

# Data Management/Curation Task Force: Year One Report and Recommendations

---

## Contents

- Executive Summary
  - Selected Accomplishments of Note
  - Key Findings and Recommendations
    - 1: Formalize the Program for Data Management/Curation
    - 2: Foster Collaborations to Grow the Culture of Data Management
    - 3: Conduct Research, Development, and Assessment Necessary for UF's Data Needs
- Priority Next Steps and Recommendations<sup>1</sup>
- Background for Charging the Data Management/Curation Task Force
- Year One Activities: Overview, Current Status, and Recommendations for Year Two
- Appendices:
  - Proposal for Digital Humanities Library Group
  - Data Survey Results

---

<sup>1</sup> Next steps and recommendations are also listed by area with the Year One Activities.

## Executive Summary

The first year of Data Management/Curation Task Force (DMCTF) focused on assessment, information gathering, planning, and development activities from January to December 2013.<sup>2</sup> The DMCTF has been enormously successful in many regards, as detailed in this report. However, based on the information gathered over the course of the past year and ongoing trends, a great deal needs to be done. Further, it needs to be done as quickly as possible given internal and external pressures for data supports.

### Selected Accomplishments of Note

The DMCTF has been extremely active in a variety of areas detailed in this report. Activities of particular importance and success include:

- Events and Trainings
  - “Big Data, Little Data: Having It All” event in fall with over 100 attendees
  - Graduate Student Events: many, focused on data outreach and training
  - Conducted trainings on data reference, the Data Management Plan Tool (DMPTool), and Best Practices in Research Data Management
  - SobekCM Digital Repository Training Series: in-person and webinar trainings with videos online for future use; faculty and graduate student attendees reported interest in SobekCM for research project data for independent and collaborative projects
- Tools Developed
  - Data Management Plan Tool (DMPTool) customized for writing data management plans for grants implemented for UF (preparing for DMPTool 2.0 coming in Spring 2014)
  - Website at <http://www.uflib.ufl.edu/datamgmt/> built around the research life cycle, including links to storage options on campus and training activities
- Assessment and Data Gathering
  - Campus-wide Data Survey in fall 2013; over 280 responses confirming the critical data needs in many areas
  - Facilitated meetings and discussions with departments, centers, and institutes on data
  - Evaluation of the Dataverse Network (DVN) with the Southeastern Universities Research Association (SURA) with data from the process supporting Research Computing implementation of new data storage system
- Building Collaboration and Awareness
  - Facilitated meetings and discussions with other campus service providers for integrated services with Research Computing, Division of Sponsored Programs, the Graduate School, and others
  - Coordinated visit by FSU Libraries for UF Research Computing Day
  - Coordinated presentation by Bess de Farber to the Research Computing Advisory Committee (Jan. 2014) to discuss collaborative processes and events, including CoLABs
  - DMCTF presentation and update to the Libraries in fall 2013

---

<sup>2</sup> DMCTF charge: <http://ufdc.ufl.edu/AA00014835/00011>

## Key Findings and Recommendations

Given the DMCTF's productive work and rapid progress, accelerating activities is not feasible without additional staffing. For future phases of work to support UF's needs, the DMCTF findings and recommendations fall within three primary areas:

1. Formalizing the program for data management
  - a. Hiring critically needed personnel
  - b. Creating a formalized structure in the Libraries, which will serve as a connected node with other groups in the Libraries and across campus
  - c. Developing standardized support for common needs, where possible and applicable
2. Fostering collaborations across campus to grow the culture of data management
3. Conducting research, development, and assessment of UF's data needs

### Key Recommendation 1: Formalize the Program for Data Management/Curation

Critically needed personnel include a Data Curation Coordinator Librarian and Data IT Expert. Both will need to be positioned to collaborate throughout the Libraries and with Research Computing for integration of services:

- The Data IT Expert could report through the Digital Development Unit, be embedded within the data team, and act as the technical contact for collaboration with Library IT, Research Computing, and other campus IT units. The Data IT Expert is needed to support testing and setup of any new tools, integrate and add additional data supports to the *IR@UF*, collaborate with Research Computing to integrate the SobekCM front-end data/digital repository system supported by the Libraries with the back-end computational processing and granularly controlled storage on Research Computing resources.
- The Data Librarian would serve on project teams and, more importantly, support other librarians in engaging in data management and curation issues<sup>3</sup>, helping to lay the foundation for creating a culture of data management as part of the human and technical infrastructure work necessary to support modern research needs and radical collaboration.<sup>4</sup>

