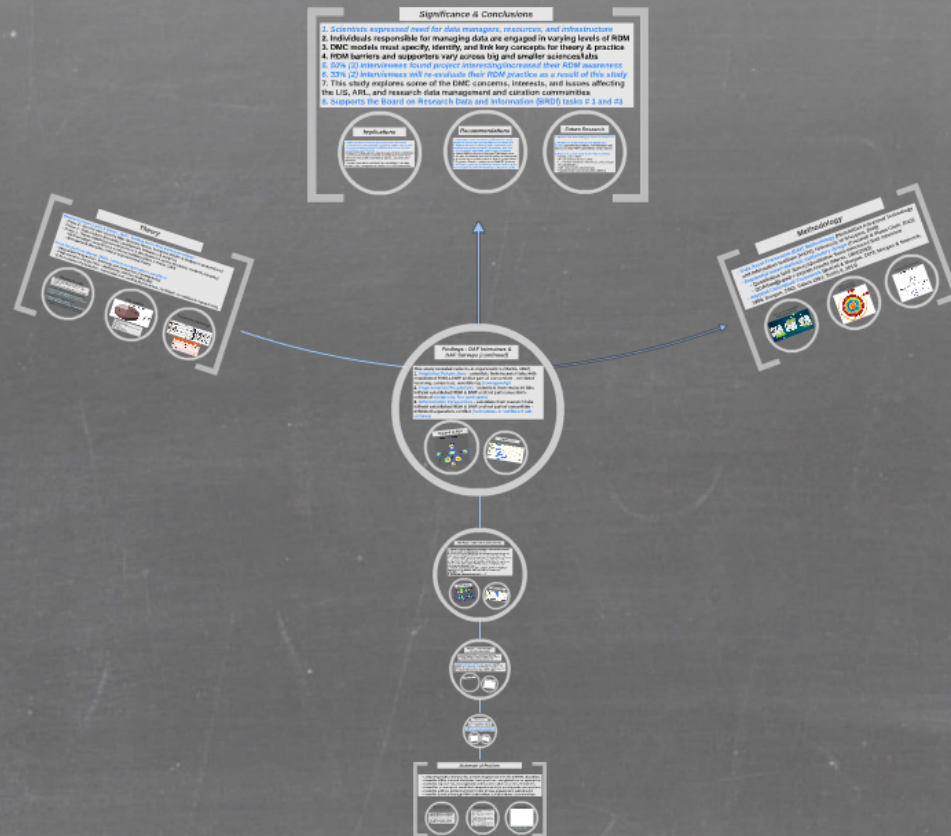


EXPLORING THE DATA MANAGEMENT AND CURATION (DMC) PRACTICES OF SCIENTISTS IN RESEARCH LABS WITHIN A RESEARCH UNIVERSITY

Plato L. Smith II, Doctoral Candidate
 Dissertation Defense
 Florida State University School of Information
 June 23, 2014, 10:00 am - 12:00 pm



EXPLORING THE DATA MANAGEMENT AND CURATION (DMC) PRACTICES OF SCIENTISTS IN RESEARCH LABS WITHIN A RESEARCH UNIVERSITY

Plato L. Smith II, Doctoral Candidate
 Dissertation Defense
 Florida State University School of Information
 June 23, 2014, 10:00 am - 12:00 pm

Statement of Problem

- Competing models/frameworks promote fragmentation in the LIS/ARL disciplines
- A need to utilize relevant standards, best practices, and guidelines as appropriate
- A need to improve data management and curation within & across disciplines
- A need for a conceptual model that integrates multiple paradigmatic perspectives
- A need to address underdeveloped theory of data management and curation
- A need to develop/leverage RDM relationships, collaborations, & partnerships

Research Purpose

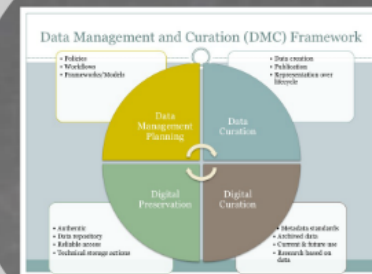
The National Science Foundation (NSF) now requires a data management plan (DMP). Even if a funding agency does not require a data management plan, developing and implementing quality research data management and curation (DMC) practices are needed across all disciplines.

The purpose of this research project is to investigate current DMC practices across multiple disciplines and explore opportunities for improving data management activities.

Research Questions

1. How do researchers create, manage, store, and preserve research data?
2. How can the identification and clarification of key DMC concepts be resolved within and across disciplines?
3. What are some of the theories, practices, and methods disciplines use to address research data management in your discipline?
4. How can multiple paradigmatic perspectives on data management and curation practices within and across disciplinary domains contribute to building DMC research & theory?

Key Concepts



FSU HSC# 2012.9198 - DMCS Opinion Preliminary Study
- 80% (45/56) Strongly agree (32) agree (13)
- Specify, identify, & interrelate key concepts within a scheme; link underlying suppositions (Merton, 1968)

Research Purpose

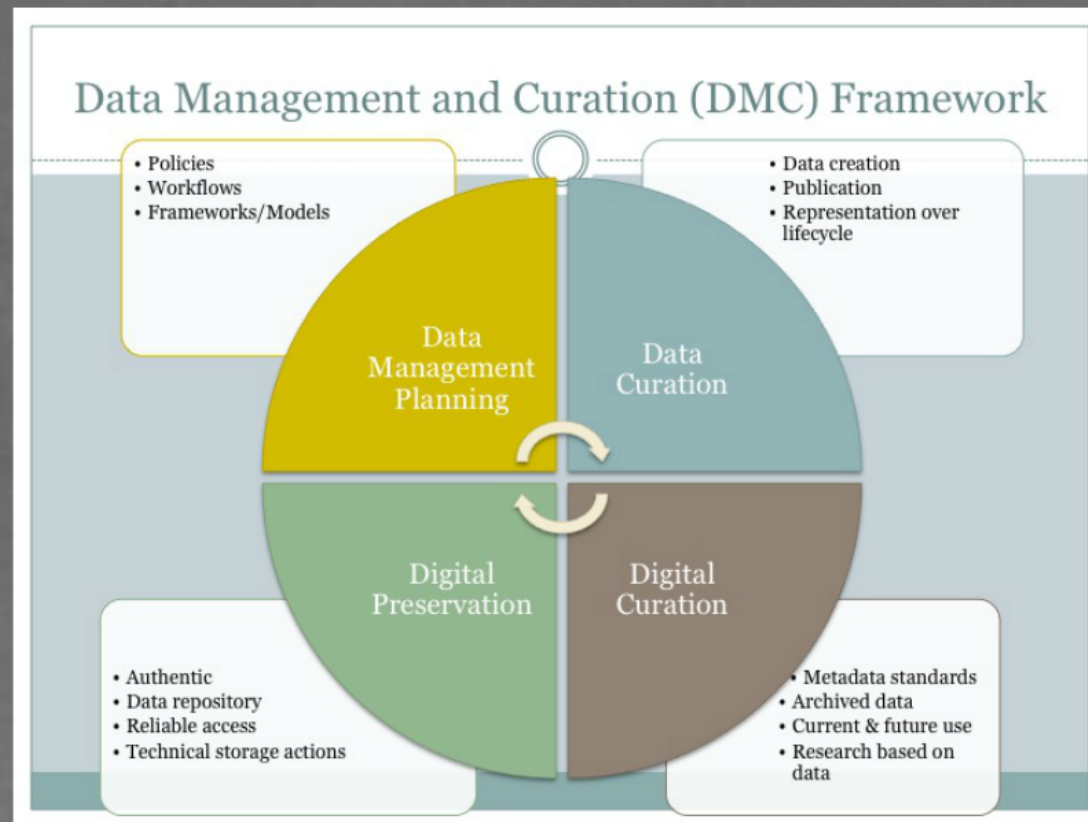
The National Science Foundation (NSF) now requires a data management plan (DMP). Even if a funding agency does not require a data management plan, developing and implementing quality research data management and curation (DMC) practices are needed across all disciplines.

