MAPPING DIGITAL PRESERVATION AND RDM CONCEPTS TOWARDS COLLECTIVE CURATION

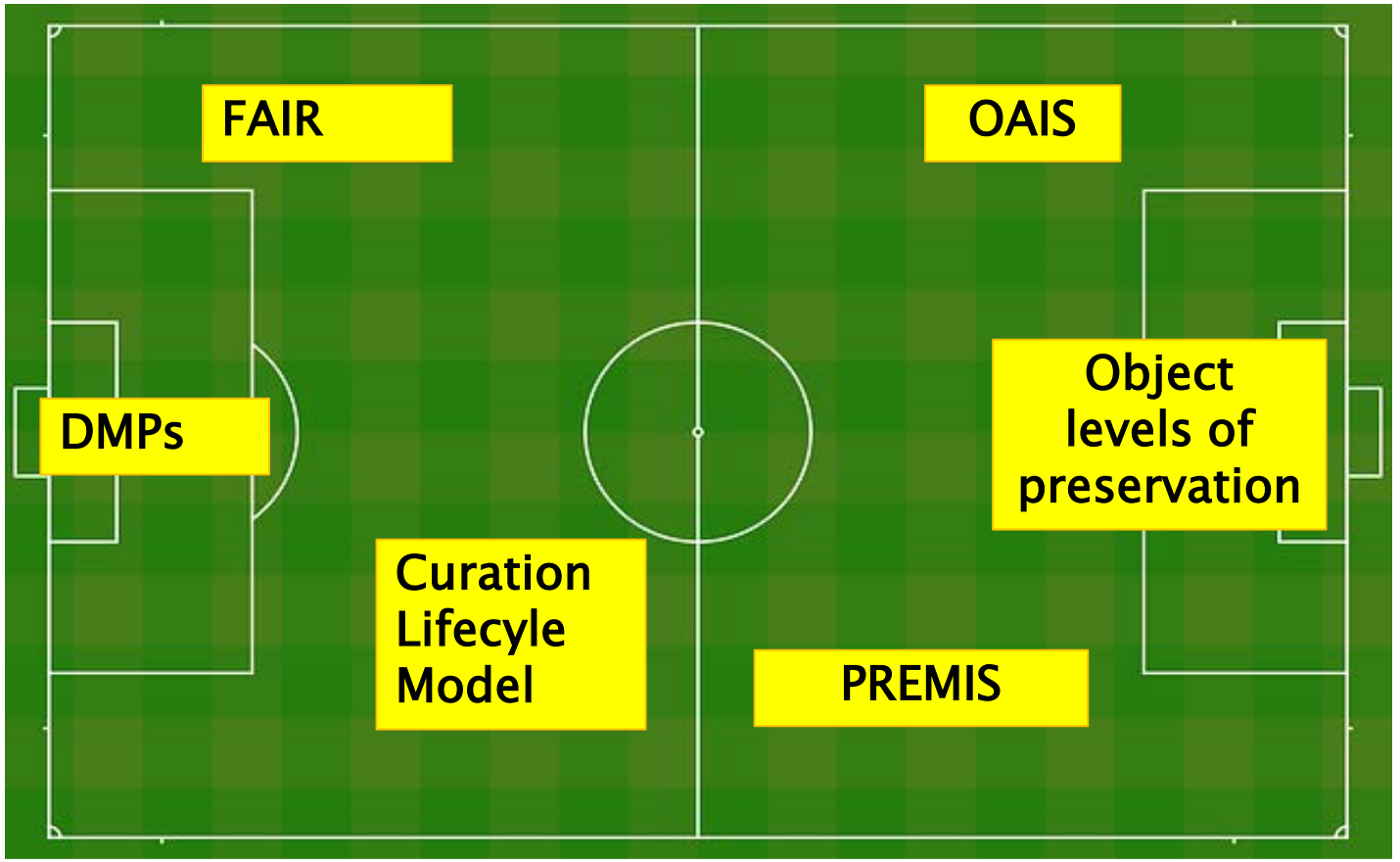
Michelle Lindlar
Pia Rudnik
Laurence Horton
Sarah Jones

https://doi.org/10.5281/zenodo.3664987

MAPPING SIX MODELS

Team RDM:
“Research Data Management Community”

DATA CREATION



FUTURE RE-USE

Team DP:
“Digital Preservation Community”

Let‘s change the game and play WITH each other!

