Active Data Management in Space 20m DG

- The group that produced OAIS and related standards are working on one which is closely aligned with the ADMP work.
- It should lead to an ISO standard which defines terms and concepts relevant to ADMP
- The RDA-ADMP group work will be able to build on this

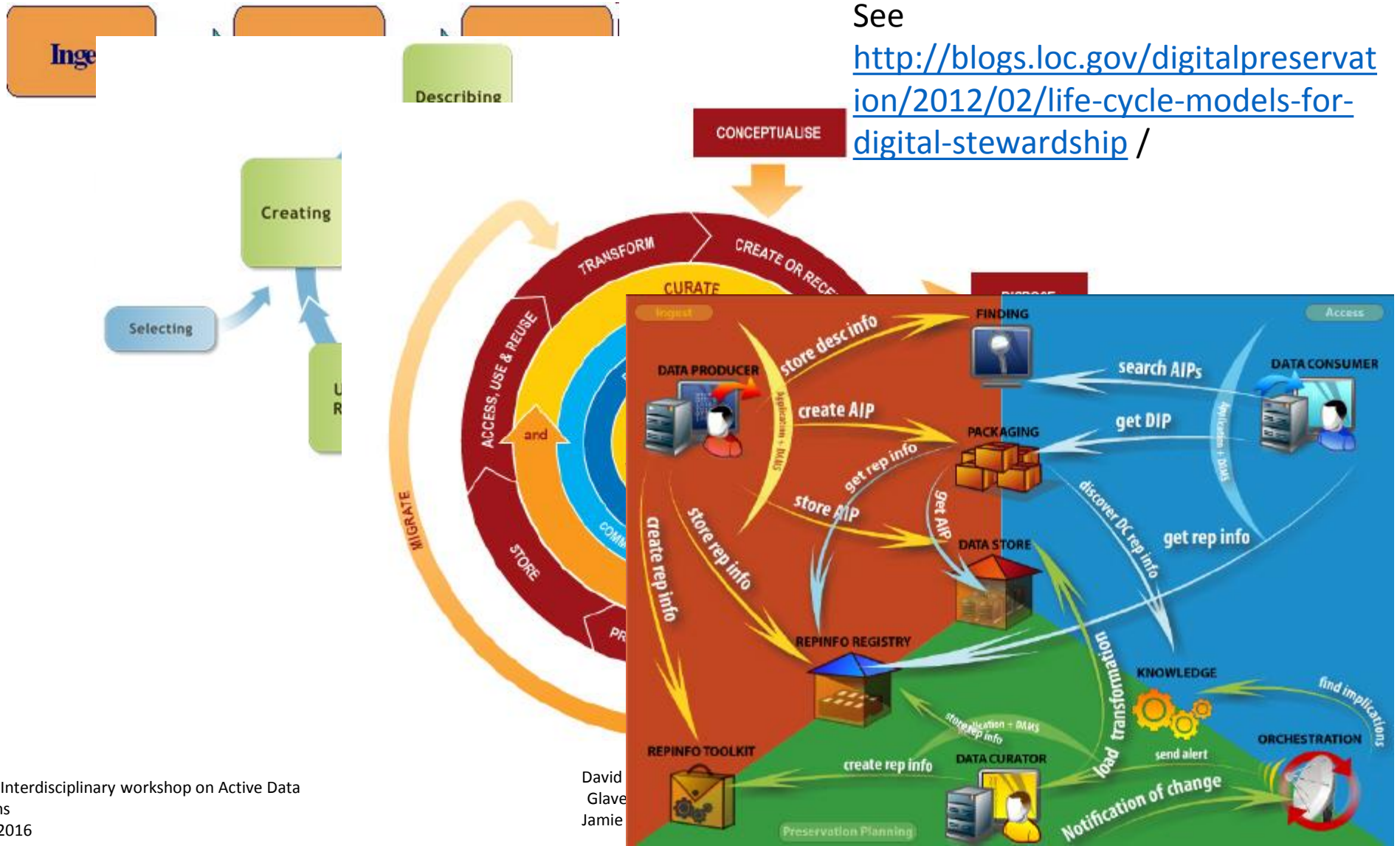
David Giaretta <david@giaretta.org>
Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>