The purpose of this research project is to investigate current DMC practices across multiple disciplines and explore opportunities for improving data management activities.

Research Questions

- 1. How do researchers create, manage, store, and preserve research data?**
- 2. How can the identification and clarification of key DMC concepts be resolved within and across disciplines?**
- 3. What are some of the theories, practices, and methods disciplines use to address research data management in your discipline?**
- 4. How can multiple paradigms perspectives on data management and curation practices within and across disciplinary domains contribute to building DMC research & theory?**

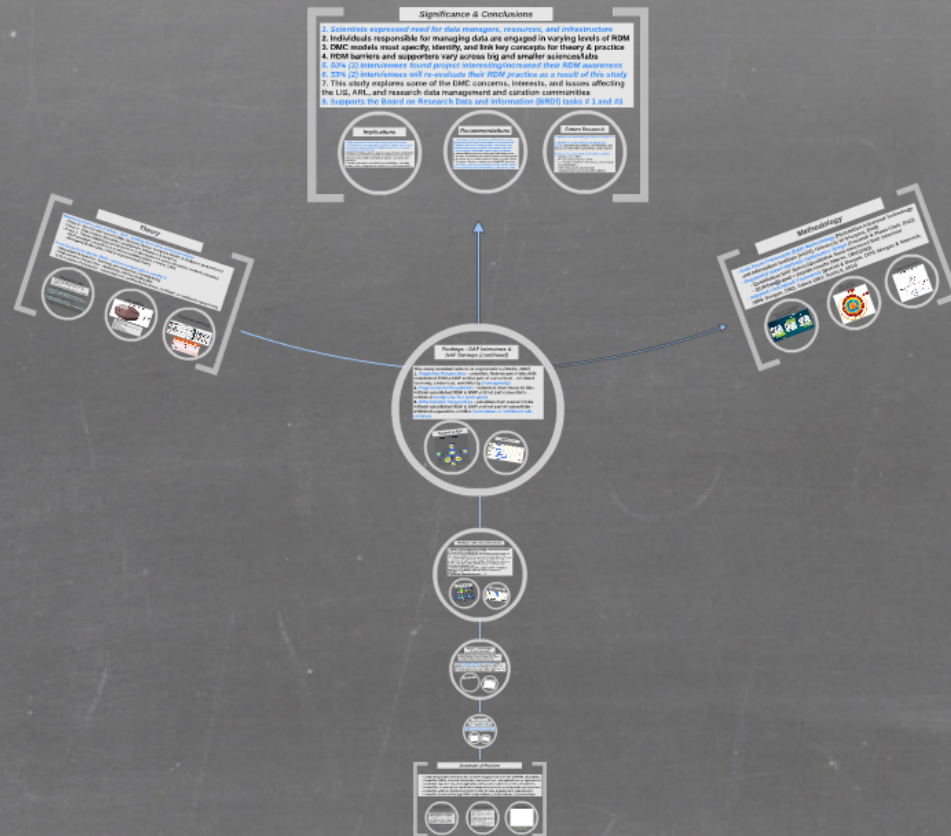
Key Concepts



FSU HSC# 2012.9198 - DMCS Opinion Preliminary Study

- **80% (45/56) Strongly agree (32)/agree (13)**

- Specify, identify, & interrelate key concepts within a scheme; link underlying suppositions (Merton, 1968)



EXPLORING THE DATA MANAGEMENT AND CURATION (DMC) PRACTICES OF SCIENTISTS IN RESEARCH LABS WITHIN A RESEARCH UNIVERSITY

Plato L. Smith II, Doctoral Candidate
 Dissertation Defense
 Florida State University School of Information
 June 23, 2014, 10:00 am - 12:00 pm

Theory

Metatriangulation (Lewis & Grimes, 1999) - building theory from multiple paradigms

- Phase 1 - Ground work (specify RQs, literature review, recognize divides & bridges in perspectives)
- Phase 2 - Data analysis (recognize paradigmatic influences & insights)
- Phase 3 - Theory building (examine patterns & discrepancies across accounts; motivate interplay)
 - Multiparadigm perspectives on theory building (Gioia & Petrie, 1990)
 - Management discipline; field of organizational theory

Three Perspectives (Martin, 1992) - cultures in organizations paradigms

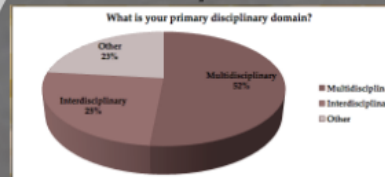
- Integration Perspective - harmony and consensus (homogeneity)
- Fragmentation Perspective - multiplicity and flux (ambiguity)
- Differentiation Perspective - separation and conflict (harmonious, conflicted, or indifferent subcultures)

Research Labs

- 58 Research centers and institutes at FSU
- 6 research labs (centers/institutes) selected
- 6/58 = 0.10
- 10% research labs at FSU sampled for this study

