



# Data Preservation Policies: from creation to exploitation

David Giaretta

[david@giaretta.org](mailto:david@giaretta.org)

[www.giaretta.org](http://www.giaretta.org), [www.iso16363.org](http://www.iso16363.org) and [www.digitalpreservation.guru](http://www.digitalpreservation.guru)

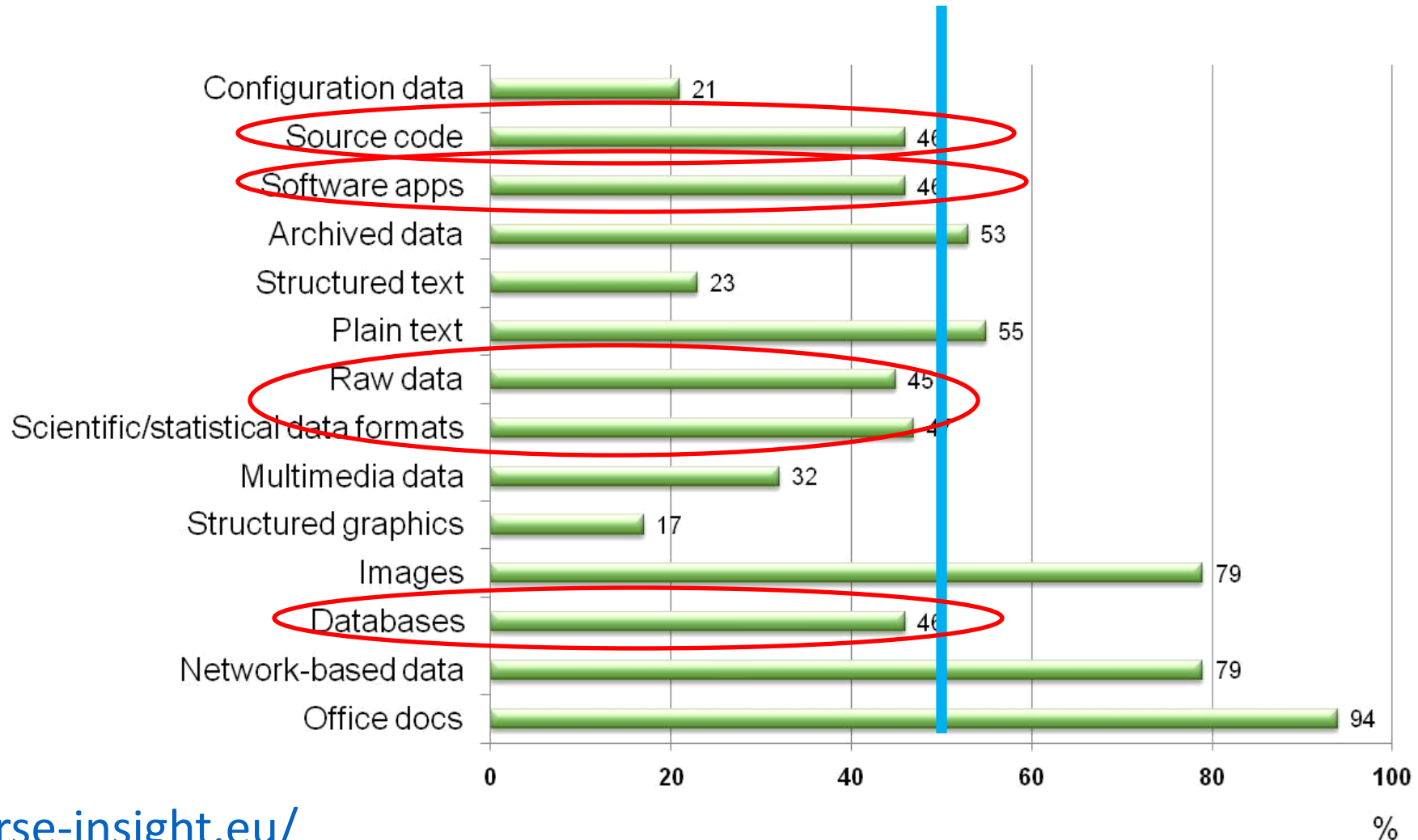
**Our Digital Future - Multidisciplinary Perspectives on Long Term Data  
Preservation and Access**