CCSDS/ISO related work

- There is a well-recognized need to capture digital information associated with a great variety of endeavours in virtually all areas of society.
- However it is widely recognized that many such endeavours are not able, for one reason or another, to leave a sufficient legacy of information so others can reuse and fully leverage the effort that has gone into the endeavour.
- The purpose … is provide guidance for projects about the Additional Information that needs to be captured and/or generated, and retained in order to ensure that the information created by the project, either as part of its main objectives, or as a by-product of achieving those objectives, can be exploited over the long term.
- …. deals with the aspects of a project, in particular the terminology used.
- Many of these terms are already used with various definitions within the target communities for this standard – e.g. space, science, records management and archival communities. It is expected that other communities can easily map this terminology to the terminology used within those communities.
- The OAIS Reference Model provided a starting point and inputs from a variety of other sources were used to arrive at the terms used within this standard.
- Aim to do the following:
  - identifies the data and Additional Information to be collected or improved at various points;
  - forms a basis for the specification of Data Management Plans
  - forms a basis for the identification and/or development of additional standards and implementation guides including those that address particular concerns in more detail;
  - forms a basis for identification and/or development of a set of software tools that will assist the development, operation and checking of the different parts of the project.

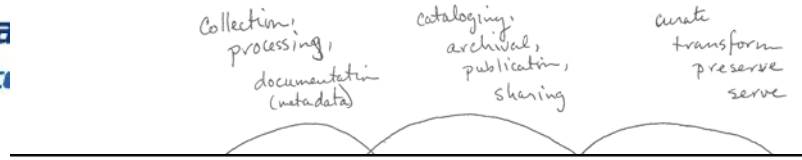
David Giaretta <david@giaretta.org>
Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>

Many models – why another?