1. Center for Advanced Power Systems (CAPS)
2. Antarctic Marine Geology Research Facility
3. Center for Ocean-Atmospheric Prediction Studies (COAPS)
4. Geophysical Fluid Dynamics (GFDL)
5. Coastal & Marine Laboratory
6. National High Magnetic Field Laboratory (NHMFL)
 - National Science Foundation (NSF) EarthCube
 - FSU HSC# 2013.11448 (Change in protocol)

Disciplines



1. Computer Science
2. Biology and Oceanography
3. Boundary-layer Meteorology and Biogeochemical cycles of water and carbon
4. Meteorology
5. Condensed Matter Physics
6. Materials Science & Physics
7. Marine Ecology, Fisheries Science

Purposive Sampling

Role	Frequency	Percent	*Other
Senior Researcher	23	23%	IT Support
Principal Investigator	29	29%	Postdoctoral research associate
Research Assistant	26	26%	Research associate
Research Technician	3	3%	Operation project manager
Research Support	3	3%	Data management
Research Student	10	10%	Postdoctoral research associate
*Other	7	7%	Postdoc
Total	101	1.01	

DAE	Surveys	Interviews
Starts	129	7
Completes	107	6
Completion Rate	83%	86%

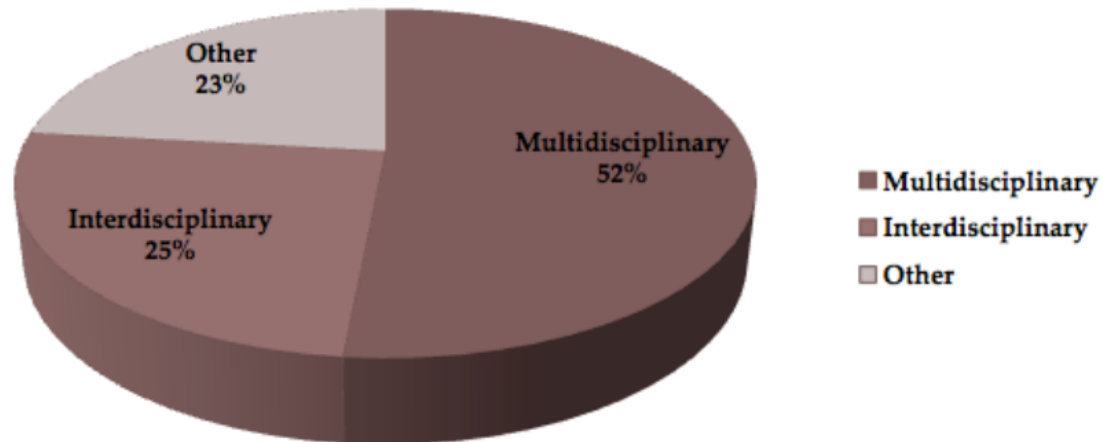
Research Labs

- 58 Research centers and institutes at FSU
- 6 research labs (centers/institutes) selected
- $6/58 = 0.10$
- 10% research labs at FSU sampled for this study

1. Center for Advanced Power Systems (CAPS)
2. *Antarctic Marine Geology Research Facility*
3. *Center for Ocean-Atmospheric Prediction Studies (COAPS)*
4. Geophysical Fluid Dynamics (GFDI)
5. *Coastal & Marine Laboratory*
6. **National High Magnetic Field Laboratory (NHMFL)**
 - *National Science Foundation (NSF) EarthCube*
 - *FSU HSC# 2013.11448 (Change in protocol)*

Disciplines

What is your primary disciplinary domain?



1. Computer Science
2. Biology and Oceanography
3. Boundary-layer Meteorology and Biogeochemical cycles of water and carbon
4. Meteorology
5. Condensed Matter Physics
6. Materials Science & Physics
7. Marine Ecology, Fisheries Science

Purposive Sampling

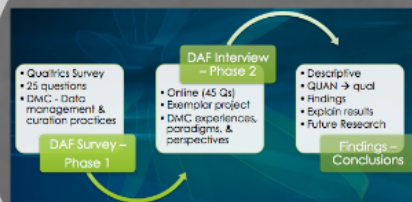
Role	Frequency	Percent	*Other
Senior Researcher	23	23%	IT Support
Principal Investigator	29	29%	Postdoctoral research associate
Research Assistant	26	26%	Research associate
Research Technician	3	3%	Operation project manager
Research Support	3	3%	Data management
Research Student	10	10%	Postdoctoral research associate
*Other	7	7%	Postdoc
Total	101	1.01	

DAF	Surveys	Interviews
Starts	129	7
Completes	107	6
Completion Rate	83%	86%

Methodology

- **Data Asset Framework (DAF) Methodology** (Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, 2009)
- **Sequential mixed-methods explanatory design** (Creswell & Plano Clark, 2011)
 - Quantitative DAF Survey/Qualitative Semi-structured DAF Interview
 - QUAN → qual = explain results (Morse, 1991/2003)
- **Adapted Conceptual Framework** (Burrell & Morgan, 1979; Morgan & Smircich, 1980; Morgan, 1983; Solem 1993; Smith II, 2013)

