



The Requirement for Automation in Large Scale Data Management

31st International Conference on Massive
Storage Systems and Technology

June 3, 2015

Dave Fellingner

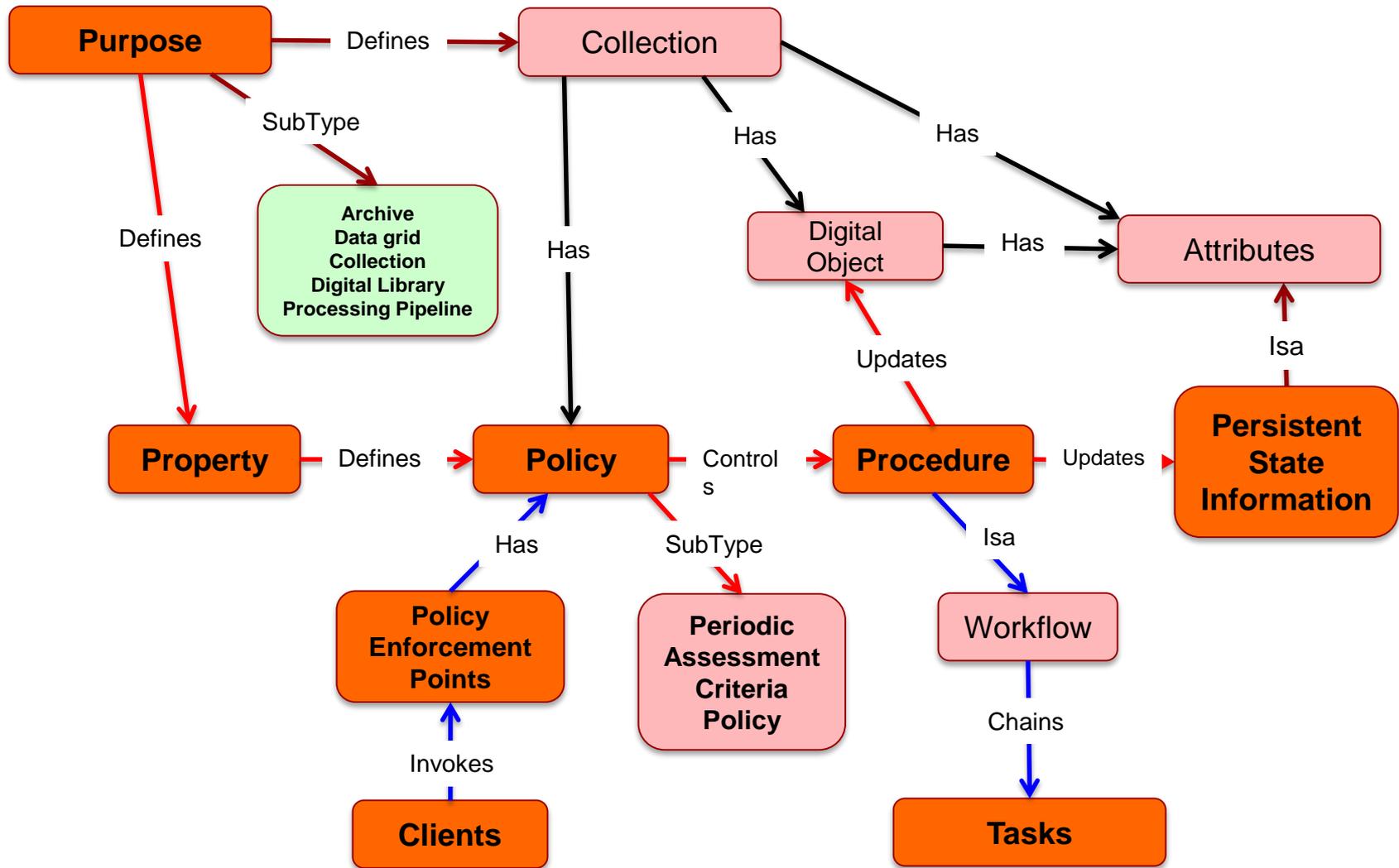
Chief Scientist, DDN dfellinger@ddn.com

2 Curation of large data sets goes beyond ACLs and retention

- ▶ All phases of data management should be rules based and automated.
- ▶ Data should be checked for format compliance upon ingestion and metadata should be extracted and cataloged.
- ▶ Access rights must be generated for both the data and catalog.
- ▶ Data retention policies must be generated based on the nature of the data.
- ▶ Distribution policies must be generated based on retention policies and a requirement for each site.
- ▶ In some cases access tracking will become a requirement since access to the data may change the retention policies.
- ▶ In all cases provenance and chain of custody must be maintained.
- ▶ HSM schemes including sparse replication and movement between tiers must be automated and based on the data type.

The Integrated Rules Oriented Data System (iRODS) is middleware that can enable this automation

Policy-based Data Management Concept Graph



Moore, R., Berg-Cross, G., Practical Policy working group, Research Data Alliance, 2014

Conclusion

- ▶ Establishing and maintaining a collection requires procedures far beyond that of establishing an indexed directory.
- ▶ The process of data ingestion must initiate an entire workflow which accesses properties, assigns policies, controls procedures, and updates persistent state information.
- ▶ The process of data retention and distribution must be policy compliant and automated in both hardware and software.
- ▶ Periodic assessments and reporting can be used to satisfy the audit requirements of a funding body.

Thank You!

Keep in touch with us

Questions?



sales@ddn.com



2929 Patrick Henry Drive
Santa Clara, CA 95054



[@ddn_limitless](https://twitter.com/ddn_limitless)



1.800.837.2298
1.818.700.4000



[company/datadirect-networks](https://www.linkedin.com/company/datadirect-networks)