**Cambridge 14/15 March 2016**

# Outline

- Requirements for policies
  - What may need to be preserved?
  - What can we afford (not) to preserve?
  - How to decide/how to justify?
- Aims for policies
  - What is needed?
  - What/How/When to capture?
- Policies to help
- Conclusions

# What may need to be preserved?



# Threats to preservation

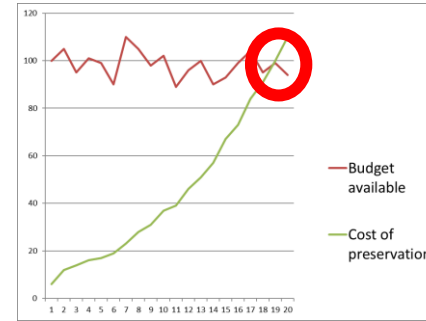
1. Users may be unable to understand or use the data e.g. the semantics, format or algorithms involved.
2. Lack of sustainable hardware, software or support of computer environment may make the information inaccessible.
3. Evidence may be lost because the origin and authenticity of the data may be uncertain.
4. Access and use restrictions (e.g. Digital Rights Management) may not be respected in the future.
5. Loss of ability to identify the location of data.
6. The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future.
7. The ones we trust to look after the digital holdings may let us down.

Plus disasters and funding issues etc

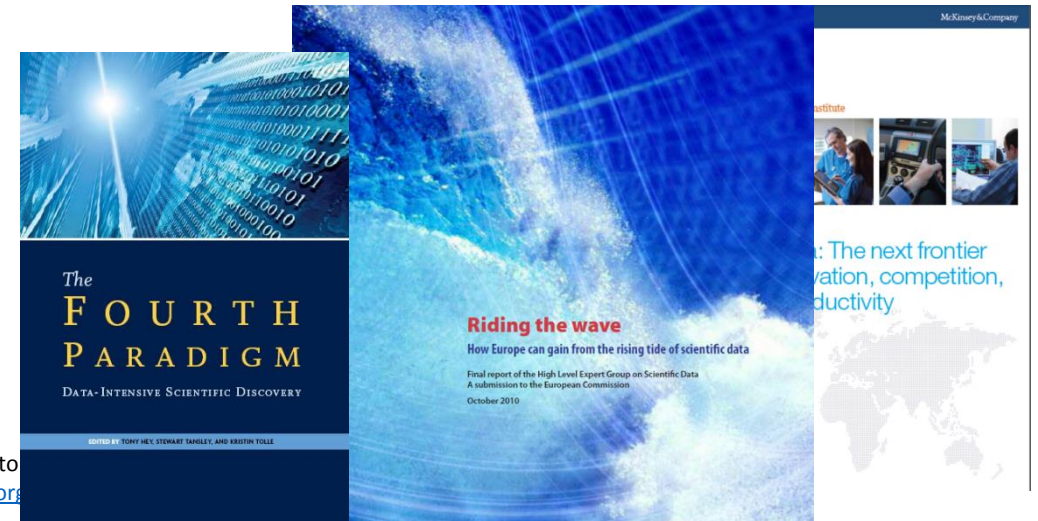
PARSE.Insight

<http://www.parse-insight.eu/>

# What can we afford (not) to preserve? Value



- EU Commissioner for the Digital Agenda said: “Data is the new Gold” but
- Gold is precious because it is rare, and does not combine
- Data is precious because there is so much and it becomes more valuable when it is combined
- McKinsey report on value of data
- EU Riding the Wave report
- New scientific paradigms



# Types of value

- Science
- Social
- Historical
- Ethical
- Aesthetic
- Scholarly
- Institutional
- Public policy
- Industry
- Governmental
  - e.g. US Executive Order 13287 ["Preserve America"](#)
  - Managing government records [directive M-12-18](#)
- Legal requirements
  - Industry, healthcare, finance .....