OUR APPROACH

- Extract key terminology from models
- Discuss meaning from RDM and DP perspective and create mapping
- Analyse misalignments in mapping

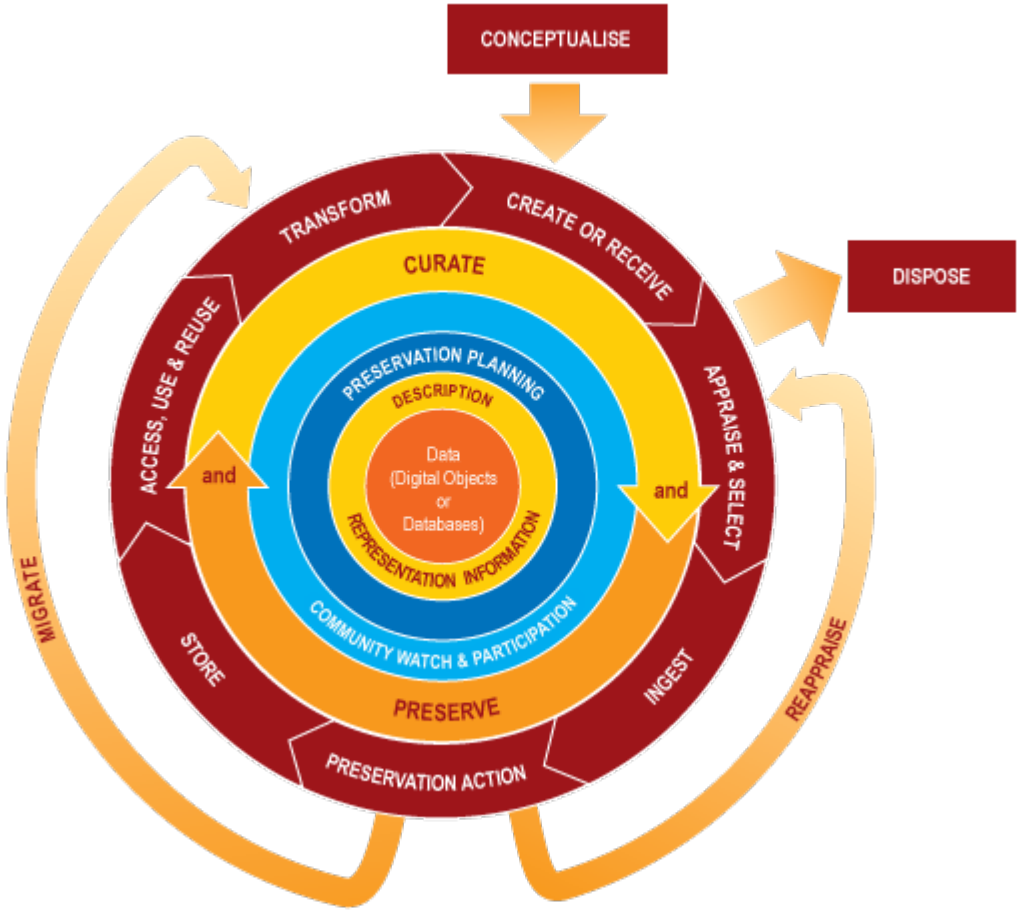
“YOU SAY POTATO, I SAY POTATO”

	DMPs	FAIR	DCC Lifecycle Model	OAIS	PREMIS	Object Levels of Preservation
What is being managed / preserved?	Data	Digital Objects	Digital Objects	Information Package (SIP, AIP, DIP)	(Digital) Object	Digital Object
				Digital Object	Bitstream	Physical Object
	Data Format	Standard Formats		Content Information	Representation File	Logical Object
					Intellectual Entity	Conceptual Object
Contextual information about target	Metadata and Documentation	Metadata Interoperable	Representation Information	Representation Information Preservation Description Information	Semantic Units	Properties of Object Classes
				Transformational Information Property / Significant Property	Significant Property	
	Persistent Identifiers	Findable / Persistent Identifiers		Data Management Functional Entity		
	Intellectual Property Rights	Usage Licence			Rights Statement	
Environment / Actors	Data Repository				Environment	
				Archive		
			Community Watch	Designated Community	Designated Community	
	Researcher		Create/receive	Producer	Agent	
		Reusable / Data User		Consumer	Agent	
			Appraise & select	Management	Agent	
Processes / Functions				Administration Functional Entity	Event	
			Ingest	Ingest Functional Entity	Event	
	Storage and Backup		Store	Archival Storage Functional Entity	Event	
	Preservation		Preservation Planning	Preservation Planning Functional Entity	Event	
	Data Sharing	Accessible	Access / Use / Reuse	Access Functional Entity	Event	
			Preservation action Transform		Event	