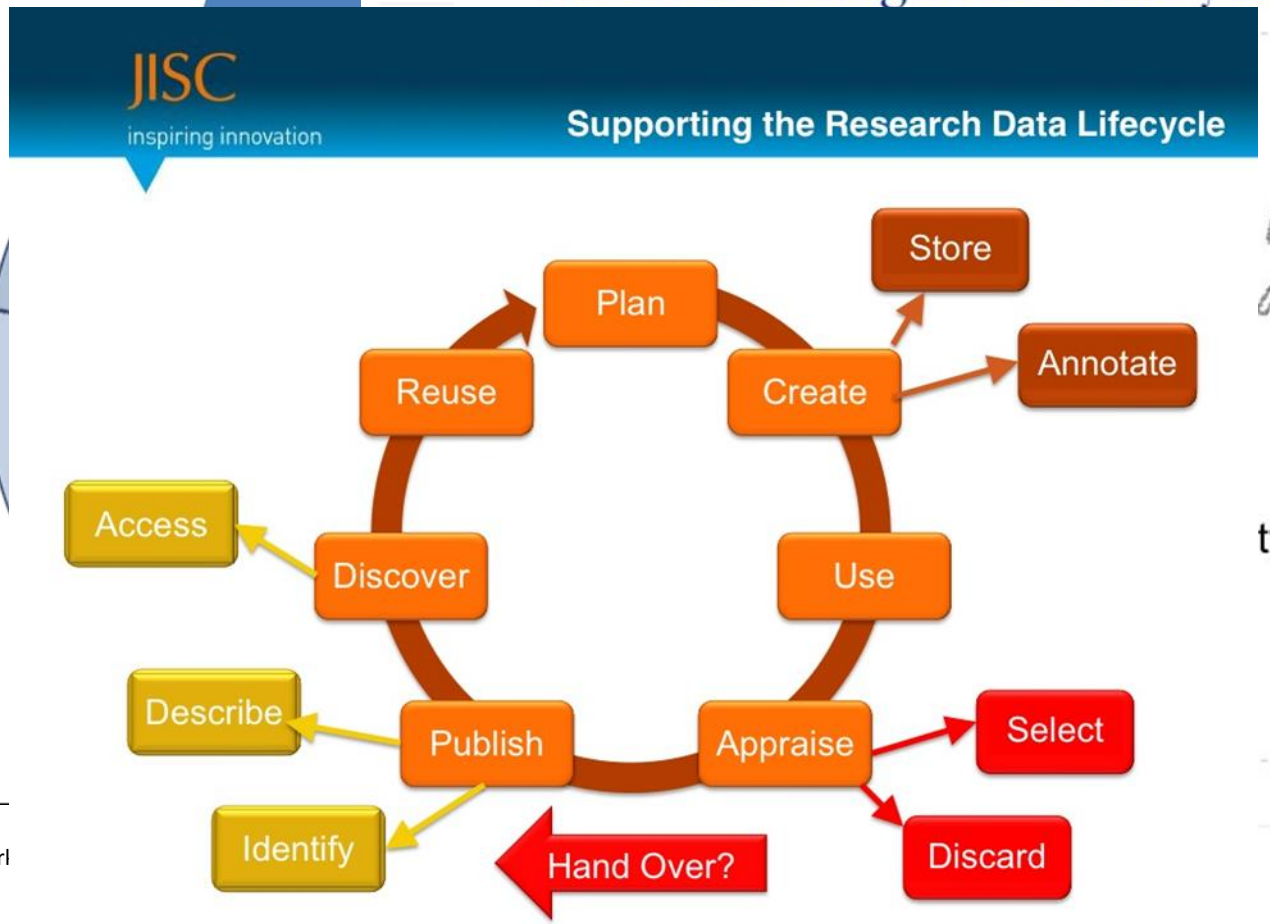


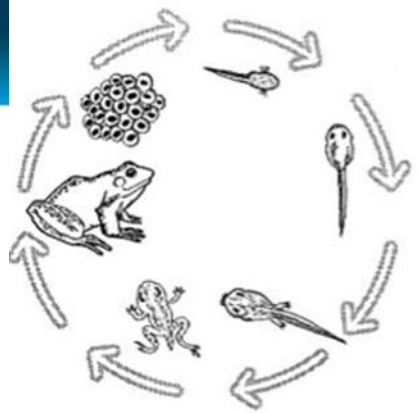
See http://blogs.loc.gov/digitalpreservation/2012/02/life-cycle-models-for-digital-stewardship /

Data Lifecycle Models and Concepts by CEOS, 2012, see http://www.ceos.org/images/DSIG/Data%20Lifecycle%20Models%20and%20Concepts%20v13.docx