# Preservation requires resources – how are they justified?

- Some objects have immediate payback (at least for now) or immediate penalties
- What about potential future/value (Blue Ribbon Task Force)?
- BUT...
- We can try to help make the objects usable
- ***Data is precious because there is so much and it becomes more valuable when it is combined***



<https://www.flickr.com/photos/55231259@N00/5288486661/>



# Semantics matter

## Metric mishap caused loss of NASA orbiter

September 30, 1999

Web posted at: 4:21 p.m. EDT (2021 GMT)

---

### In this story:

[Metric system used by NASA for many years](#)

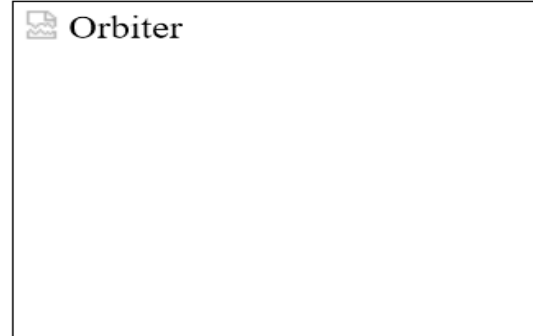
[Error points to nation's conversion lag](#)

[RELATED STORIES, SITES](#) ↓

---

By Robin Lloyd  
CNN Interactive Senior Writer

(CNN) -- NASA lost a \$125 million Mars orbiter because a Lockheed Martin engineering team used English units of measurement while the agency's team used the more conventional metric system for a key spacecraft operation, according to a review finding released Thursday.



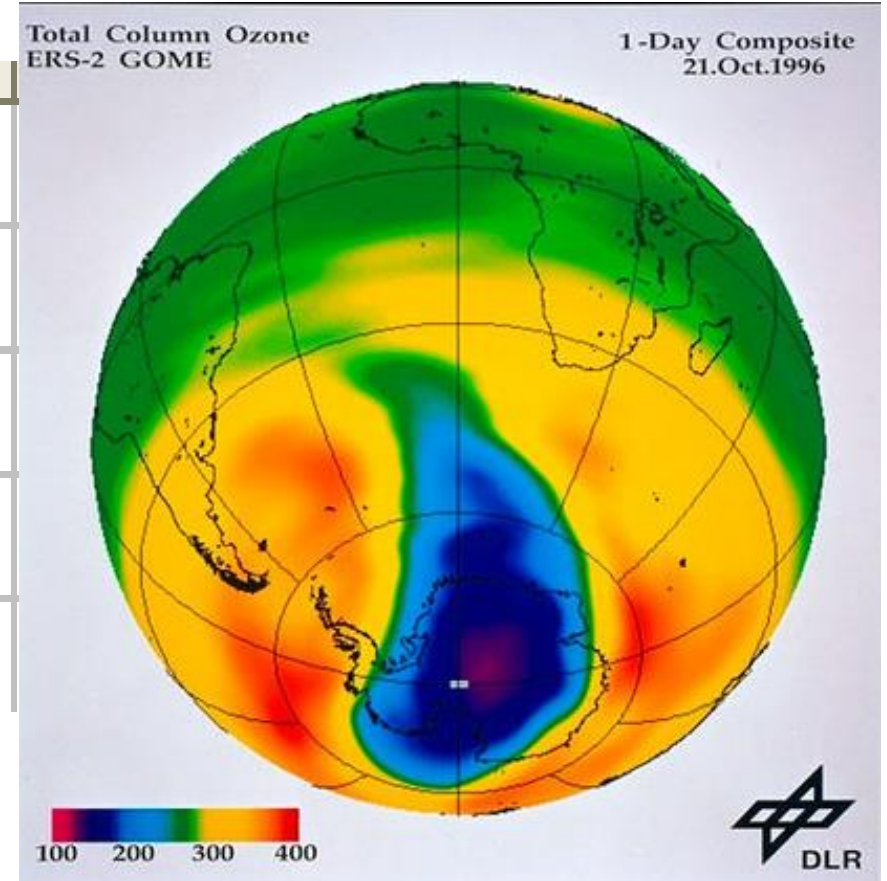
NASA's Climate Orbiter was lost September 23, 1999

