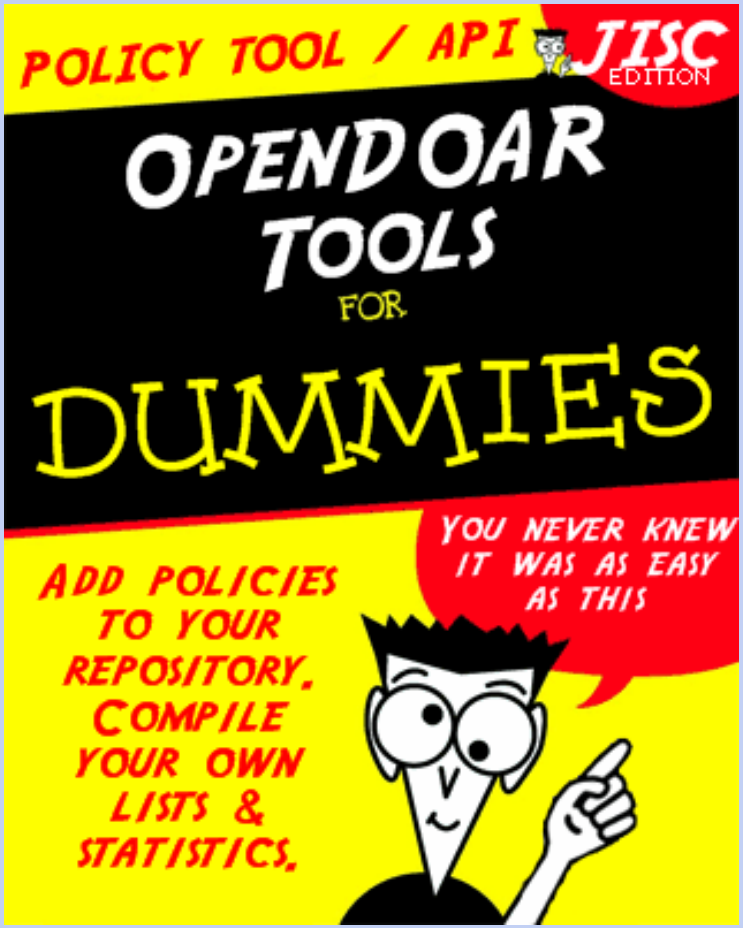
Source: http://www.wepreserve.eu/events/dr-2007/programme/presentations/wednesday/DCCDPEnestor_2007_3.pdf