Further formalization of the program for data management/curation is needed once the new positions are hired. Many of UF's peer institution libraries have already created formal data units within their libraries. Structured in various ways<sup>5</sup>, they are designed to connect various areas and experts from existing library structures and support new data activities. For UF, the most successful structure would connect existing experts, serve as the home for new positions and data activities, and serve as a critical

---

<sup>3</sup> Carlson, Jake (2013). "Opportunities and Barriers for Librarians in Exploring Data: Observations from the Data Curation Profile Workshops." *Journal of eScience Librarianship* 2(2): <http://dx.doi.org/10.7191/jeslib.2013.1042>

<sup>4</sup> For radical collaboration defined in relation to UF and research, see: *Vision for Research Computing at UF* (Feb 2011), <http://www.it.ufl.edu/wp-content/uploads/2012/03/research-computing-vision.pdf>

<sup>5</sup> Examples include: Publishing & Curation Services, Office of Digital Scholarly Publishing, Data Management Consulting Group, Scholars' Lab, Scholarly Commons.

hub to connect with other groups in the Libraries and across campus. The DMCTF is already working to build this hub structure through connecting with other Library groups, as explained in the next recommendation.

### **Key Recommendation 2: Foster Collaborations to Grow the Culture of Data Management**

With or without additional personnel, the DMCTF has a critical role to play in developing a culture of data management within the Libraries and extending through collaboration with other campus service groups to the full campus. Ideally, the DMCTF will continue to focus efforts on specific work by the group and work done in collaboration with various library groups:

- Collaborating on data-related activities: Library Instruction Committee for data instruction.
- Identifying and activating other best-fit collaborations: Born Digital Archival Content Working Group, Grants Committee<sup>6</sup>, and others.
- Providing an excellent hub framework and configuration for supporting new groups for data much as the new Digital Humanities Library Group.<sup>7</sup>

In addition to the DMCTF, data specific groups, and core collaborators outside of the libraries like Research Computing, all Liaison Librarians have a critical role as the primary contacts for data and as leaders on consultative teams within the libraries for data and other digital scholarship needs.<sup>8</sup> To further foster collaboration and grow the culture of data management, the Libraries need ongoing work internally with Liaisons, and through to external groups. For external group collaboration, the DMCTF recommends approaching best fit, targeted groups to build collaboration on data. Additionally, the Libraries have a wealth of expertise to offer for all aspects of data management and curation, including support on attribution and fair-cite initiatives for attribution and credit for data scientists and computational team members for data projects.

### **Key Recommendation 3: Conduct Research, Development, and Assessment Necessary for UF's Data Needs**

Because data management and curation impact all types of scholarly work, all scholarly areas and fields, and all aspects of academic institutions, ongoing research, development, and assessment are needed to ensure support for immediate campus data needs and build support for future needs. Currently, the DMCTF can best support this ongoing work in collaboration with other library and campus groups. For 2014, the DMCTF can best conduct assessment and outreach on data in collaboration with Research Computing in their efforts for outreach and promotion of Research Computing and HiPerGator.

---

<sup>6</sup> The DMCTF is collaborating with the Libraries' Grants Management Committee, which is discussing next steps on data and grant related supports. In December 2013, DMCTF members, others in the Libraries, and the Grants Manager met with the Division of Sponsored Programs on collaboration related to data, grants, etc.

<sup>7</sup> The Digital Humanities Library Group (<http://ufdc.ufl.edu/AA00014835/00030/pdf>) developed from the DMCTF.

<sup>8</sup> For liaison roles on data and digital scholarship project teams, see: <http://ufdc.ufl.edu/AA00017119/00021/pdf>

## Priority Next Steps and Recommendations

For the next phase of activities, the DMCTF has identified next steps for the DMCTF in collaboration with others. The DMCTF has also identified recommendations, which are dependent upon administrative approval or prioritization.