<http://edition.cnn.com/TECH/space/9909/30/mars.metric.02/>



# Data contains numbers etc – need meaning

latitude	longitude	Ozone
132	50	34.9
178	50	45
190	50	78

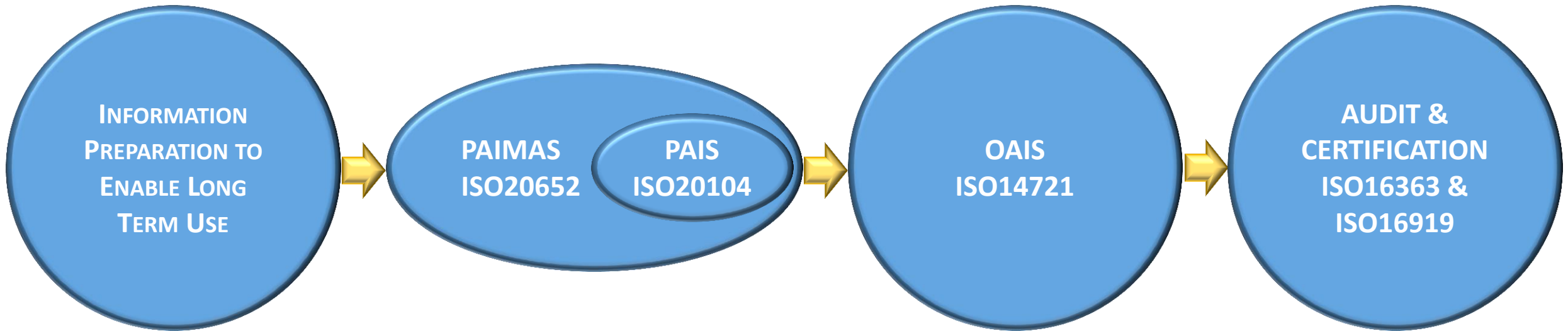


FORMATS can be tricky and can change

BUT

SEMANTIC are even trickier

# What is needed?



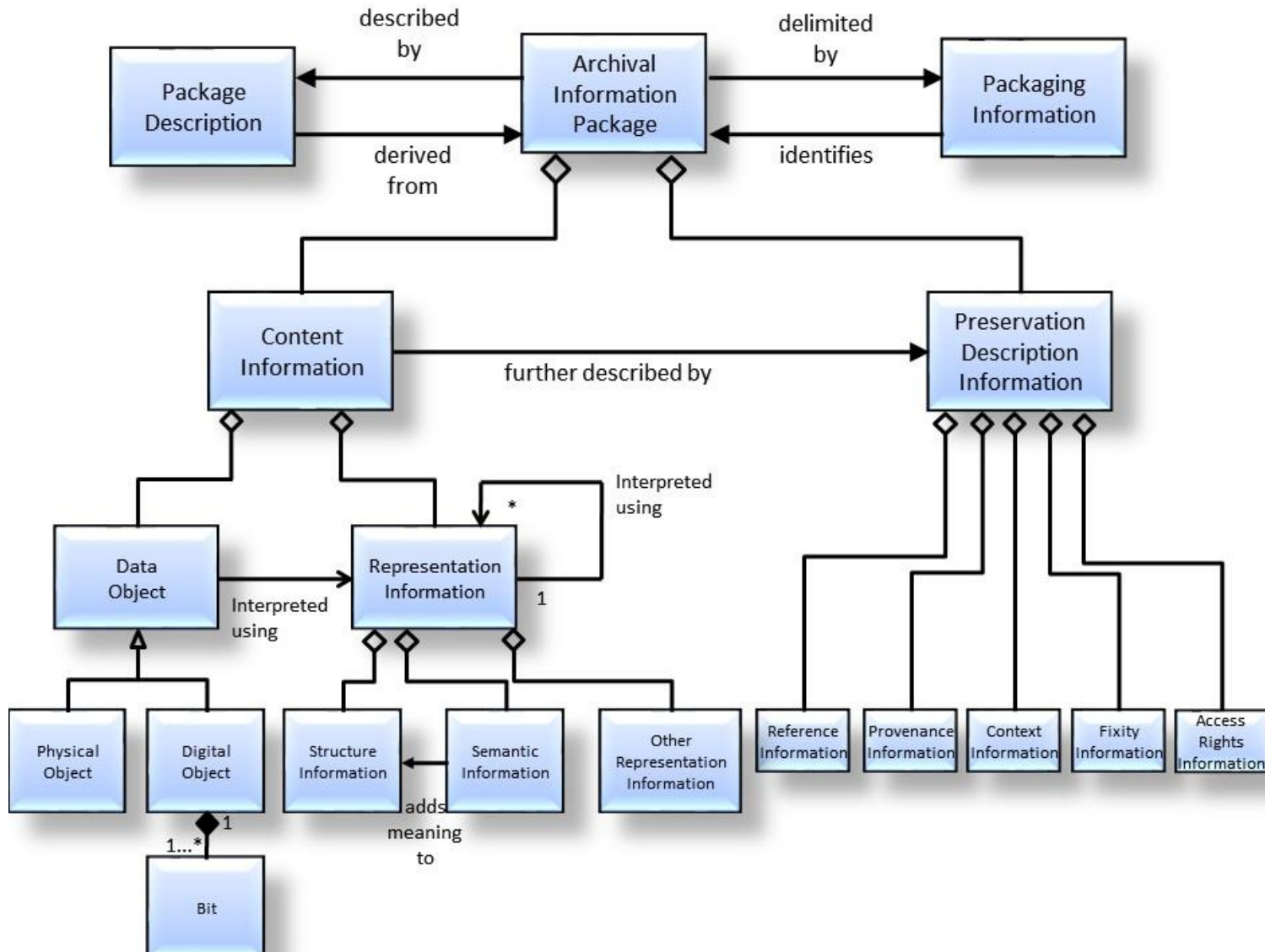
Helps to ensure that the Additional Information needed is collected/ created