Same term, different meanings

Different terms, same meanings

CURATION LIFECYCLE MODEL



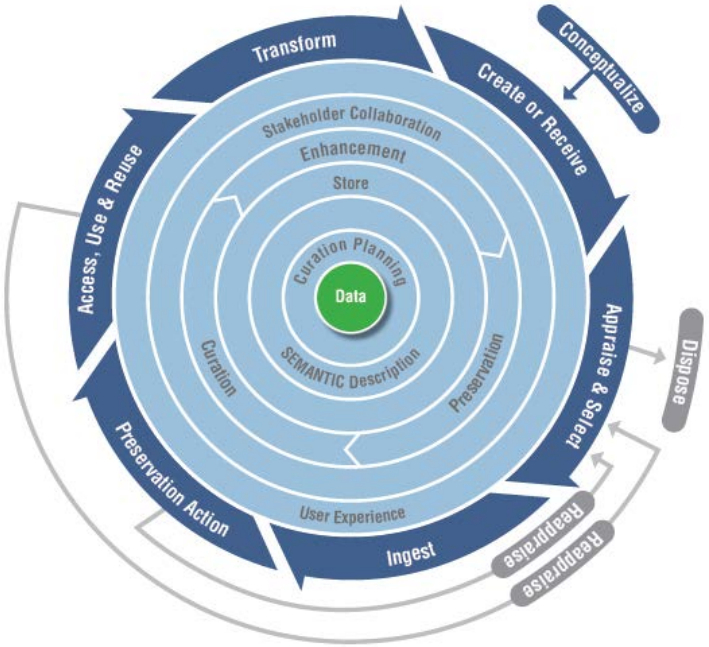
Key Terminology

- Management / Preservation Target:
  Digital object
- Contextual Information:
  Representation information
- Environment/Actors:
  Create, Appraise, Community watch
- Processes / Function:
  Ingest, Preserve, Access

http://www.dcc.ac.uk/resources/curation-lifecycle-model

KEY OBSERVATIONS

- Very comprehensive model:
  focuses on content creation and curation
- Too complex to present to researchers
  as many steps focus on role of curator
- Been widely adopted and iterated
  suggesting ongoing relevance



DATA MANAGEMENT PLANS

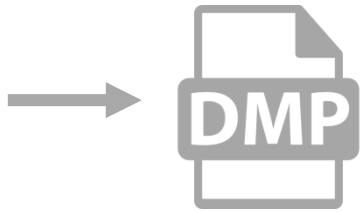
Key Terminology

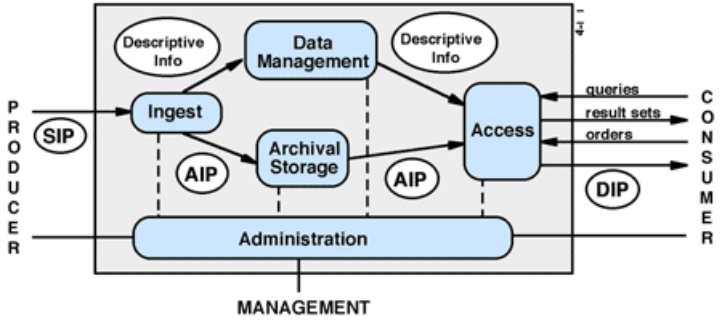
- Management / Preservation Target:
  Data, Data formats
- Contextual Information:
  Metadata, Persistent identifiers, Intellectual Property Rights
- Environment/Actors:
  Data Repositories, Researcher
- Processes / Functions:
  Storage and backup, Preservation, Data sharing

http://www.dcc.ac.uk/resources/data-management-plans/checklist

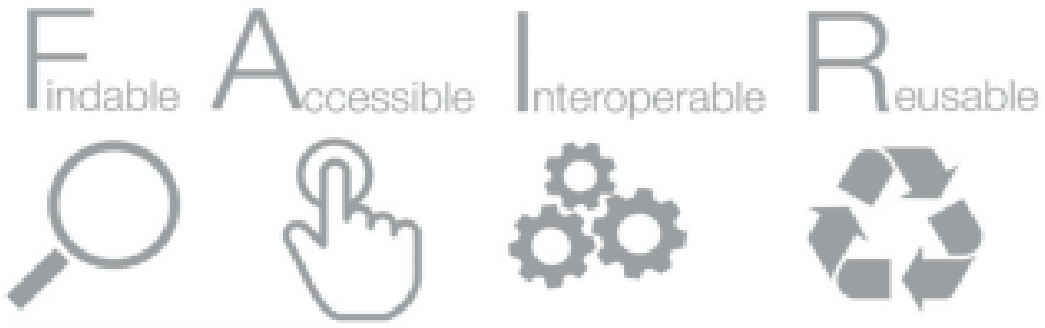
KEY OBSERVATIONS

- Potential for DMPs to be a bridge between content creation and curation communities
- DMPs no longer just researcher concern – increasing onus on role of institution
- Desire to extract information from DMPs to inform actions related to preservation / access
- Suggestion that DMP style liaison could be a useful pre-step to OAIS