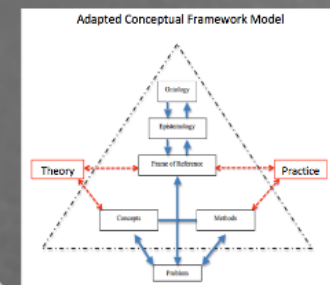
Data Asset Framework (DAF) Methodology



DCC Curation Lifecycle Model



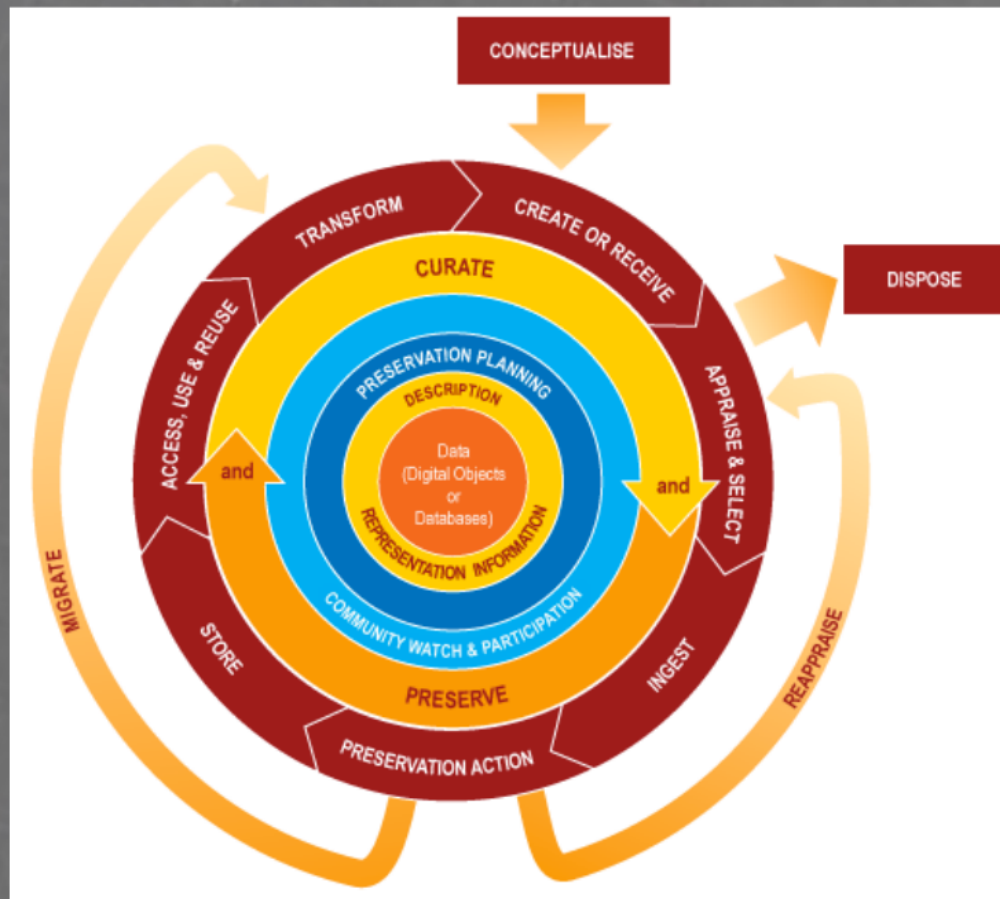
Conceptual Framework



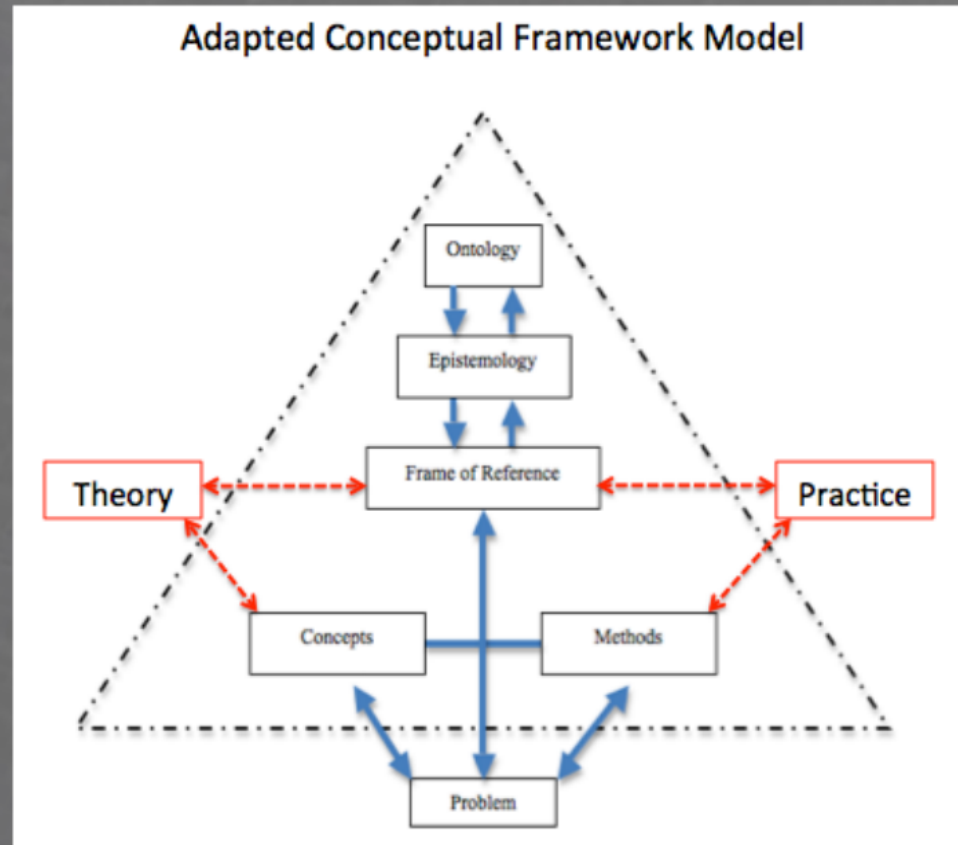
Data Asset Framework (DAF) Methodology

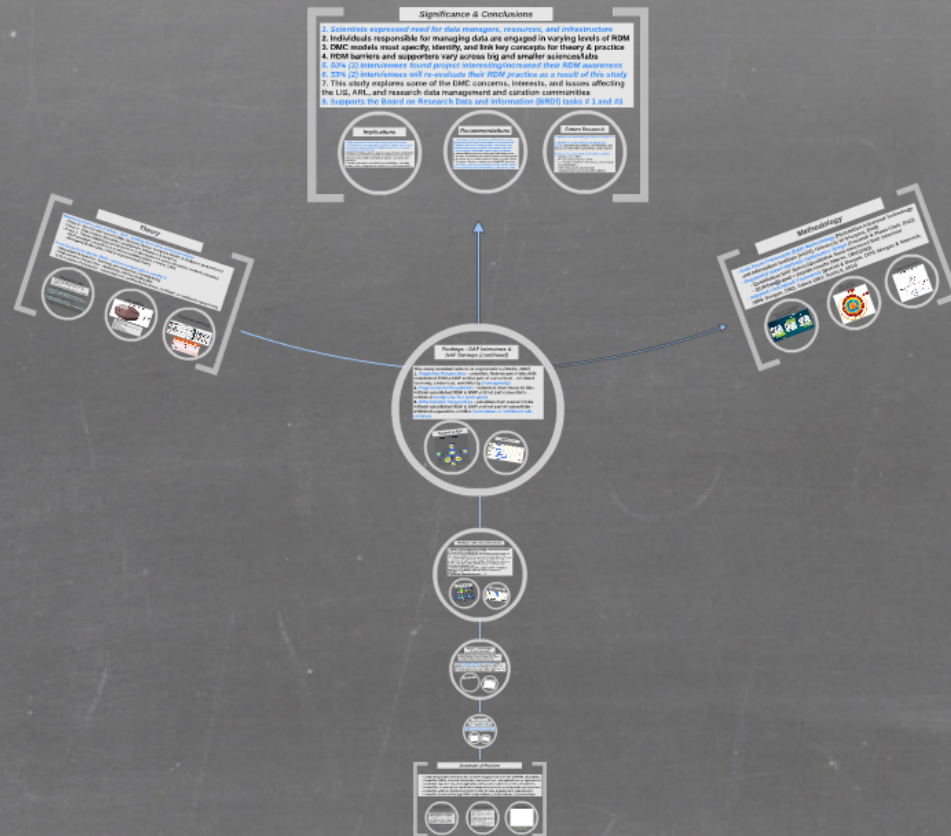


DCC Curation Lifecycle Model



Conceptual Framework





EXPLORING THE DATA MANAGEMENT AND CURATION (DMC) PRACTICES OF SCIENTISTS IN RESEARCH LABS WITHIN A RESEARCH UNIVERSITY

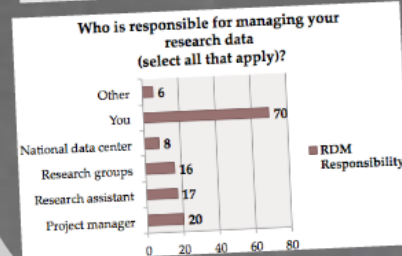
Plato L. Smith II, Doctoral Candidate
 Dissertation Defense
 Florida State University School of Information
 June 23, 2014, 10:00 am - 12:00 pm

Findings - DAF Surveys - FSU HSC# 2013.3073

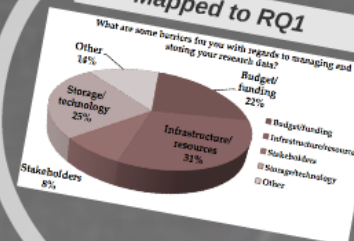