Defines a mechanism to transfer the Primary Information and the Additional Information to the archive

Defines how the information should be preserved

Defines how to check that the information is being preserved

# OAIS Archival Information Package



- Defines the information needed to preserve the data object
- Much of the Additional Information can only be collected during the creation of the data

Collection, processing, documentation (metadata)      Cataloging, archival, publication, sharing      Curate, transform, preserve, serve

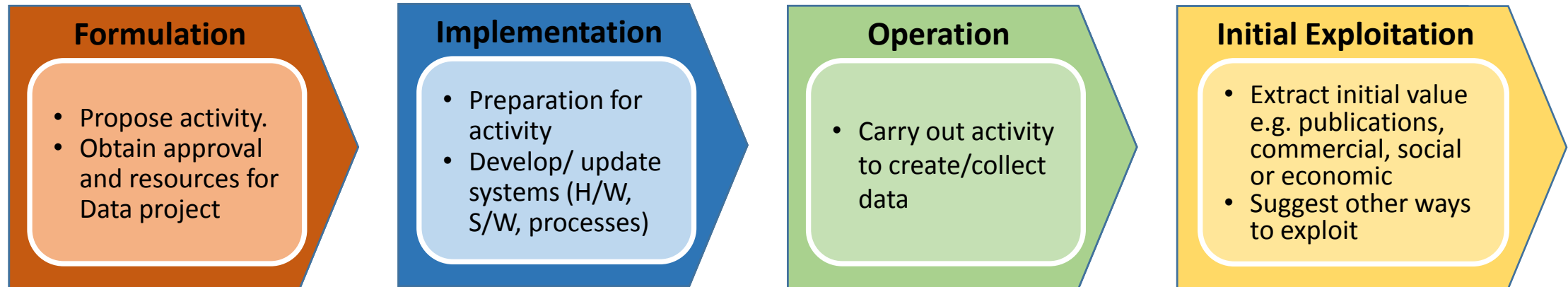
Data Lifecycle Models and Concepts by CEOS, 2012, see <http://www.ceos.org/images/DSIG/Data%20Lifecycle%20Models%20and%20Concepts%2>