The Geoarchiving Process Lifecycle

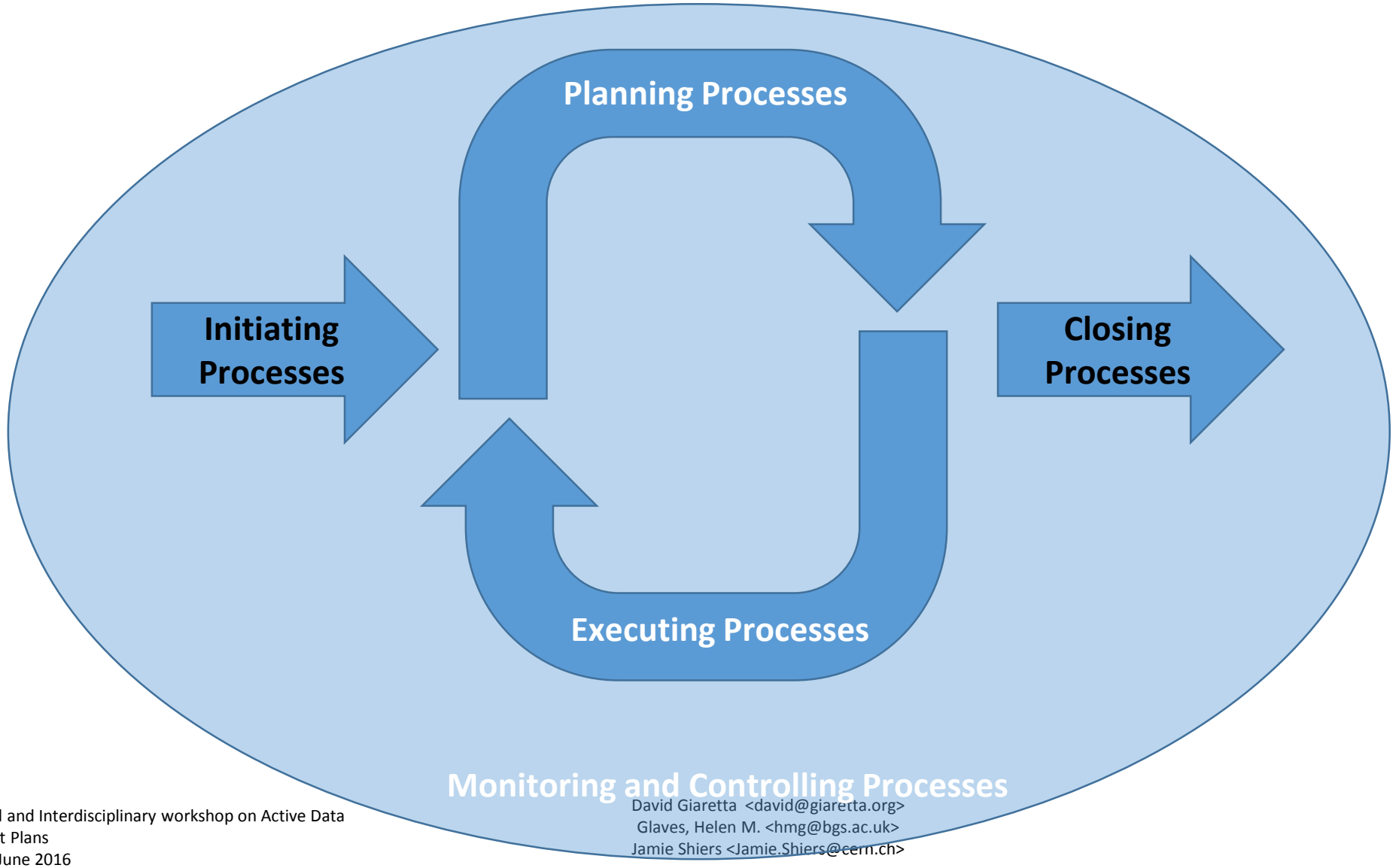




http://www.ipcc.ie/lifecycle.html

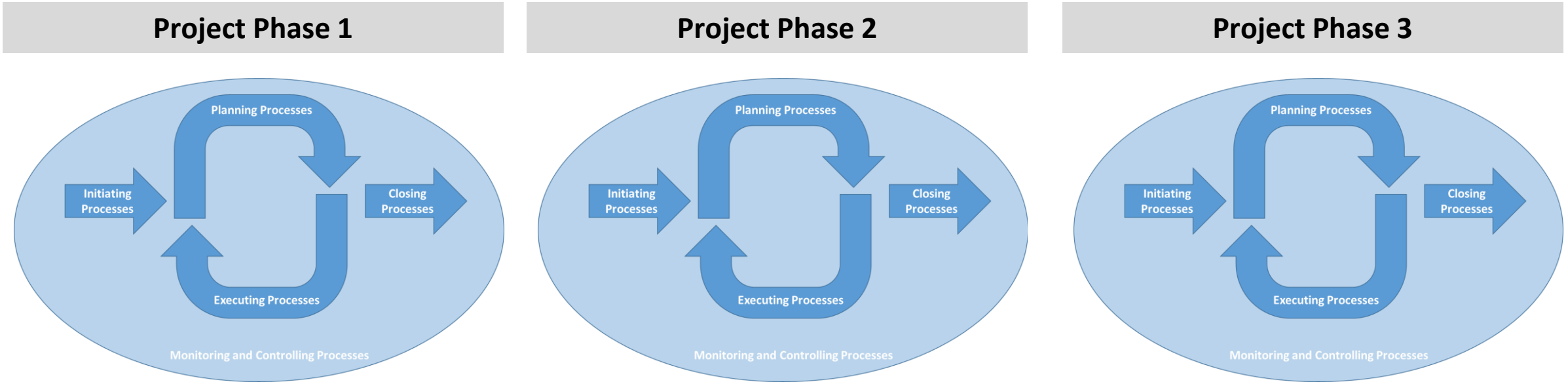
ty

Project Management Body of Knowledge (PMBOK) concepts



David Giaretta <david@giaretta.org>
Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>