- A.** Deep-C, NCAR, NCDC, NGDC, NODC, NSIDC, ORNL DAAC, NHMFL (*managed data centers/repositories*)
- B.** Lab PC & server, Box/Dropbox, Google Drive, Figshare
- C.** NetCDF, THREDDS, ISO, OPeNDAP, OAIS, ITIL, IEEE

- 1.) 44% (38) were asked by a funder to produce a DMP; 56% (48) were not
- 2.) 35% (28) use standards, best practices, & guidelines; 65% (53) do not
- 3.) 25% (21) No; 48% (41) Maybe; 27% (23) Yes - benefit from data curator
- 4.) 22% (19) Yes; 17% (15) No; 60% (52) Don't know - campus RDM policy
- 5.) 59% (51) ext. HD, 47% (40) computer, and 41% (35) HEI/unit - backup

Mapped to RQ1

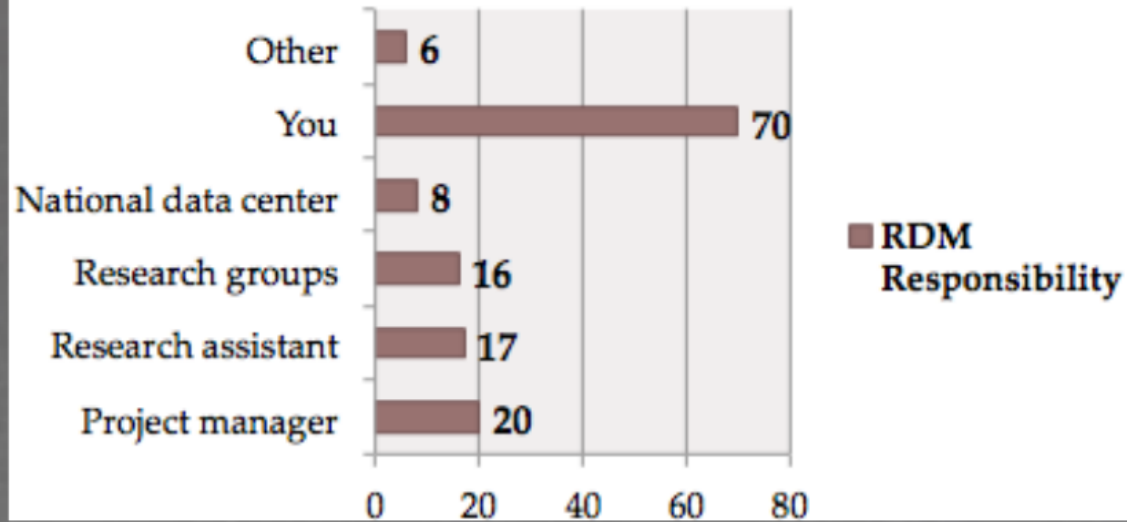


Mapped to RQ1



Mapped to RQ1

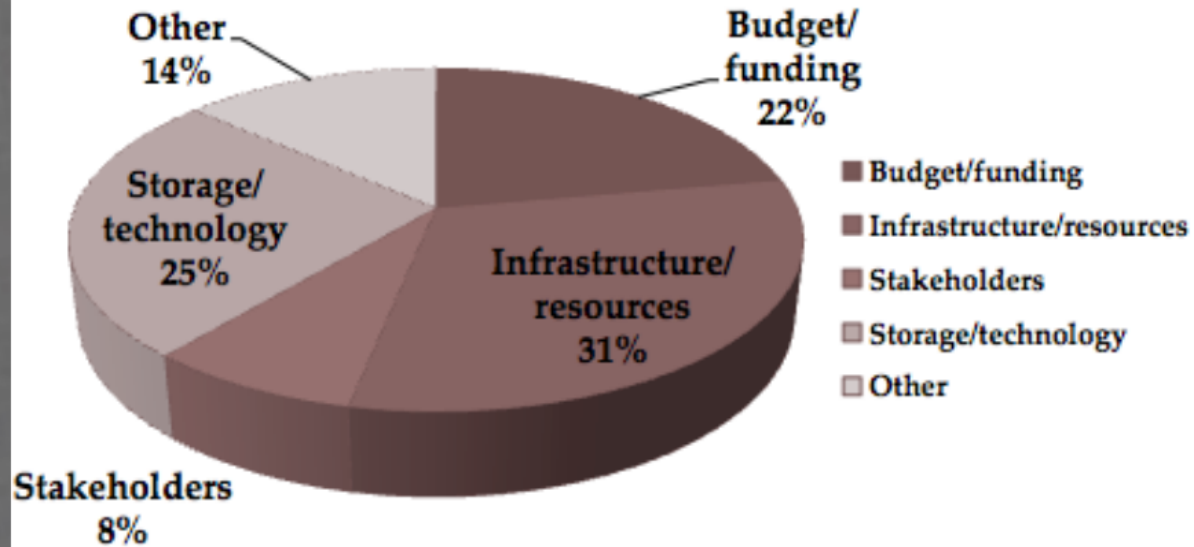
**Who is responsible for managing your research data
(select all that apply)?**