FAIR



https://www.force11.org/group/fairgroup/fairprinciples

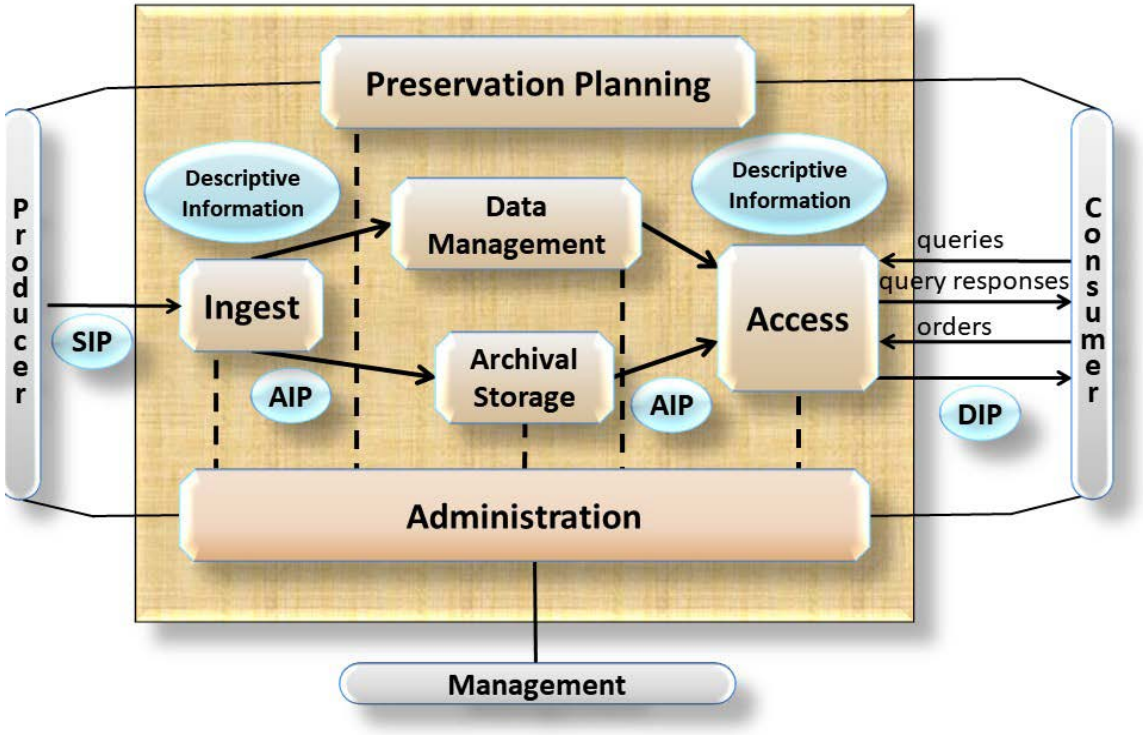
Key Terminology

- Management / Preservation Target: Digital objects, Standard formats
- Contextual Information: Interoperable / Metadata, Usage licence, Findable / Persistent identifiers
- Environment/Actors: Reusable / Data User
- Processes / Function: Accessible

KEY OBSERVATIONS

- Wide adoption but concepts are nothing new
- Missing digital preservation and data quality concepts
- Deposit in trusted repositories is a critical step that should be emphasised
- Could FAIR Data Objects will lead to sustainable Archival Information Packages?