Giaretta Associates

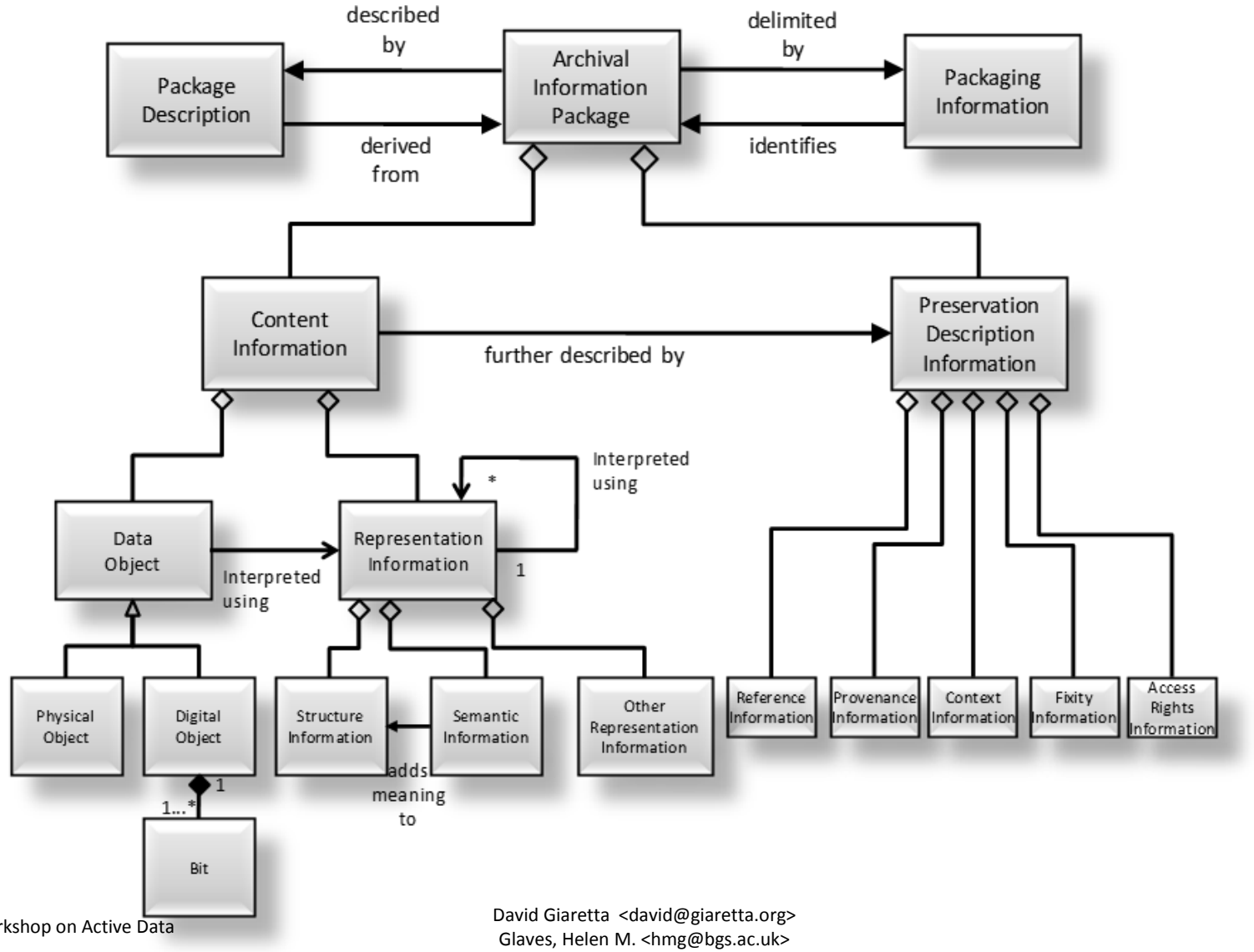


Example of project with three phases

David Giaretta <david@giaretta.org>
Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>

OAIS Archival Information Package

Logical container for everything needed for Long Term Preservation



David Giaretta <david@giaretta.org>
Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>

Additional Info. to collect/update at each type of process – some examples

Additional Information Topic	Detailed area	Initiating	Planning	Executing	Closing
Content Data	Inventory of data produced/ expected	Rough idea	Increasingly detailed	Becoming complete	Complete
	Types of data (raw, processed, etc.) which should be preserved?	Rough idea	Increasingly detailed	Becoming complete	Complete
	Type of data e.g. images, tables – which generic interfaces?	Rough idea	Increasingly detailed	Becoming complete	Complete
	Volume that would require preservation	Rough idea	Increasingly detailed	Becoming complete	Complete
	Quality constraints	Rough idea	Increasingly detailed	Becoming complete	Complete
	Quality checks which may be performed on the data by non-experts	Rough idea	Increasingly detailed	Increasingly detailed	Complete

David Giaretta <david@giaretta.org>
Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>