Detailed next steps and recommendations are included within the section on year one activities. Selected, priority next steps and recommendations are included here for ease of review.

### Next Steps

- Present at the Selectors Meeting on the Library Liaison Team Model for data and data resources and support available from the DMCTF
- Schedule presentations at library faculty departmental meetings
- Integrate data management whenever possible with existing groups:
  - Library Instruction (for trainings and possible for-credit data courses)
  - Library Liaisons/Selectors (for integrated reference/consultation, referral, and promotion of existing data resources, information gathering for new needs, and overall communication with all of UF for data needs)
  - Born Digital Archival Content Working Group
  - Grants Committee
  - SobekCM user group
  - Authors@UF for data associated with publications
  - Support developing new groups as needed:
    - Digital Humanities Library Group (new interest group launched in 2014)
- Implement an annual survey in fall on data (add questions on Research Computing, HiPerGator)
- Pursue integration with the Division of Sponsored Programs; specifically:
  - DMPTool Training: promote and integrate with DSP trainings
  - *IR@UF* implementation of an authority system for ORCID and other identifiers<sup>9</sup>
- Support Librarians in delivering trainings for the DMPTool<sup>10</sup> and other trainings
- Continue development of the DMPTool with the version 2 release in 2014
- Submit a grant proposal to fund name disambiguation support by integrating an authority system in SobekCM (for the *IR@UF* and all UF resources), with future plans for integration with the Division of Sponsored Programs.
- Develop funding proposals to integrate data support with the Libraries and Research Computing's systems
- Investigate creating an introductory data course (akin to a data literacy course).
- Hold events, specifically targeting an "Elegance of Data" event
- Collaboratively write a brief (1 page) news article with Research Computing for UFIT news

---

<sup>9</sup> Discussed in the meeting with DSP, with Emerging Technologies grant proposal targeted for 2014.

<sup>10</sup> The DMCTF discussed the DMPTool trainings with the Office of Research and Research Computing. Both will refer and promote this centralized service as provided by the Libraries.

- Collaborate with Research Computing on outreach for resources and services from Research Computing and HiPerGator.

## Recommendations

The DMCTF recommends a variety to activities to grow the campus-wide culture of data management in collaboration, and to grow the culture of data management within the Libraries. The DMCTF recommends incorporating data into existing Liaison activities, and has created a many resources to support these activities.<sup>11</sup>

### Recommendations in collaboration with Research Computing:

- Strengthen the collaboration with Research Computing, and work to cross-promote Research Computing as a core, library-related resource. Specific activities for the next phase include collaboration on the outreach plan (to be released in spring 2014).
- Build on the collaboration with Research Computing to connect with campus IT groups and campus IT professionals supporting data, targeting specific “best match” groups for shared data needs:<sup>12</sup>
  - Informatics Institute
  - Selected IT groups and individuals
  - Selected campus units/groups: Samuel Proctor Oral History Program, ICBR, and other campus units/groups selected to expand connections into new areas while also strengthening existing collaborations on data.
- Implement a pilot program for a “Visualization Open House” or Visualization Lab” in Marston 308 (or other appropriate space), with set lab hours for specific days/times each week for at least one full semester, with the GIS Librarian holding the open hours, with Research Computing if available

### Recommendations for Liaison Librarians:

- Contact their departments and faculty to:
  - Provide an update on data services from the libraries (draft text<sup>13</sup>)

---

<sup>11</sup> For instance, see this draft text on data support and services: <http://ufdc.ufl.edu/AA00019190>. Other resources include standard, template slides and handouts on data management resources from the Libraries and Research Computing, and a guide for supporting data at the reference desk.

<sup>12</sup> The “best matches” must be involved with data, but also include those in areas of dramatic change who are positioned to enable radical collaboration. The age of Big Data blends previously separated areas for librarians (e.g., functional, technical, and subject expertise) in a parallel manner to the blended expertise necessary for research computing (e.g., systems administration, programming, database administration, experimental processes, concerns and needs specific to academic research as with understanding the difference in acceptable error rates in research versus commercial web searching, and how to support the rigorous requirements for research). The blurred roles and goals for the related “application engineer” roles are also useful in identifying potential “best matches” (see <http://llc.oxfordjournals.org/content/26/2/217.full>).