OAIS



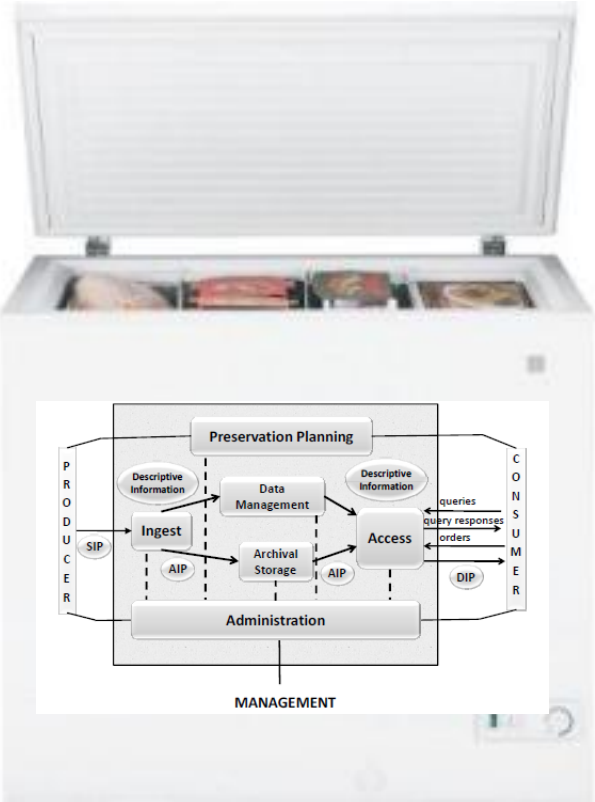
Key Terminology

- Management / Preservation Target:
  Information Packages (SIP, AIP, DIP), Digital Object, Content Information
- Contextual Information:
  Representation Information, Preservation Description Information, Transformational Information Property (Significant Property)
- Environment/Actors:
  Archive, Producer, Consumer, Management, Designated Community
- Processes / Function:
  Ingest / Archival Storage / Preservation Planning / Access / Data Management / Administration Functional Entities

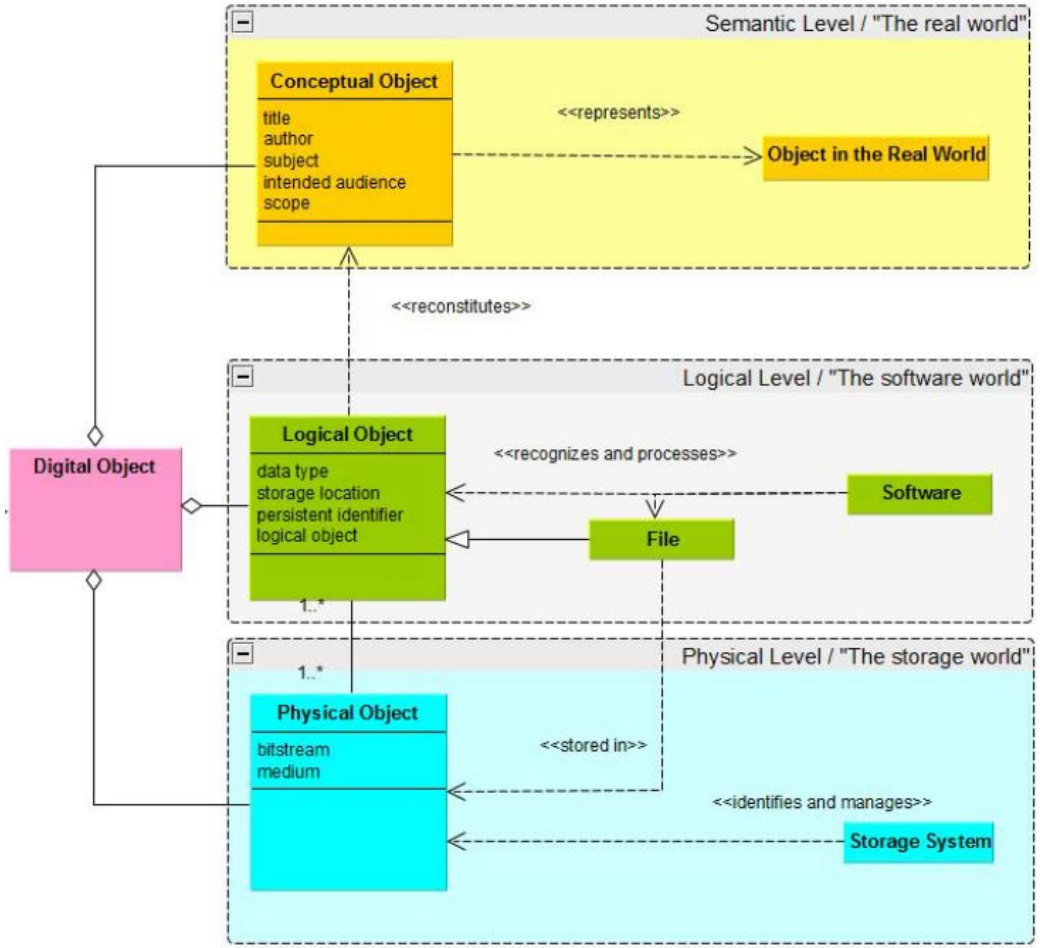
https://public.ccsds.org/Pubs/650x0m2.pdf (2012)

KEY OBSERVATIONS

- Heavily rooted in DP, little adoption in RDM
- Reference model = generic, not a blueprint for an implementation
- Complex, formalised model = not easy to understand for other domains
- Not without critique in DP domain
- RDM activity not really covered, but of high relevance to OAIS processes



OBJECT LEVELS OF PRESERVATION



Key Terminology

- Management / Preservation Target:
  Digital Object, Physical Object, Logical Object, Conceptual Object
- Contextual Information:
  Properties of Object Classes
- Environment/Actors:
  None
- Processes / Function:
  None