Additional Information Topic	Detailed area	Initiating	Planning	Executing	Closing
Representation Information	Choice of data format	Rough idea	Increasingly detailed	Becoming complete	Complete
	Format definitions and formal descriptions	Rough idea	Increasingly detailed	Becoming complete	Up to date and accumulating
	Semantics of the data elements	Rough idea	Increasingly detailed	Becoming complete	Almost complete
	Data dictionaries and other semantics	Rough idea	Increasingly detailed	Becoming complete	Up to date and accumulating
	Information Model	Rough idea	Increasingly detailed	Becoming complete	Complete
	Other Data Documentation	Rough idea	Increasingly detailed	Becoming complete	Up to date and accumulating
	Applicable standards	Rough idea	Increasingly detailed	Becoming complete	Complete
	Hardware and Software Dependencies	Rough idea	Increasingly detailed	Becoming complete	Up to date and accumulating
	Other software which may be used on the data		Increasingly detailed	Increasingly detailed	Growing
	Calibration and system test tools and system test data that will be delivered.	Rough idea	Increasingly detailed	Becoming complete	Up to date and accumulating
	Relationships between data items	Rough idea	Increasingly detailed	Complete	Complete

Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>

Additional Information Topic	Detailed area	Initiating	Planning	Executing	Closing
Reference Information	DOI or other unique identifiers	Rough idea	Becoming complete	Up to date and accumulating	Up to date and accumulating; New methods could be introduced
	Rules, methods, tools for referencing data	Rough idea	Becoming complete	Up to date and accumulating	Up to date and accumulating; New methods could be introduced
	What standards will be used to ~~format,~~ identify and reference the data and metadata	Rough idea	Becoming complete	Up to date and accumulating	Up to date and accumulating; New methods could be introduced
	What may be used in future to identify the Information	Fairly firm	Increasingly detailed	Increasingly detailed	Evolving

David Giaretta <david@giaretta.org>
Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>

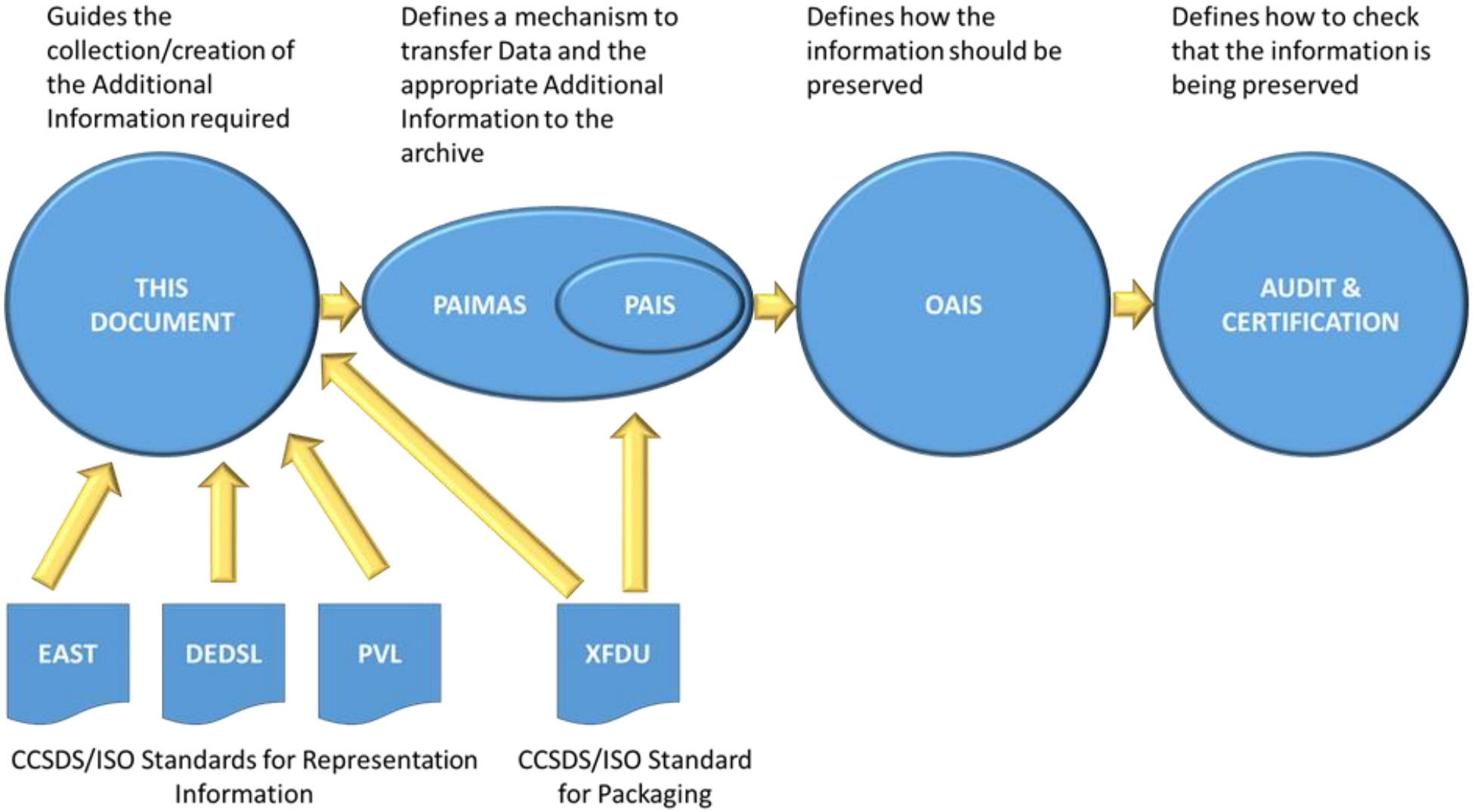
Additional Information Topic	Detailed area	Initiating	Planning	Executing	Closing
Provenance Information	Record of origins of the project e.g. in a CRIS system	Fairly firm	Complete	Completed	Complete
	Documentation about the hardware and software used to create the data, including a history of the changes in these over time		Rough Idea then Increasingly detailed	Becoming complete	Up to date and accumulating
	Processing workflow	Rough idea	Increasingly detailed	Becoming complete	Complete
	Processing inputs		Rough Idea then Increasingly detailed	Becoming complete	Complete
	Processing parameters	Rough idea	Increasingly detailed	Becoming complete	Complete
	Who was responsible for each stage of processing		Increasingly detailed	Becoming complete	Complete
	When each stage was performed		Increasingly detailed	Becoming complete	Complete
	Record of any special hardware needed	Rough idea	Increasingly detailed	Becoming complete	Complete
	Calibration	Rough idea	Becoming complete	Complete	Complete
	System Testing	Rough idea	Becoming complete	Up to date and accumulating	Up to date and accumulating; New methods could be introduced
	Resident Archives			Rough idea	Becoming complete
	Who was responsible for each stage of processing (Fixity)		Up to date and accumulating	Up to date and accumulating	Up to date and accumulating