- Request information regarding “Dinky Databases”<sup>14</sup> using the example request text<sup>15</sup>, sharing all findings with the DMCTF
- Request information on data-related courses in the department, and share about data management support resources from the Libraries specifically relevant for the course
- Update LibGuides and appropriate websites to add a tab or page for data with the links and materials below:
  - Data website, <http://library.uflib.ufl.edu/datamgmt/>
  - IR@UF, <http://library.ufl.edu/uflir>
  - DMPTool
  - Research Computing, <http://rc.ufl.edu>
  - HiPerGator, <http://www.hpc.ufl.edu/2013/06/hipergator/>
  - Appropriate subject repositories (or DataBib for finding resources)
  - Additional, appropriate resources for data identified by the Liaisons
- Include and integrate data into existing and new:
  - Presentations and trainings (e.g., introductions and overviews, new graduate student and postdoc orientations,<sup>16</sup> resources materials, basic handouts, etc.)<sup>17</sup>
  - Events (explicitly mentioning data support from the Libraries and Research Computing resources)
  - Courses taught by librarians
  - Processes for retiring faculty research materials
- Watch the SobekCM training videos to support data needs using SobekCM when appropriate.
  - Join the SobekCM user group to best support data, digital collections, and user needs.<sup>18</sup>

---

<sup>13</sup> <http://ufdc.ufl.edu/AA00019190/00001>

<sup>14</sup> <http://ufdc.ufl.edu/l/AA00014835/00022/pdf> and <https://docs.google.com/spreadsheet/ccc?key=0AoYPOTobTSykdDNUeEprN01DQzIBdE10T19BRnRLRUE#gid=0>

<sup>15</sup> See example call for information: <http://cms.uflib.ufl.edu/datamgmt/contactus>

<sup>16</sup> [http://guides.uflib.ufl.edu/grad\\_orientation](http://guides.uflib.ufl.edu/grad_orientation) and <http://guides.uflib.ufl.edu/postdoc>

<sup>17</sup> <http://guides.uflib.ufl.edu/content.php?pid=49813&sid=435652>

<sup>18</sup> SobekCM User Group: <https://groups.google.com/forum/#!categories/sobekcm-discuss/general>

## Background for Charging the Data Management/Curation Task Force

The Office of Science and Technology Policy's data mandate, funder requirements for data management plans and activities, and significant ongoing changes to research and teaching practices in the digital or the data age make data management a priority. Data management needs parallel and intersect with changes in academic libraries and librarianship which are enabling new types of data support.<sup>19</sup> Academic institutions are facing a critical, urgent need for data management alongside of needs and opportunity with academic libraries for creating, sustaining, and transforming cultures of data management.

Recognizing the need and opportunity, in 2011 UF created Research Computing to collaborate with other campus groups to enable radical collaboration<sup>20</sup> and the Research Computing Advisory Committee tasked a subcommittee on data lifecycle management. Research Computing continues to develop technological infrastructure at UF and in collaboration in the State and beyond. In 2012, the UF Smathers Libraries and Research Computing collaboratively developed the *Strategic Agenda for E-Science and BIGDATA* to address the human infrastructure at UF and beyond, to create a culture of data management and enable radical collaboration. Following this, the Data Management/Curation Task Force (DMCTF) began in 2013 with representatives from the Libraries, Research Computing, and the Office of Research. The DMCTF was charged to assess needs, make recommendations, and develop support for the role of the Libraries in campus-wide data management and curation.

While the initial DMCTF charge focused on assessment and information gathering, the DMCTF members agreed that the charge had to be extended. The extension was to better support campus data needs (immediate and long-term) with currently available resources, begin building additional supports in the best manner possible as integrated with other groups across campus, and to undertake activities and develop plans for growing the full campus-wide culture of data management.<sup>21</sup>

---

<sup>19</sup> For a longer review, see: <http://www.clir.org/pubs/reports/pub160/pub160.pdf>

<sup>20</sup> <http://www.it.ufl.edu/wp-content/uploads/2012/03/research-computing-vision.pdf>

<sup>21</sup> Q1 Report: <http://ufdc.ufl.edu/l/AA00014835/00001/pdf>

## Year One Activities: Overview, Current Status, and Recommendations for Year Two

The DMCTF charge included advisory and operational activities. Those activities, the status, and next steps are detailed below. Where the charge listed specific advisory and operational activities as separate groupings, the activities needed to support data management blend assessment, advisory, testing, and operational activities. Because of this, the advisory and operational activities are grouped together, where useful.