KEY OBSERVATIONS

- Layers of objects are recognised in DP practice / other models
- Granular approach to “Digital Object”, not found in other models
- Suitable as basic model to map risks, mitigation strategies and responsibilities
- Neglects actors and processes

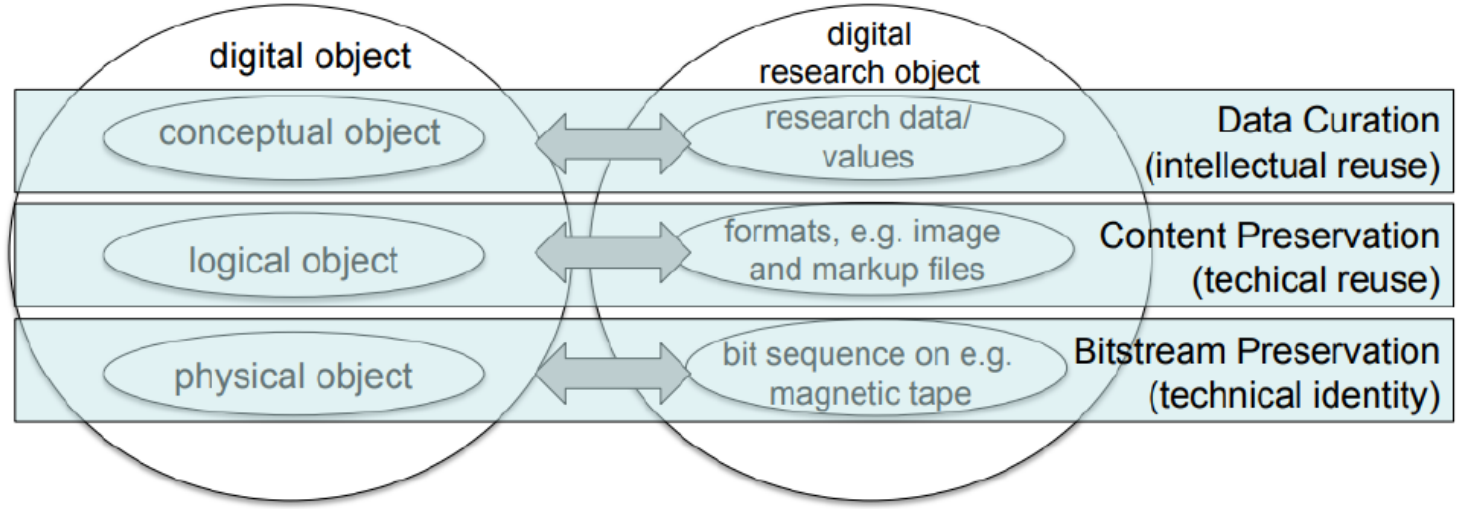
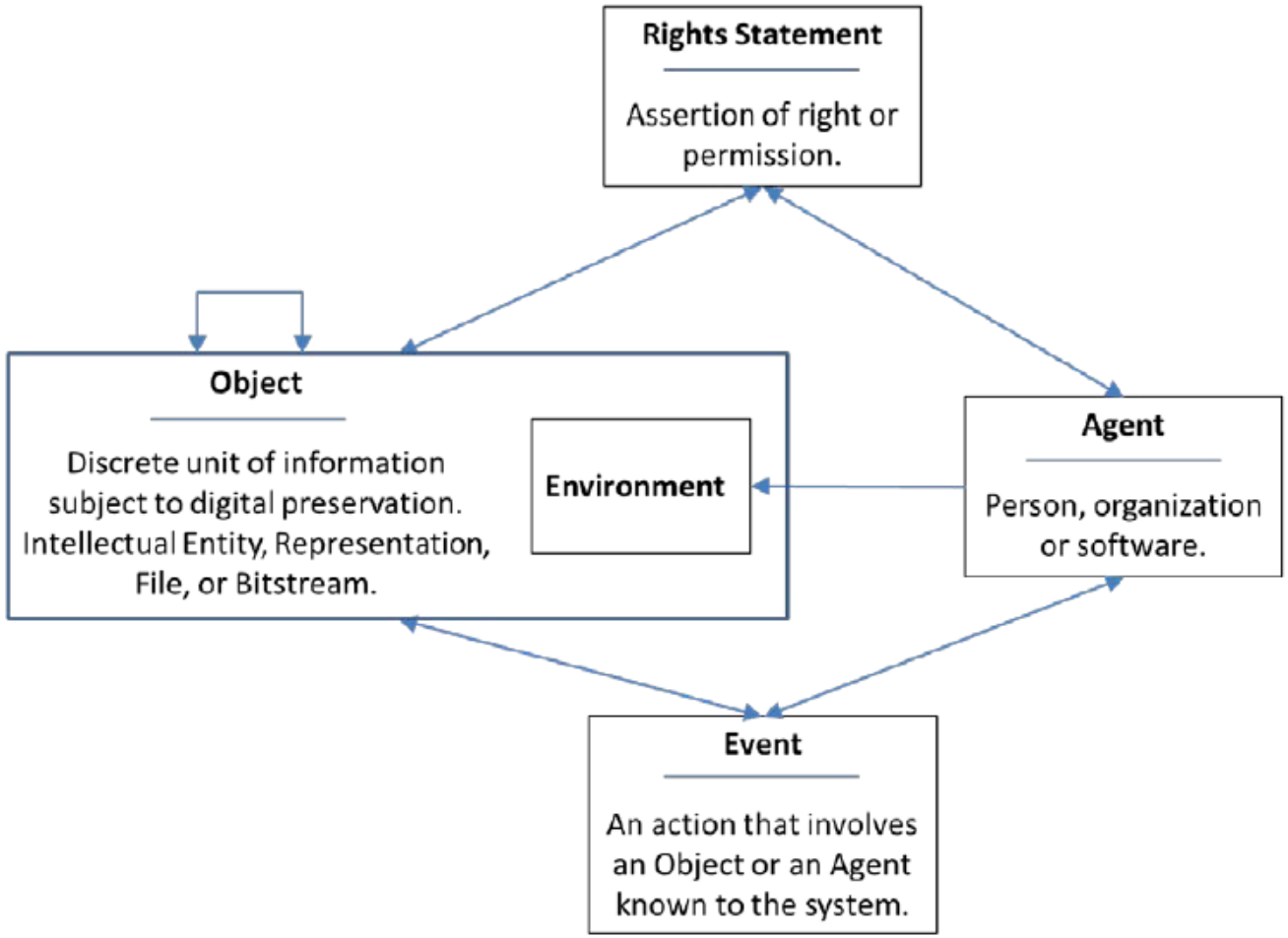