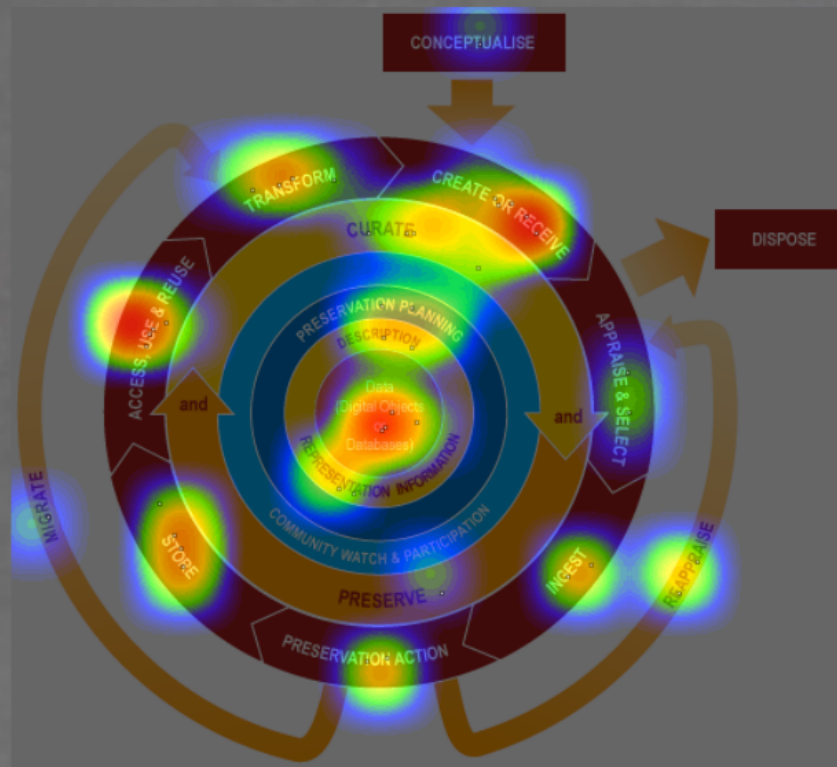
Stakeholder
8%

Mapped to RQ1

What are some barriers for you with regards to managing and storing your research data?



Mapped to RQ2



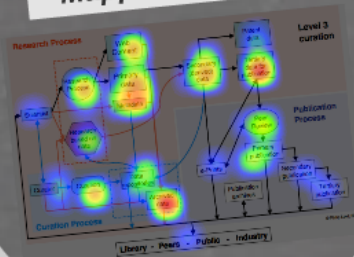
Mapped to RQ2

Region	Response	%
Conceptualize - DMP	6	100%
Data - DMP	10	167%
Curate - Data Curation	4	67%
Community - DMP	-	0%
Preserve - Digital Curation	1	17%
Appraise - Digital Curation	6	100%
Store - Digital Preservation	6	100%
Transform - Digital Curation	5	83%
Access, Use & Reuse - DC	4	67%
Dispose - DMP	-	0%
Other	-	0%
Statistic	Value	
Total Responses	6	

Findings - DAF Interviews (cont.)

1. "Capture of original measurement information and metadata about the process and tools for QA/QC purposes..." - P1
2. "At this point, mostly having good documentation and putting data in a repository or making it discoverable and citable in some way..." - P4
3. "...I am not familiar with any specific theories or concept for research data management, it all evolved as best practice out of experience." - P5
4. "...We ensure that they have best practices for instrument exposure on the vessel and seek to collect sufficient metadata to understand the observations being made." - P3
5. "There is very little theory involved. Basically there is no funding or reward for such an effort so data is only held or managed until publication..." - P6
6. "No theories of systematic approach..." - P7

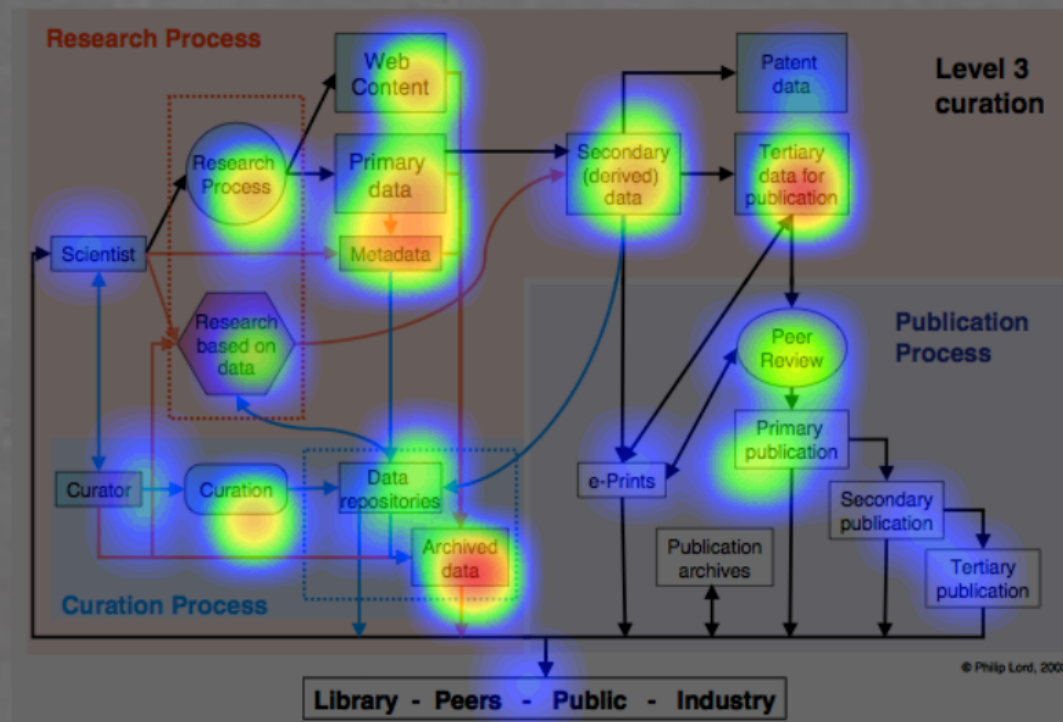
Mapped to RQ3



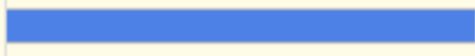

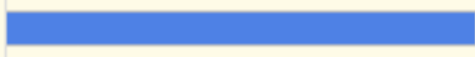


Mapped to RQ3

Region	Responses	%
Research Process - Data Collection	30	42.9%
Collection - Digital	7	100%
Publication - Access	8	114%
Other - Library	10	143%
Other - Public - Industry	1	14%
Total Responses	7	