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

ty

# What/How/When to capture?

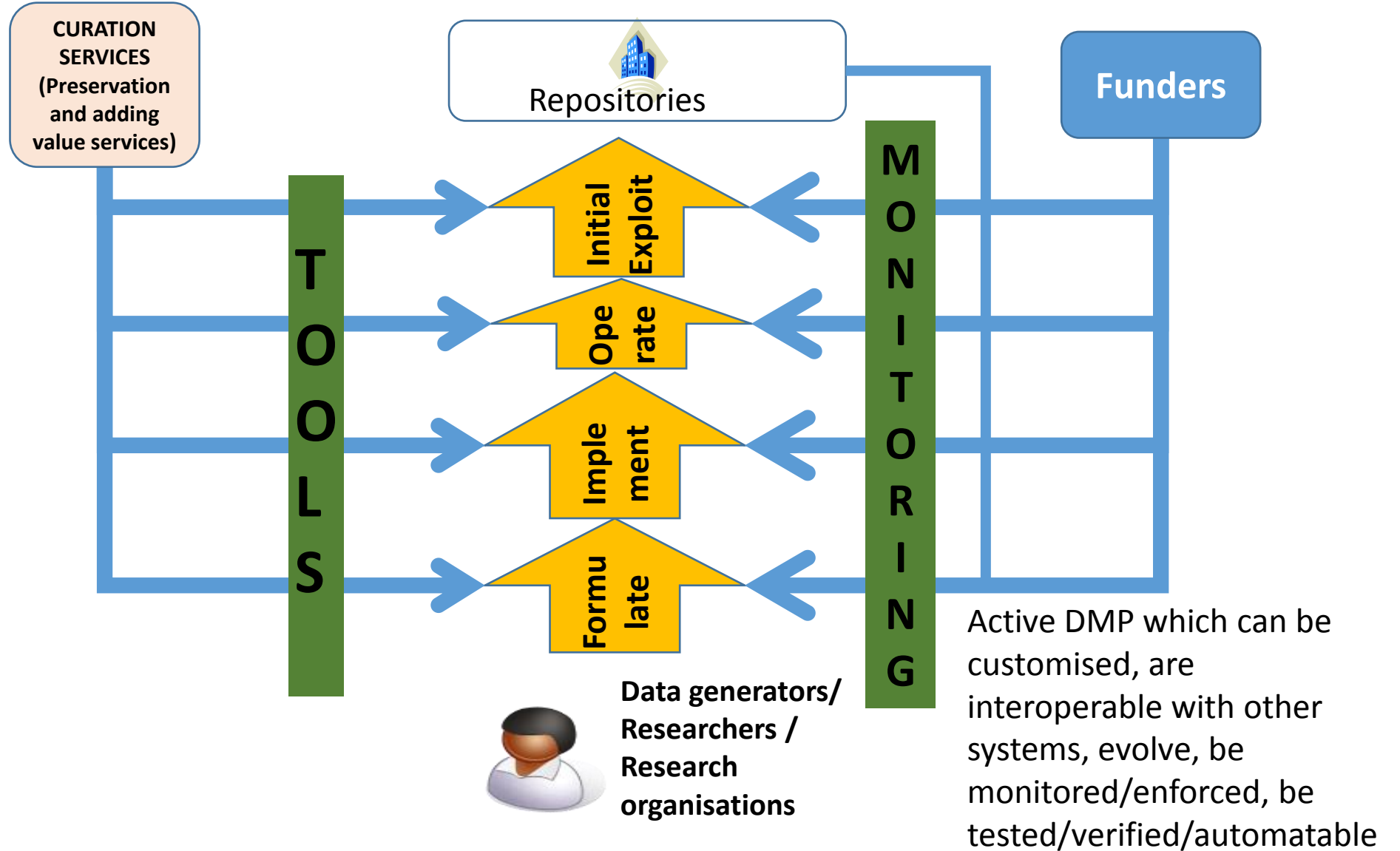


[http://cwe.ccsds.org/moims/default.aspx#\\_MOIMS-DAI](http://cwe.ccsds.org/moims/default.aspx#_MOIMS-DAI)

# Policies to help

- Current Data Management Plans must be improved
- ACTIVE Data Management Plans should be required
  - RDA <https://www.rd-alliance.org/groups/active-data-management-plans.html>
- Must monitor the creation of the various types of Additional Information (“metadata”) needed
- Ensure data is eventually placed in an ISO certified archive

Topic	Issue
Representation Information	Choice of data format
	Format definitions and formal descriptions
	Semantics of the data elements
	Data dictionaries and other semantics
	Information Model
	Other Data Documentation
	Applicable standards
	Hardware and Software Dependencies
	Other software which may be used on the data
	What calibration and system test tools and system test data will be delivered.
Relationships between data items	



# Conclusions

- Data is/are different from documents
  - Semantics (e.g. meaning of the numbers) cannot be ignored
  - “format” and “emulate or migrate” not enough
- People who create the data know a lot about it
  - We need to ensure they capture the Additional Information
  - Details will be better defined over the period it is created
- Will improve the chances that the data can be re-used/combined and increased in value
  - Consistent with Preservation best practice and ISO certification
- Helps to justify preservation and aids in appraisal







END

[david@giaretta.org](mailto:david@giaretta.org)

[www.giaretta.org](http://www.giaretta.org) / [www.iso16363.org](http://www.iso16363.org) / [www.digitalpreservation.guru](http://www.digitalpreservation.guru)

[http://cwe.ccsds.org/moims/default.aspx#\\_MOIMS-DAI](http://cwe.ccsds.org/moims/default.aspx#_MOIMS-DAI)

<https://www.rd-alliance.org/groups/active-data-management-plans.html>