### Charge: Formally assess, through surveys, interviews, and focus groups, campus-wide data management needs and current support resources and activities

**Status:** Formal and informal methods were used to help bridge translation gaps (e.g., some researchers do not identify their data as “data”) and included: focused group discussions; formal interviews; informal information gathering through trainings, workshops, facilitated discussions, reference meetings, etc.; responses and discussions from the “Big Data, Little Data” event attendees; and a campus-wide survey with 288 responses).

#### Next steps:

- Implement an annual survey in fall on data (add questions on Research Computing, HiPerGator)
- Continue to assess and develop recommendations for data needs and to foster a culture of and comprehensive support for a data management
- Continue to assess and develop recommendations for specific areas of need; identify appropriate projects and develop plans to build to recommendations
- Continue work on “Dinky Databases” as an identified project:
  - Dinky databases<sup>22</sup> are a known problem where researchers have small datasets and databases on personal, departmental, and other servers. With the current uneven support, the problem is undefined. With the problem defined, the next steps would be to assess these needs to develop support options and a plan for implementing support that continues to build toward the long-term goals for providing better researcher support and better centralized data support for possible new opportunities (e.g., for some fields, this is a project to define a service as with ArcServer for GIS).
  - The DMCTF recommends that Library Liaisons be asked to contact their departments and researchers on this problem, and to collect and share information with the DMCTF<sup>23</sup> and also with Research Computing.
  - After liaison contact, the DMCTF could coordinate and establish the needed collaborative teams, including collaborating with the appropriate IT (Library and

---

<sup>22</sup> <http://ufdc.ufl.edu/l/AA00014835/00022/pdf> and <https://docs.google.com/spreadsheet/ccc?key=0AoYPOTobTSykdDNUEprN01DQzIbDE10T19BRnRLRUE#gid=0>

<sup>23</sup> See example call for information: <http://cms.uflib.ufl.edu/datamgmt/contactus>

Research Computing) units to migrate databases to the *IR@UF* when appropriate, inform and collaborate on development for SobekCM (which powers the *IR@UF*) for appropriate database support, and collaborating on additional supports as appropriate which may include developing other centralized solutions. The DMCTF could then collaboratively support migrating the databases for permanent support and leveraging of capacity with the appropriate solutions.

### Charge: Review and consider the best practices and models of peer institutions

**Status:** The DMCTF reviewed the rich variety and abundance of activities by peer institutions. Programmatic activities, plans, and best practices are still largely in development with wide variation based on institutional particularities and resource availability. Many institutions are engaged in activities and have developed structures that are informative for UF's best practices. For instance, Purdue has an integrated grant process so that when grant proposals are submitted, the librarians are notified and can follow up regarding data management and work to ensure ongoing integrated technical operations and open communication.<sup>24</sup>

The Libraries at Emory University<sup>25</sup> and Notre Dame University<sup>26</sup> created new centers for data support (providing central, integrated support for data management/curation, GIS, digital scholarship, and related). Similarly, Pennsylvania State University created the new Publishing and Curation Services Unit.<sup>27</sup> Each of these library units offer consultative services and the units act as central nodes that connect to other campus areas for collaboration and to provide integrated campus-wide support.

In the UF Libraries' *Fiscal Year 2013-2014 Budget Review* report for RCM, the UF Libraries identified the core need for a Data Curation Librarian and IT Experts within the libraries,<sup>28</sup> noting that UF's Libraries fill the role of the intellectual ombudsman as they bring disciplines together in a rapidly changing environment. The need for these positions continues to grow. Based on peer institutions, the preferred model includes new data positions and a formalized library unit that provides data and related services.

Aside from new personnel, best practices from academic libraries include integration with other campus groups and within the libraries themselves.