Mapped to RQ3



Mapped to RQ3

Region		Response	%
Research Process – Data Curation		30	429%
Curation – Digital Curation		7	100%
Curation – Digital Preservation		8	114%
Publication - Access		10	143%
Other – Library – Peers – Public - Industry		1	14%
Statistic			Value
Total Responses			7

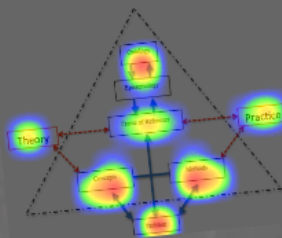
Findings - DAF Interviews & DAF Surveys (continued)

This study revealed *cultures in organizations* (Martin, 1992)

1. **Integration Perspectives** - scientists from research labs with established RDM & DMP and/or part of consortium - exhibited harmony, consensus, consistency (**homogeneity**)
2. **Fragmentation Perspectives** - scientists from research labs without established RDM & DMP and not part consortium - exhibited **multiplicity, flux (ambiguity)**
3. **Differentiation Perspectives** - scientists from research labs without established RDM & DMP and not part of consortium - exhibited separation, conflict (**harmonious or indifferent sub-cultures**)

Mapped to RQ4

Adapted Conceptual Framework Model

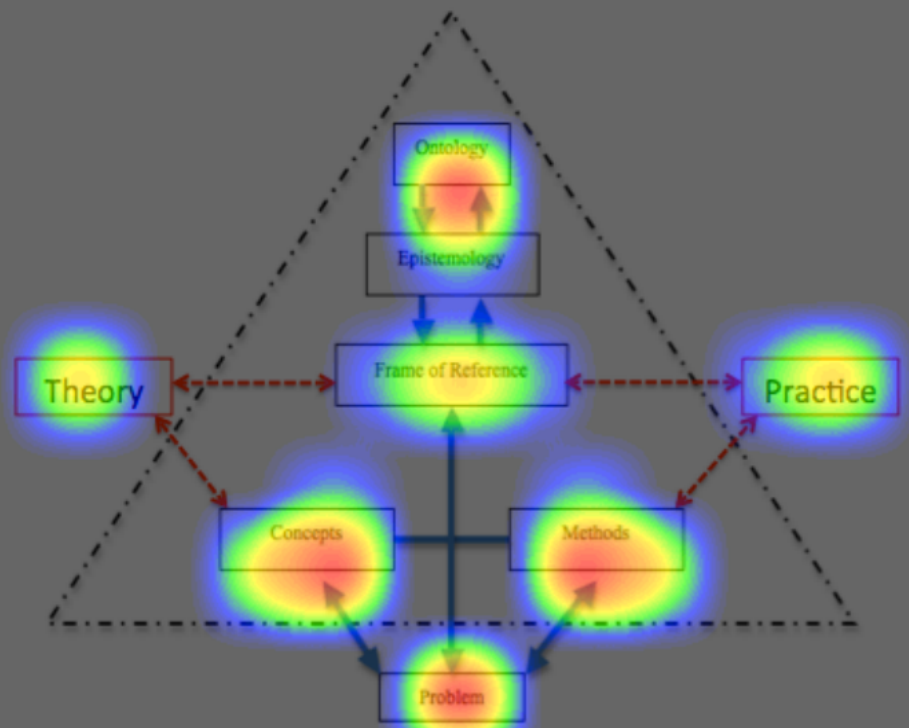


Mapped to RQ4

Region	Response	%
Region #1 - Ontology	3	45%
Region #2 - Epistemology	1	14%
Region #3 - Frame of Reference	4	57%
Region #4 - Practice	3	43%
Region #5 - Theory	6	80%
Region #6 - Culture	2	28%
Region #7 - Methods	6	80%
Region #8 - Tradition	3	43%

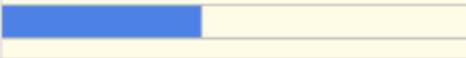
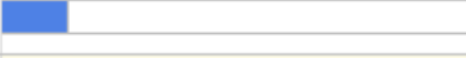
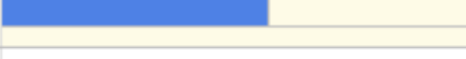

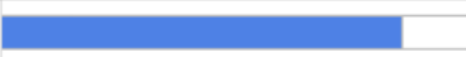
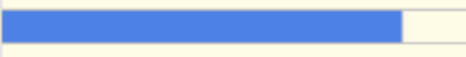

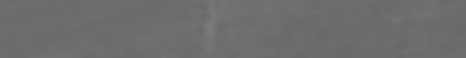
Mapped to RQ4

Adapted Conceptual Framework Model



R
of
Re
Pra
Reg
Regi
Conc
Region
Method
Region
Problem

Mapped to RQ4

Region		Response	%
Region #1 - Ontology		3	43%
Region #2 - Epistemology		1	14%
Region #3 - Frame of Reference		4	57%
Region #4 - Practice		3	43%
Region #5 - Theory		2	29%
Region #6 - Concepts		6	86%
Region #7 - Methods		6	86%
Region #8 - Problem		3	43%

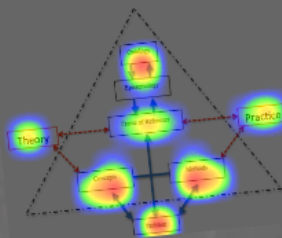
Findings - DAF Interviews & DAF Surveys (continued)

This study revealed *cultures in organizations* (Martin, 1992)

1. **Integration Perspectives** - scientists from research labs with established RDM & DMP and/or part of consortium - exhibited harmony, consensus, consistency (**homogeneity**)
2. **Fragmentation Perspectives** - scientists from research labs without established RDM & DMP and not part consortium - exhibited **multiplicity, flux (ambiguity)**
3. **Differentiation Perspectives** - scientists from research labs without established RDM & DMP and not part of consortium - exhibited separation, conflict (**harmonious or indifferent sub-cultures**)

Mapped to RQ4

Adapted Conceptual Framework Model



Mapped to RQ4

Region	Response	%
Region 01 - Ontology	3	45%
Region 02 - Epistemology	1	14%
Region 03 - Frame of Reference	4	57%
Region 04 - Practice	3	43%
Region 05 - Theory	6	86%
Region 06 - Context	2	29%
Region 07 - Methods	6	86%
Region 08 - Tradition	3	43%