OpenDoar Tool for Policies



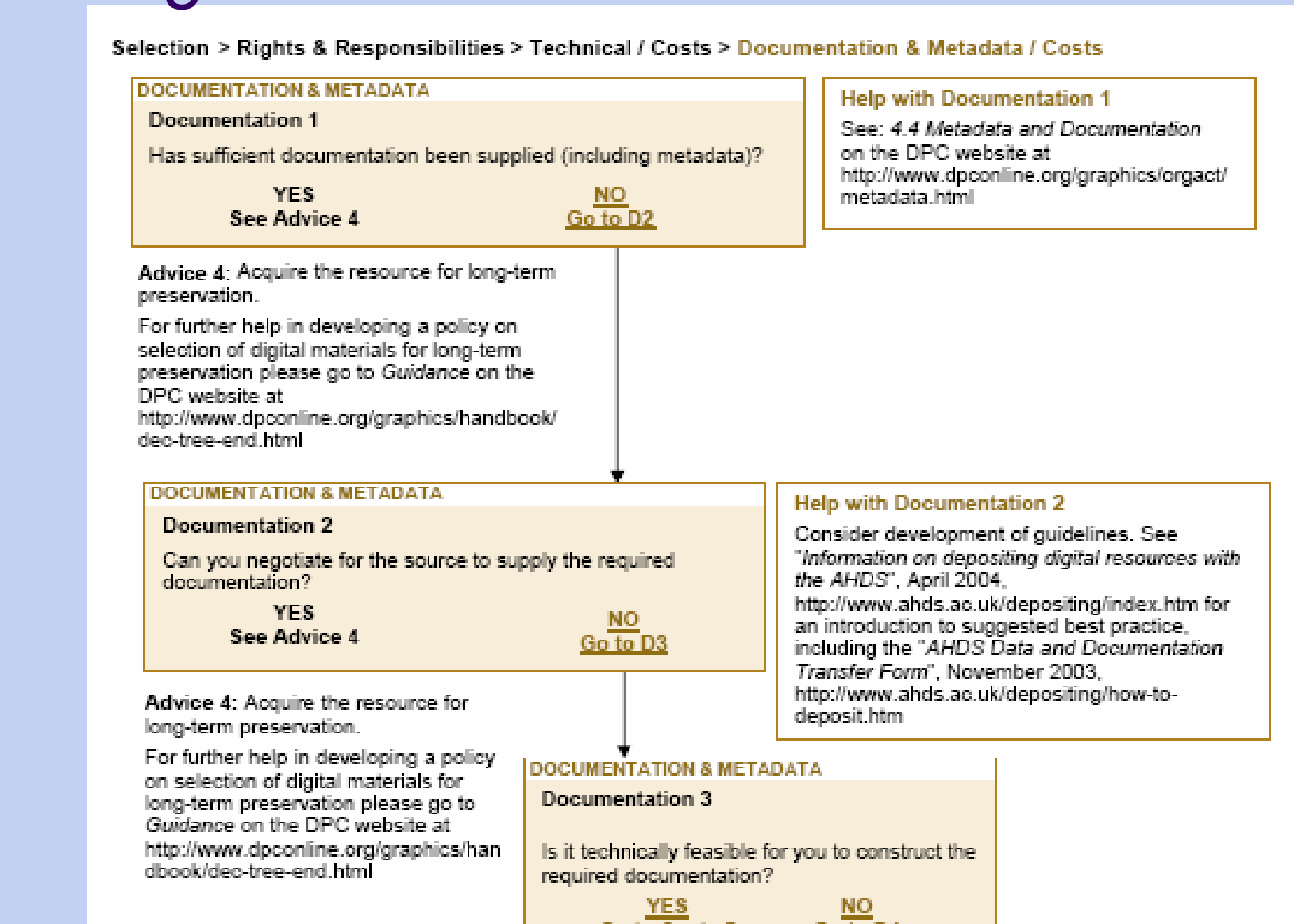
- http://www.opendoar.org/tools/en/policies.php
- Standard policies for Open Access repositories
- "simple (policy building) tool to help repository adminstrators to formulate and/or present their repository's policies. It provides a series of check boxes and pick lists for all the key policy options, which can be very quickly selected."
- generates source code for an EPrints static HTML page
- recommended options for minimum compliance with the aims of the Open Access movement

OpenDoar tool: policy coverage



- Metadata Policy- for information describing items in the repository.
- Data Policy - for full-text and other full data items.
- Content Policy - for types of document and dataset held.
- Submission Policy - concerning depositors, quality and copyright.
- Preservation Policy - Retention period; Functional preservation; File preservation; Withdrawal policy; Withdrawn items; Version control; Closure policy