#### Next steps:

- Pursue integration with the Division of Sponsored Programs; specifically:
  - DMPTool Training: promote and integrate with DSP trainings

---

<sup>24</sup> This has been discussed with the Division of Sponsored Programs, and should be available in mid-2014, following a system upgrade by DSP for their data systems.

<sup>25</sup> See: <http://digitalscholarship.emory.edu/about/> and <http://digitalscholarship.emory.edu/research/data%20management.html>

<sup>26</sup> See: <http://library.nd.edu/cds/> and <http://library.nd.edu/cds/expertise/DataManagement.shtml>

<sup>27</sup> See: <http://www.libraries.psu.edu/psul/pubcur.html>

<sup>28</sup> See page 27: <http://ufdc.ufl.edu/IR00001359/00001>

- *IR@UF* implementation of an authority system for ORCID and other identifiers<sup>29</sup>
- Integrate data management whenever possible with existing groups:
  - Library Instruction (for trainings and possible for-credit data courses)
  - Library Liaisons/Selectors (for integrated reference/consultation, referral, and promotion of existing data resources, information gathering for new needs, and overall communication with all of UF for data needs)
  - Born Digital Archival Content Working Group
  - Grants Committee
  - SobekCM user group
  - Authors@UF for data associated with publications
- Support developing new groups as needed:
  - Digital Humanities Library Group (new interest group launched in 2014)

### **Charge: Develop and implement templates and support training and services for the DMPTool (Data Management Plan Tool) and other resources**

**Status:** The DMPTool is a tool that provides a structured guide for the process of writing a data management plan, with tailored supports for different funding agencies and programs. The DMCTF collaborated with others in the Libraries to deliver trainings and instruction focused on and featuring the DMPTool.<sup>30</sup> The DMCTF collected example data management plans<sup>31</sup> from UF for use by others as models in writing their plans.

#### **Next Steps:**

- Continue to deliver hands-on trainings with the DMPTool are needed on an ongoing basis<sup>32</sup>
- Continue development of the DMPTool with the version 2 release in 2014
- Continue to collaborate with the Library Instruction Committee on data training and instruction for librarians, faculty, and students.
  - Specific training of interest includes training on the DMPTool (interest in overviews on the tool, and in hands-on in a guided process of using the tool to create a data management plan for a specific granting opportunity). Trainings needed include in-person and online trainings as well as ongoing work to ensure DMPTool training is part of an integrated data management training program from the Libraries, integrated with Library Instruction overall, and integrated with other campus groups involved in data management including the Division of Sponsored Programs.

<sup>29</sup> Discussed in the meeting with DSP, with Emerging Technologies grant proposal targeted for 2014.

<sup>30</sup> Presentation slides: <http://ufdc.ufl.edu/AA00017906/00007/pdf>

<sup>31</sup> Example UF DMPs: [http://ufdc.ufl.edu/contains/?t=%22Data+Management+Plan+\(+DMP+\)%22&f=SU](http://ufdc.ufl.edu/contains/?t=%22Data+Management+Plan+(+DMP+)%22&f=SU)

<sup>32</sup> The DMCTF discussed the DMPTool trainings with the Office of Research and Research Computing. Both will refer and promote this centralized service as provided by the Libraries.

**Charge: Recommend a framework for liaisons and subject specialists to incorporate data instruction and consultation into their workflows**

**Charge: Develop materials and sessions for training of liaisons, subject specialists, and other library staff to prepare them to support campus data management services**

**Charge: Develop training and outreach materials to be used by liaisons, subject specialists, and other library staff in their work with clients**

**Recommendation for a Framework:** In the age of Big Data, the roles for librarians (as functional, technical and subject experts) are being blurred with librarians adding fluency in the other areas to their primary expertise. With broad and deep areas of knowledge, Librarians are expert collaborators for the academic and research community, and libraries serve to support communities of expert collaborators with the space and other resources required to support connecting to information and other resources for research. Data instruction and consultation logically extend from existing expertise and activities.

The DMCTF recommends an integrative framework that recognizes the existing strengths, leveraging and building on known areas of expertise, workflows, activities, products, and capacities. The DMCTF recommends developing this integrative framework by focusing on specific activities (explained in the next sections) to strengthen the foundation for next steps in development and expansion.