Significance & Conclusions

- 1. Scientists expressed need for data managers, resources, and infrastructure**
- 2. Individuals responsible for managing data are engaged in varying levels of RDM**
- 3. DMC models must specify, identify, and link key concepts for theory & practice**
- 4. RDM barriers and supporters vary across big and smaller sciences/labs**
- 5. 50% (3) interviewees found project interesting/increased their RDM awareness**
- 6. 33% (2) interviewees will re-evaluate their RDM practice as a result of this study**
- 7. This study explores some of the DMC concerns, interests, and issues affecting the LIS, ARL, and research data management and curation communities**
- 8. Supports the Board on Research Data and Information (BRDI) tasks # 1 and #3**

Implications

- 1. Raise organized research data management awareness**
- 2. Stimulate data management cognition & cogent data policies**
- 3. Improve departmental and institutional level research data management accountability**
- 4. Support funding agencies data management plan requirement**
- 5. Increase data access, resource discovery, data use & reuse**
- 6. Increase the usefulness of data to science, research, and education**
- 7. Assist scientists/research labs in contributing to the wider research data management & curation scholarly communities**

Recommendations

- 1. Leverage current resources, collaborations, and campus & consortium partnerships to maximize ROI**
- 2. Expand the use of existing tools, resources, and infrastructure across projects, disciplines, and labs**
- 3. Develop DMC education, workshops, & projects**
- 4. Lower RDM barriers and increase RDM supporters**
- 5. Include all stakeholders in DMC policy developments**
- 6. Connect big & smaller sciences across projects/labs**
- 7. Register a research project as an EXSEDE gateway**
- 8. Develop a gateway to publicly-funded research RCN (i.e. Research Council UK Gateway to Research (GTR))**

Future Research

- 1. Update DAF methodology & increase sample size**
- 2. Perform an audit of physical & digital data assets; approved repositories; non-managed data stores; working RDM relationships/collaborations**
- 3. Expand research study to non-FSU scientists**
 - Deep-C Consortium
 - Florida Climate Institute (FCI)
 - Los Alamos National Laboratory & University of Florida (NHMFL)
 - NSF Antarctic Data Consortium
 - National Snow & Ice Data Center (NSIDC)



Implications

1. Raise organized research data management awareness
2. Stimulate data management cognition & cogent data policies
3. Improve departmental and institutional level research data management accountability
4. Support funding agencies data management plan requirement
5. Increase data access, resource discovery, data use & reuse
6. Increase the usefulness of data to science, research, and education
7. Assist scientists/research labs in contributing to the wider research data management & curation scholarly communities

Recommendations

- 1. Leverage current resources, collaborations, and campus & consortium partnerships to maximize ROI*
- 2. Expand the use of existing tools, resources, and infrastructure across projects, disciplines, and labs*
- 3. Develop DMC education, workshops, & projects*
- 4. Lower RDM barriers and increase RDM supporters**
- 5. Include all stakeholders in DMC policy developments**
- 6. Connect big & smaller sciences across projects/labs**
- 7. Register a research project as an EXSEDE gateway**
- 8. Develop a gateway to publicly-funded research RCN (i.e. Research Council UK Gateway to Research (GtR))*

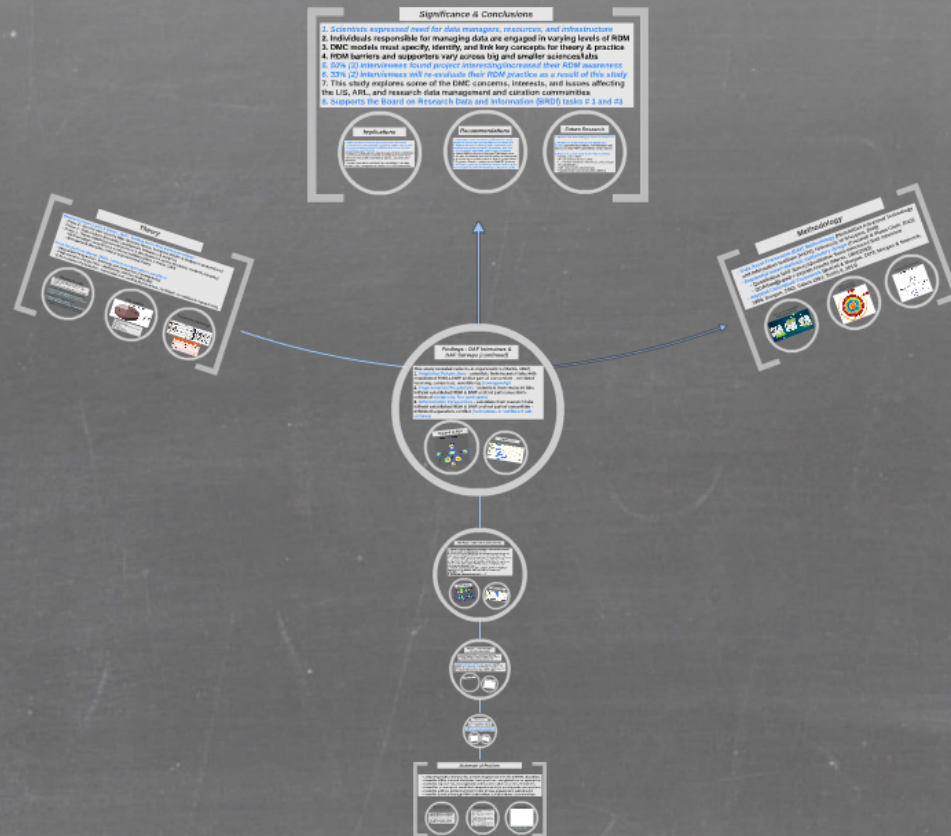
Future Research

1. Update DAF methodology & increase sample size

2. Perform an audit of physical & digital data assets; approved repositories; non-managed data stores; working RDM relationships/collaborations

3. Expand research study to non-FSU scientists

- Deep-C Consortium
- Florida Climate Institute (FCI)
- Los Alamos National Laboratory & University of Florida (NHMFL)
- NSF Antarctic Data Consortium
- National Snow & Ice Data Center (NSIDC)



EXPLORING THE DATA MANAGEMENT AND CURATION (DMC) PRACTICES OF SCIENTISTS IN RESEARCH LABS WITHIN A RESEARCH UNIVERSITY

Plato L. Smith II, Doctoral Candidate
 Dissertation Defense
 Florida State University School of Information
 June 23, 2014, 10:00 am - 12:00 pm