Figure 4: Curation Levels and Digital Objects in WissGrid (Ludwig, 2009

PREMIS

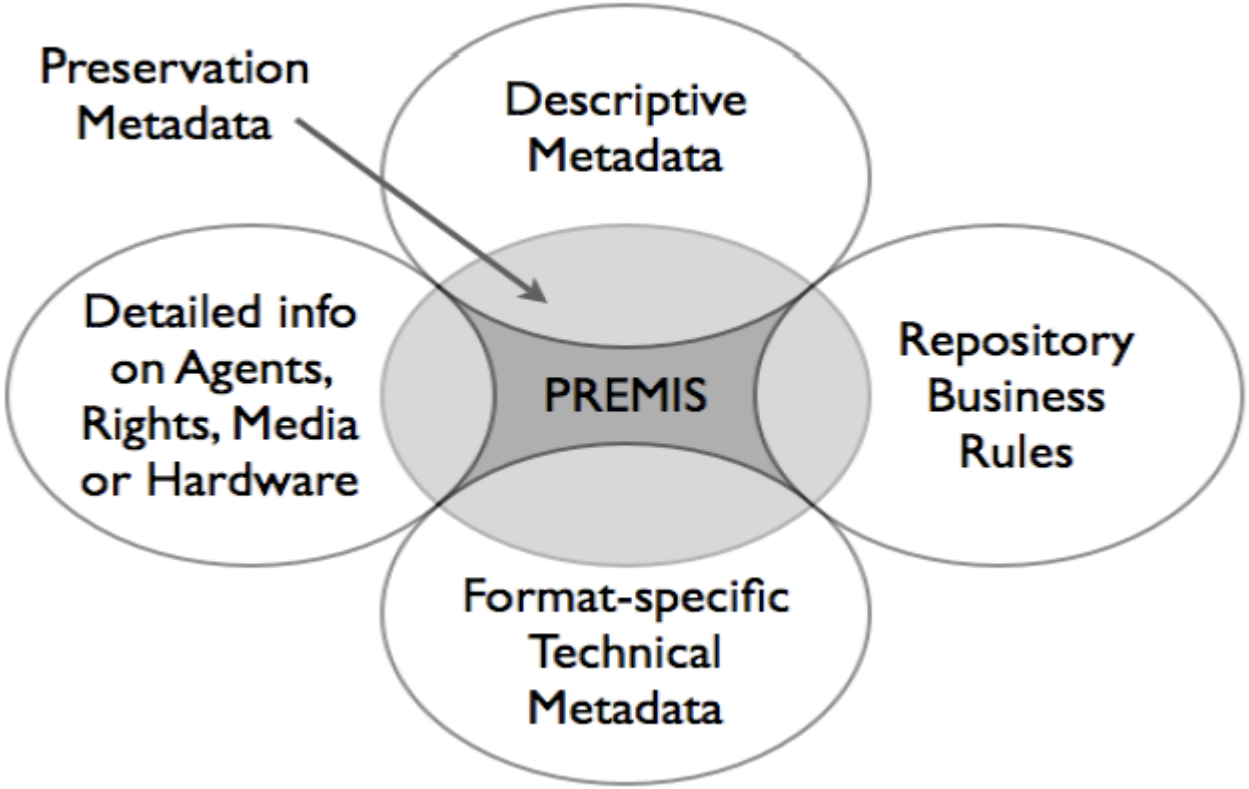


Key Terminology

- Management / Preservation Target:
  Digital Object, Intellectual Entity, Representation, File, Bitstream
- Contextual Information:
  Semantic Units, Significant Property, Rights Statement
- Environment/Actors:
  Environment, Designated Community, Agent
- Processes / Function:
  Event

KEY OBSERVATIONS

- De-facto standard for preservation metadata
- Captures information required to manage objects over the long-term
- Captures audit trails within a repository / archive
- Data Dictionary may be perceived as complicated



WHAT DID WE LEARN?

Image by Johannes Plenio
https://unsplash.com/photos/voQ97kezCx0

DIGITAL OBJECTS

- DP models have a more granular understanding of a Digital Object than RDM models, which seem rather process focused
- Misalignment in “data formats” (DMP) / “standard formats” (FAIR) to “logical object”
  → “this format is sustainable”
  vs.
  “these are the properties we want to save” (significant properties)

Image: Erwin Verbruggen https://www.flickr.com/photos/ipres2018/31543864697/ CC BY-SA 2.0

METADATA

- Both domains aware of importance of metadata
- DP models include metadata about changes processes to data / audit trails
