



RESEARCH DATA ALLIANCE

Practical Policies Results & Impact

Rainer Stotzka, Reagan Moore

Research Infrastructure meet Research Data Alliance
Workshop, Amsterdam, 26.-27.5.2015

research data sharing without barriers
rd-alliance.org

Scenario

- Create research data repository
- Data
- Some assertions: policies & rules attached to the data
 - Contextual metadata extraction
 - Access rules
 - Regular integrity checks
 - etc.

Policy:

Assertion or assurance that is enforced about a collection or a dataset

Problem

Computer actionable policies

- Enforce management
- Automate administrative tasks
- Validate assessment criteria
- Automate scientific analysis
- etc.

A generic set of policies that can be revised and adapted by user communities and site managers does not exist.

- *Domain scientists* who want to build-up a collection or a repository
- *Data centers* for automating policies

Goals

- To bring together practitioners in policy making and policy implementation
- To identify *typical application scenarios* for policies such as replication, preservation etc.
- To *collect* and to register practical policies
- To enable *sharing*, revising, adapting, and re-using of computer actionable policies



Survey of 30 Institutions for Highest Priority Policies

Policy	Importance
Integrity	217
Preservation	150
Access control	126
Provenance	108
Data Management plans	99
Publication	75
Replication	66
Data staging	52
Federation	37
Metadata sharing	23
Regulatory	16
Collection properties	7
Identifiers	7
Data sharing	7
Versioning	7
Licensing	6
Format	6
Data Life Cycle	6
Arrangement	5
Processing	5

In close cooperation with the Engagement Group



Outcomes Policy Templates: Practical Policy Working Group, September 2014

Version: August 24, 2014

<https://www.rd-alliance.org/filedepot?cid=104&fid=556>

Templates

- Interactions of policies and DO attributes
- Policy descriptions
- Technology independent
- Reviews of the provided policy areas in progress

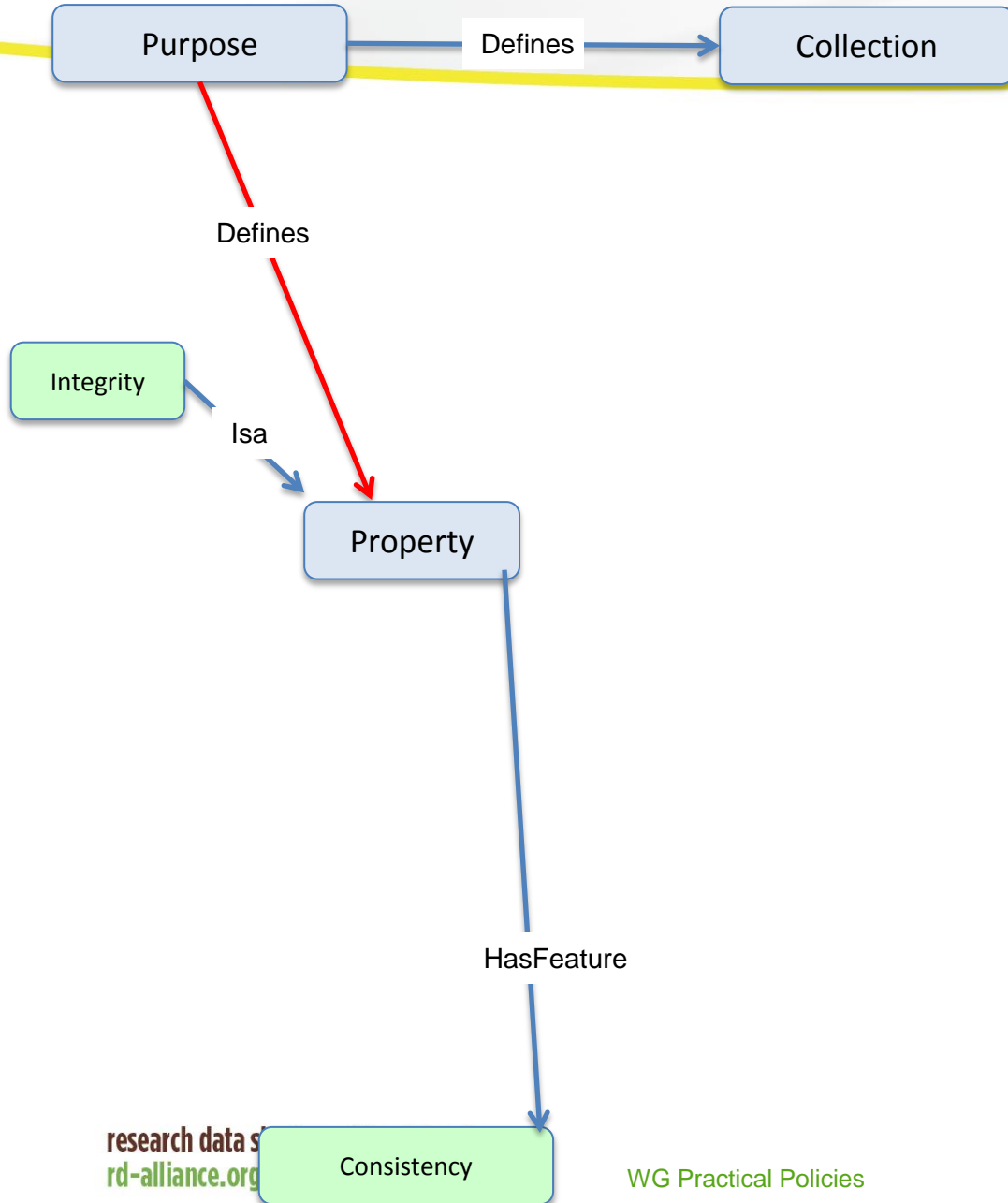
Implementations: Practical Policy Working Group, September 2014