**Status for Sessions, Training, and Outreach Materials:** The Data website<sup>33</sup> is a core resource for supporting campus data management services. The DMCTF hosted a number of trainings, workshops, and instruction sessions on various topics (Research Computing, various within the “Big Data, Little Data” event, Zotero training<sup>34</sup>), and created additional data orientation and overview materials.<sup>35</sup>

**Recommendation:** The DMCTF recommends a variety to activities to incorporate data into existing Liaison activities. The DMCTF has created a variety of resources to support these activities.<sup>36</sup>

The DMCTF recommends that all librarians:

- Incorporate the slides (as refined to be most appropriate) on data into all appropriate library presentations and instruction. The DMCTF recommends that all Liaison and Subject Specialists update their LibGuides and appropriate websites to add a tab or page for data with the links and materials below:
  - Data website, <http://library.uflib.ufl.edu/datamgmt/>

---

<sup>33</sup> <http://library.ufl.edu/datamgmt>

<sup>34</sup> Announcement: <http://ufdc.ufl.edu/AA00016055/00001/pdf> ; presentations from “Big Data, Little Data”: <http://ufdc.ufl.edu/AA00017906/> ; Zotero LibGuide: <http://guides.uflib.ufl.edu/profile.php?uid=2284>

<sup>35</sup> Data curation exploration slides: <http://ufdc.ufl.edu/AA00013885/00001/pdf> and short promotional handout: <http://ufdc.ufl.edu/AA00017342/00001/downloads> (PSD file)

<sup>36</sup> For instance, see this draft text on data support and services: <http://ufdc.ufl.edu/AA00019190>. Other resources include standard, template slides and handouts on data management resources from the Libraries and Research Computing, and a guide for supporting data at the reference desk.

- IR@UF, <http://library.ufl.edu/ufir>
- DMPTool
- Research Computing, <http://rc.ufl.edu>
- HiPerGator, <http://www.hpc.ufl.edu/2013/06/hipergator/>
- Appropriate subject repositories (or DataBib for finding resources)
- Additional, appropriate resources for data identified by the Liaisons
- Incorporate links for relevant data management resources into all presentations and trainings, including introductions and overviews, as with the new graduate student and postdoc orientations,<sup>37</sup> as well as into all resources materials, including basic handouts.<sup>38</sup>
- Contact their departments to provide an update on data services from the libraries (draft text<sup>39</sup>)
- Contact their departments to request information and feedback regarding on their small databases (“dinky databases”) with research data (draft text<sup>40</sup>).
  - Following contact regarding the dinky databases, the DMCTF recommends that the DMCTF collaborates with librarians who receive information on dinky databases and data projects to create a data project page, with short entries on a variety of data projects and associated researchers. The simple narrative (not directory listing) information and the data project page<sup>41</sup> would provide an accessible sampler guide to different types of data projects across campus and would provide an accessible entry point for thinking about data projects and will promote understanding of what “data” means for different research fields, options for managing it, and way to connect/provide context for case studies.<sup>42</sup>
- Include and integrate data into existing events whenever possible, explicitly mentioning data support from the libraries and Research Computing resources. In many cases, this may simply require acknowledging that data is already included and integral to the event. For instance, events like various events in InfoCommons, GIS Day, Digital Humanities Day, and other Digital Humanities events<sup>43</sup> may not always mention data explicitly as data, but data is a central concern and part of both.
- Include and integrate data into existing and new courses taught by librarians. The DMCTF will collaborate with the Instruction Committee and individual librarians to add information on data management support from the Libraries and Research Computing to be included in existing courses which could include: GIS, Introduction to Library and Internet Research (IDS4930),

---

<sup>37</sup> [http://guides.uflib.ufl.edu/grad\\_orientation](http://guides.uflib.ufl.edu/grad_orientation) and <http://guides.uflib.ufl.edu/postdoc>

<sup>38</sup> <http://guides.uflib.ufl.edu/content.php?pid=49813&sid=435652>

<sup>39</sup> <http://ufdc.ufl.edu/AA00019190/00001>

<sup>40</sup> Draft Text for Requesting Information on Small or “Dinky” Databases:

<http://ufdc.ufl.edu/AA00014835/00022/pdf>

<sup>41</sup> Possibly like that for DH: <http://cms.uflib.ufl.edu/DigitalHumanities/UFDigitalHumanitiesProjects>

<sup>42</sup> The DMCTF could expand selected entries into case studies to show projects where data is excellently managed, to explain the different aspects of data management with links to resources on campus providing support, etc. This could help to communicate what the technologies do and how the resources connect. This would also support the Data Science Projects course in CSE, which requires a list of projects with half page explanations for students to use in finding data projects to work on for the class.