Jamie Shiers <Jamie.Shiers@cern.ch>

Additional Information Topic	Detailed area	Initiating	Planning	Executing	Closing
Issues Outside the Information Model	Schedule of deliveries	Fairly firm	Increasingly detailed	Complete	
	Cost	Fairly firm	Increasingly detailed	Complete, but may Evolve	Complete, but may Evolve
	Pointers to the components to be transferred to the archive		Fairly firm	Complete	Complete, but may Evolve
	Potential preservation aims of the archive	Rough idea	Increasingly detailed	Increasingly detailed	Complete
	Potential risks to preservation and exploitation of the data	Fairly firm	Increasingly detailed	Complete, but may Evolve	Complete, but may Evolve
	The target archives and designated community for the solicitation.	Fairly firm	Complete	Complete, but may Evolve	Complete, but may Evolve
	The budget for archiving.	Fairly firm	Complete	Complete, but may Evolve	Complete, but may Evolve
	The schedule for major project milestones and deliveries to the archive.	Fairly firm	Complete	Complete, but may Evolve	Complete, but may Evolve
	Change Management		Complete	Complete, but may Evolve	Complete, but may Evolve
	The mechanism for communication between project and archive.	Fairly firm	Complete	Complete, but may Evolve	Complete, but may Evolve

Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>

Relationship between standards



David Giaretta <david@giaretta.org>
Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>

Relationship to RDA ADMP

- CCSDS document
  - would define terminology and basic concepts which are consistent with OAIS etc
  - would become an ISO standard
- RDA ADMP
  - can contribute to the review of the document and example cases
  - could/should use ISO terms and concepts to
    - define exchange formats and semantics for plans – ideally machine processable
    - specify/collect/implement tools to create the Additional Information
    - reach out to funders, data creators and archives

David Giaretta <david@giaretta.org>
Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>

DISCUSSION

David Giaretta <david@giaretta.org>
Glaves, Helen M. <hmg@bgs.ac.uk>
Jamie Shiers <Jamie.Shiers@cern.ch>