Version: August 24, 2014

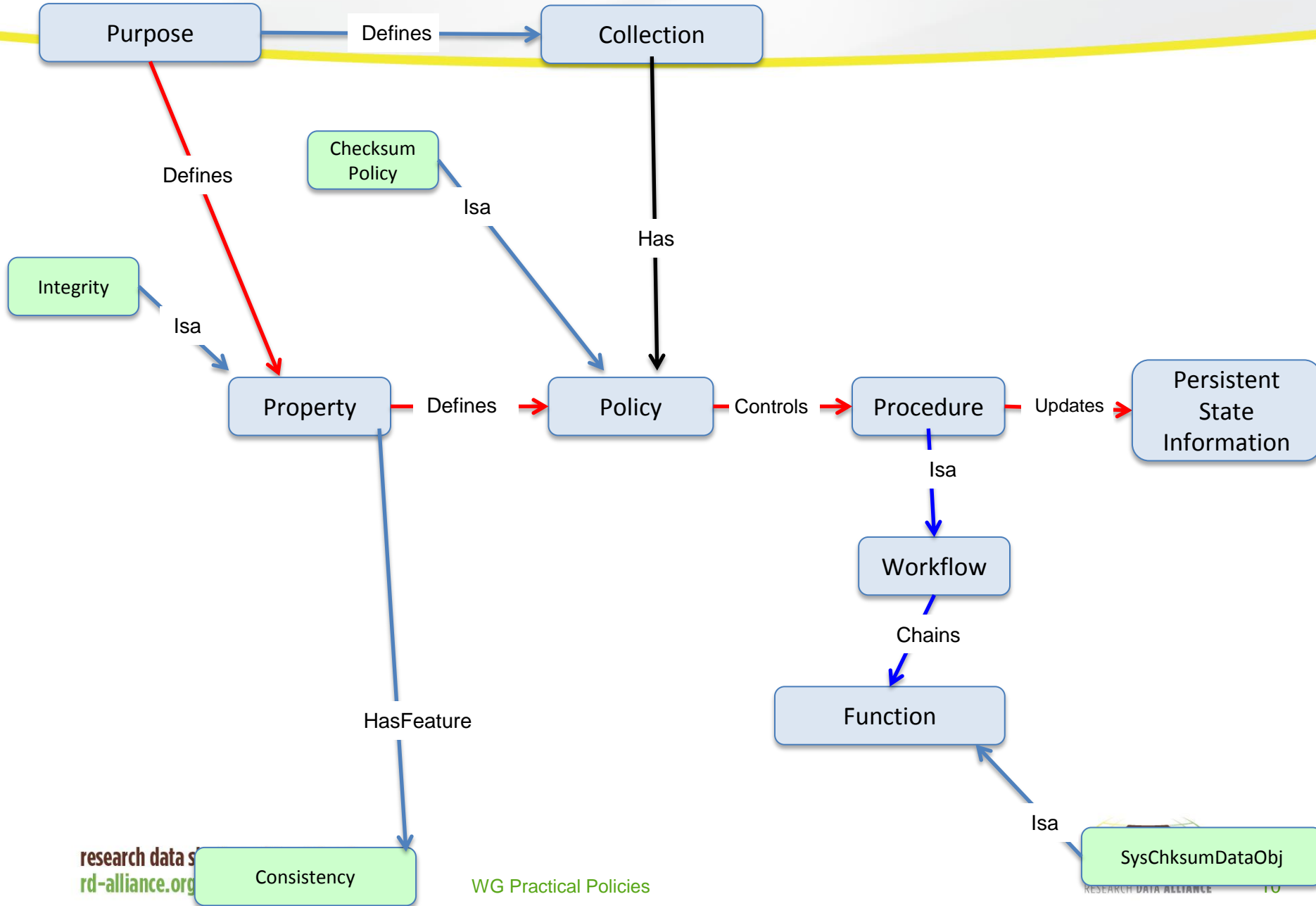
<https://www.rd-alliance.org/filedepot?cid=104&fid=553>