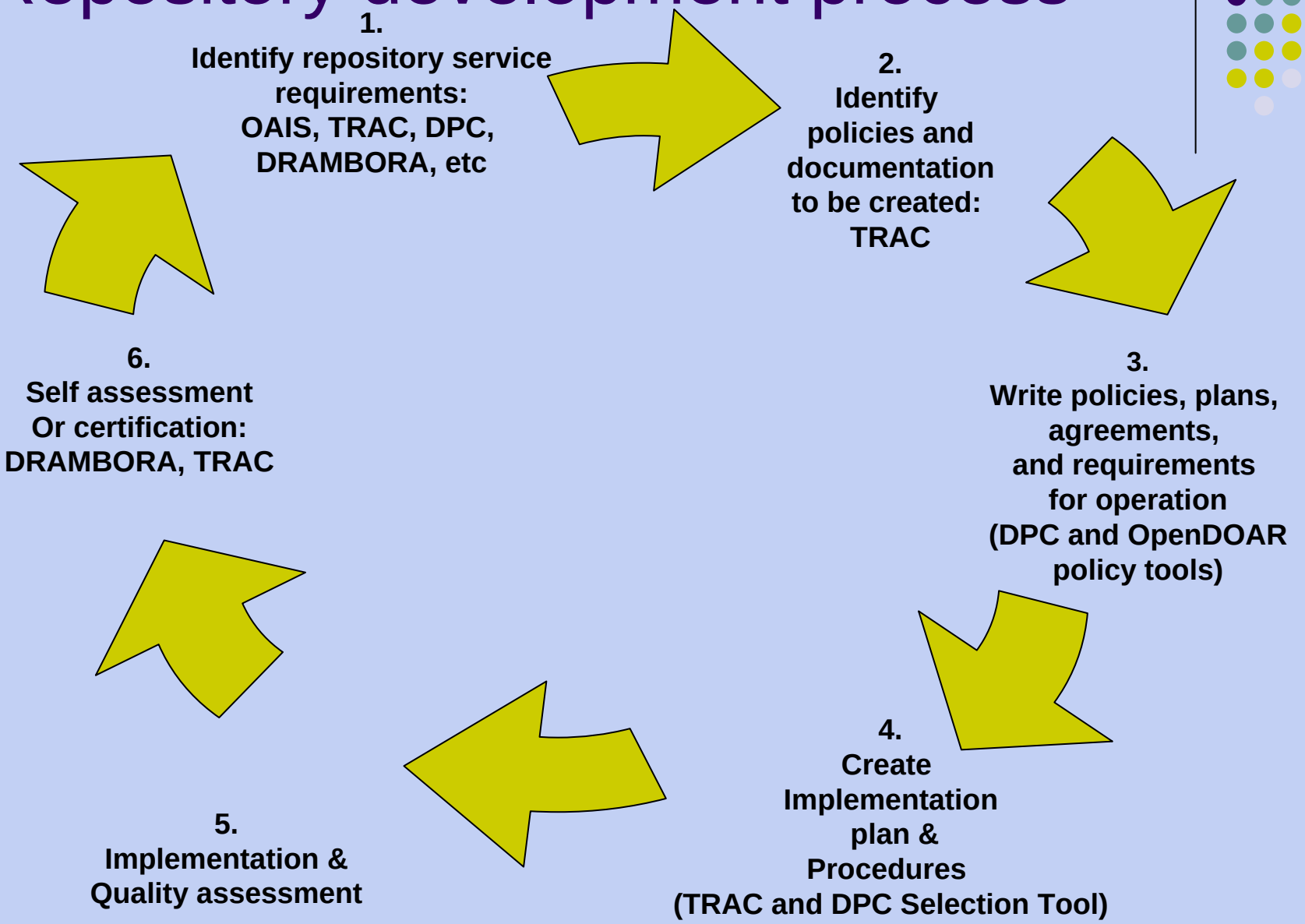
DPC Decision Tree for Selection of Digital Materials: Metadata



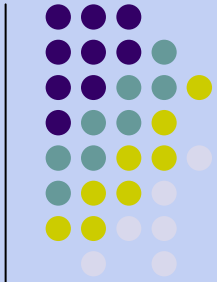
Fit to purpose

PURPOSE	DPC Handbook	TRAC	DRAMBORA	DPC-Tool for Selection	OpenDOAR Tool for Policies
Define requirements	x	x	x	x	x
Build policies	x	x	x	x	x
Self assessment		x	x		
Quantify risks and costs			x	x	
Communicate trustworthiness		x	x		
Help compile evidence for audit		x	x		x

Repository development process



PLEDGE:
from checklist to policy areas



Massachusetts Institute of Technology, University of California at San Diego Libraries, San Diego Supercomputer Center

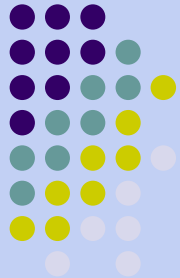
- **Step 1 = define policy areas (e.g. upper level TRAC); determine a list of which policies must be implemented (TRAC specifications)**

  Each policy area's title includes a policy number, a name and references to the earlier PLEDGE Policy List and the TRAC Checklist.

  - Example: ***CU-0008 Content Access (SA28, PR14, AR05, AR18, B5.1, B5.2, C3.3)***

Source: MIT Policy Definitions for Policies of Interest to Data. Robert Wolfe. http://pledge.mit.edu/images/b/bd/Federationpoliciesv3.pdf

PLEDGE:
from checklist to policy areas to policy statements



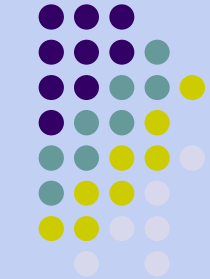
- **Step 2 = create policy declaration statements for each policy area; state the requirements for operation, not technical specifics**

  The abstract policy description is a general statement of a repository's policy characteristics and requirements in a given policy area.

  EXAMPLE: *Repository has documented and implemented access policies (authorization rules, authentication requirements) consistent with deposit agreements for stored objects. Repository ensures that agreements applicable to access conditions are adhered to. Repository access management system fully implements access policy.*

Source: MIT Policy Definitions for Policies of Interest to Data. Robert Wolfe. http://pledge.mit.edu/images/b/bd/Federationpoliciesv3.pdf

PLEDGE:
from policy statements to references



- **Step 3 = each entity in a policy statement is defined in language descriptions: humans and machine readable references**

  Descriptive statements take the subject-predicate-object form and are used to assign ownership of each policy.