- Significant properties as key element currently missing in RDM

significant property = feature of a digital object which needs to be preserved across object transformations

ACTORS & PROCESSES

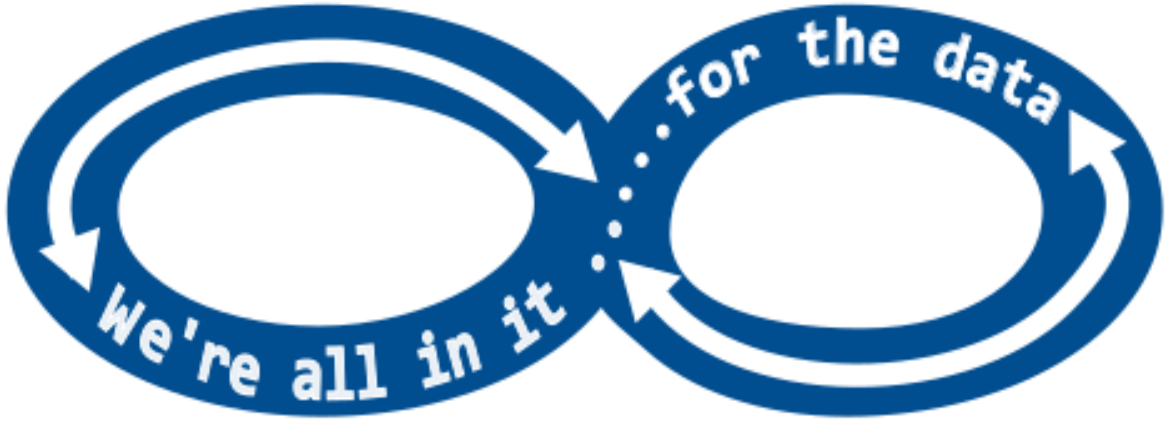
- RDM more producer, DP more consumer oriented (looking back vs. looking forward)
- RDM has more granular understanding of context the object was created in
- Currently lacking in DP is consideration of ethics
  → cultural framework in which digital artefacts are produced, stored and consumed needs to be considered in DP as well!

Image: Erwin Verbruggen https://www.flickr.com/photos/ipres2018/45570593535/ CC BY-SA 2.0

RECOMMENDATIONS

- Applying the DMP concept in digital preservation to encourage early engagement with content creators and consider what needs to be preserved
- Apply the Designated Community concept in RDM to understand who is most likely to use the data and how, using this information to inform choices of file formats, standards and preservation approaches.
- Apply digital preservation concepts to FAIR to ensure Digital Objects remain usable over time and are effectively preserved.

THANK YOU! QUESTIONS?



M. Lindlar – TIB Hannover
Michelle.lindlar@tib.eu
mickylindlar

Sarah Jones – DCC
Sarah.Jones@glasgow.ac.uk
sjDCC

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

TIB

DCC