- Examples for implementations:
 - English language descriptions
 - iRODS
 - GPFS
- ~50 pages

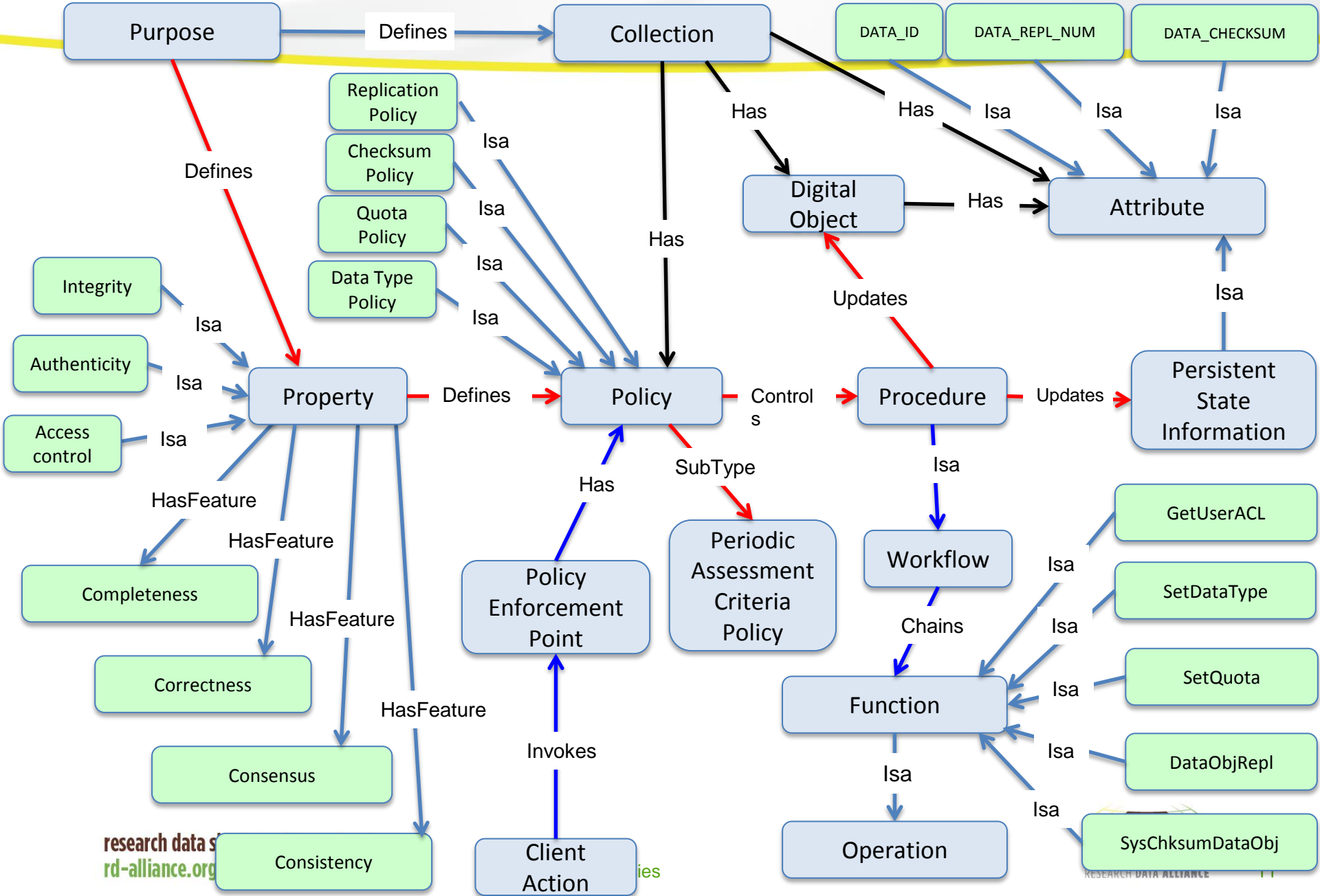
Concept Graph by Reagan Moore



Concept Graph by Reagan Moore



Concept Graph by Reagan Moore



Example: Contextual metadata extraction and management policies

Scenarios:

- Extract metadata from an associated document, e.g. DICOM.
- Extract metadata from a structured document, e.g. FITS for astronomy, netCDF, and HDF
- Extract metadata by parsing patterns within the text within a document.
- Identify a feature present within a file and label the file with the location of the feature that is present within the file.

Contextual metadata for provenance for description for structural for representation for administration for event	Constraint	State attributes for Constraint
	On file	File_name
	On collection	Collection_name
	On user	User_ID
	On storage	Storage_name
	Operations	State Attributes for Operation
	Extract metadata	Attribute_name
		Attribute_value
		Attribute_unit
		Source_file
		Source_collection
	Register metadata	Attribute_name
		Attribute_value
		Attribute_unit
		Destination_file
		Destination_collection
		Metadata_creation_time
	Verify metadata load	Metadata_modification_time
		File_name
		Attribute_name
Attribute_value		
Verify metadata names	Attribute_unit	
	Attribute_name HIVE_reserved_vocabulary	
Set ACL on metadata	File_ID	
	Metadata_ID	
	ACL_type	

Adoption

Target Communities:

- Groups managing data collections
- Data centers

First adopters are the institutions/organizations who contributed to the results, e.g. RENCI, KIT, OSC, DARIAH, RZG, etc.:

- EUDAT
- CESNET
- DataNet Federation Consortium

Impact

Result: List of of policy categories and policies

- Improved data center **administration**: quantify the management steps, automate enforcement
- By **sharing** policies, communities can interoperate and share data more effectively
- **Transparency**: basis of establishing trust
- Implemented policies: can be used as examples and be adapted to specific requirements and other data management systems

What's next?

Policy Templates: starting point

- Generic policy *description language*
- *Registry* system for practical policies
 - Registration → Data Policy Manager Architecture (EUDAT)
 - Citation
- Prefabricated practical policies modules
 - Data Policy Manager Architecture (EUDAT)
- Use policies to configure community-specific repositories

Conclusions

- “Outcomes Policy Templates: Practical Policy Working Group, September 2014”
<https://www.rd-alliance.org/filedepot?cid=104&fid=556>
- “Implementations: Practical Policy Working Group, September 2014”
<https://www.rd-alliance.org/filedepot?cid=104&fid=553>