  Descriptive statements are Statements of Assertion according to the iRods policy ontology.

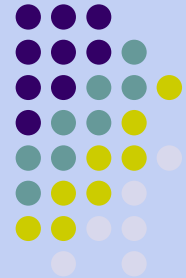
  Example: *DSpace @ MIT has access policy at http://libraries.mit.edu/dspacemit/build/policies/access.html*

PLEDGE: from policy statements to logical statements

- **Step 4 = deontic statements: logical statements define actors, actions, and constraints that enforce a policy statement.** These statements are what the repository will enforce, thus enforcing the policy.

No.	Actor	Action	Constraints	Deontic Type
1	Authorized DSpace Depositor	Deposit Item	1) Item has MIT accessible version, or 2) Item has sponsor prohibition, or 3) Item has time limit block	Permission
2	Repository	Audit MIT Accessible Version		Obligation
3	User	Access MIT Accessible Version	4) User is member of MIT community, and 5) Item has MIT accessible version	Permission

PLEDGE:
from logical statements to rules language (iRods)

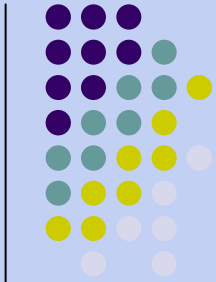


- **Step 5 = Policy Expression Languages: translate the English statements into syntax (RDF)**

In this step policy metadata is translated into a rules language that repository software can interpret.

http://pledge.mit.edu/images/b/bd/Federationpoliciesv3.pdf
http://pledge.mit.edu/images/d/d2/Dspacepoldefproc.pdf

references:
10 Core



- ***Ten core requirements of digital preservation repositories***
  http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=92

- ***10 Core Requirements For Trustworthy Digital Archives.***

  Susanne Dobratz and Astrid Schoger (ppt) at the DCC/DPE/DRIVER/nestor Joint Workshop, Berlin, 27-28 November 2007. http://www.wepreserve.eu/events/dr-2007/programme/presentations/wednesday/DCCDPEnestor_2007_3.pdf

references: TRAC

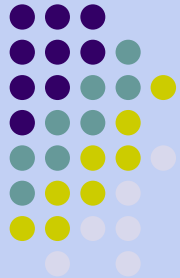
- ***Trustworthy Repositories Audit & Certification (TRAC) Criteria and Checklist***
  http://www.crl.edu/PDF/trac.pdf

- ***TRAC report template from Disruptive Technology Librarian*** -
  http://dltj.org/2007/08/trac-cc-report-template/

- ***MOIMS-rac*** (MOIMS-Repository Audit and Certification BOF): "A self-nominating working group working to propose an ISO standard for the auditing of Digital Repository's; using the TRAC checklist as the tool on which to base their proposal, and are currently revising the checklist for this purpose. *The TRAC checklist will be revised early in 2008 based upon the suggestions of this group*." (CRL website)
  http://wiki.digitalrepositoryauditandcertification.org/bin/view/Main/WebHome

- ***TRAC information on the CRL website:***
  http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91

references:
nestor and DRAMBORA



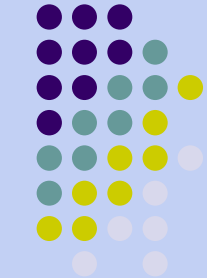
- ***nestor Catalogue of Criteria for Trusted Digital Repositories***
  http://www.nbn-resolving.de?urn:nbn:de:0008-2006060703
  http://www.langzeitarchivierung.de/

- ***DCC/DPE Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)***
  http://www.repositoryaudit.eu/download

- ***OpenDOAR Policies Tool***
  http://www.opendoar.org/tools/en/policies.php

references:

Digital Repository Audit and Certification BOF



Crosswalk file between TRAC, Nestor and DCC. Robin Dale. Mapping of Audit & Certification Criteria for CRL Meeing (15-16 January 2007)

http://wiki.digitalrepositoryauditandcertification.org/pub/Main/ReferenceInputDocuments/TRAC-Nestor-DCC-criteria_mapping.doc

Mandatory requirements working document:

http://wiki.digitalrepositoryauditandcertification.org/bin/view/Main/MandatoryReqts

Comparison Chart of RLG/NARA TRAC with NESTOR CCTDR, DCC/DPE DRAMBORA, ISO27001 and OECD Guidelines. Katia Thomaz. http://wiki.digitalrepositoryauditandcertification.org/pub/Main/DocAnalyses/ComparisonChart.doc