<sup>43</sup> <http://library.uflib.ufl.edu/DigitalHumanities/DigitalHumanitiesWorkingGroup>

African Studies Bibliography Graduate Course<sup>44</sup>, Anthropology Bibliography Graduate Course Preserving History Undergraduate Internship Course in Archives, etc.

- Pursue new presentation and training opportunities on data in collaboration with resource experts like the *IR@UF* Coordinator, Scholarly Communications Librarian, etc.
- Contact their faculty regarding data-related courses to share data management support resources from the Libraries specifically relevant for the course. For instance, Computer Science Engineering will begin a new Data Science 3 course sequence (each course is 3 credit hours) in spring 2014. The notes towards the sequence currently list the courses as: Elements of Data Science; Advanced Topics in Data Science; Projects in Data Science.<sup>45</sup> In Digital Worlds, an existing course is focused on research computing and data: Digital Worlds: Interdisciplinary Research Seminar, an introduction to research computing (DIG6840C).
- Watch the SobekCM training videos to then support data needs using SobekCM when appropriate. The Smathers Libraries use and support the SobekCM Open Source Software for digital collections, digital libraries, data collections, and data sets. SobekCM's exceptional support for data, user needs, and curator and collection manager needs makes it an excellent choice for data management, and an excellent system to know for comparison when needed.<sup>46</sup>
  - Join the SobekCM user group to best support data, digital collections, and user needs.<sup>47</sup>
- Review the data sets in the *IR@UF* where data sets accompanying electronic theses and dissertations (ETDs) are submitted directly into the *IR@UF's ETD Collection* along with all ETDs.<sup>48</sup>

The DMCTF recommends ongoing work by the DMTCF in collaboration with others to:

- Investigate creating a new data course. An introductory course on data, data science, and research computing (what could be called a course on data literacy) does not yet exist at UF. The Research Computing Advisory Committee (RCAC) has identified that a primary data course is needed.<sup>49</sup> The DMCTF recommends that the DMCTF and Library Instruction Committee collaborate on the possibility of developing and teaching the new data literacy course.<sup>50</sup>
- Host additional data-focused events, ideally with additional data activities included. The DMCTF recommends a new "Elegance of Data" event, possibly a poster session, and possibly modeled in some ways on the prior "Elegance of Science" program. For additional data activities for events

---

<sup>44</sup> <http://guides.uflib.ufl.edu/content.php?pid=6493&sid=1480100>

<sup>45</sup> [http://www.it.ufl.edu/wp-content/uploads/2012/03/RCAC\\_minutes\\_2013Nov4.pdf](http://www.it.ufl.edu/wp-content/uploads/2012/03/RCAC_minutes_2013Nov4.pdf)

<sup>46</sup> SobekCM Training Series: <http://ufdc.ufl.edu/AA00019186/00001/pdf>.

<sup>47</sup> SobekCM User Group: <https://groups.google.com/forum/#!categories/sobekcm-discuss/general>

<sup>48</sup> <http://ufdc.ufl.edu/ufetd>

<sup>49</sup> [http://www.it.ufl.edu/wp-content/uploads/2012/03/RCAC\\_minutes\\_2013Nov4.pdf](http://www.it.ufl.edu/wp-content/uploads/2012/03/RCAC_minutes_2013Nov4.pdf)

<sup>50</sup> This includes consideration on whether a series of online modules or credit course is best, and the process, where it could be that a series of online modules is used to build into a credit course or found sufficient on their own, etc.

like this, the DMCTF recommends specifically including the *IR@UF*, perhaps requiring all participants to submit their materials to the *IR@UF*.<sup>51</sup>

**Future Recommendations:** Pursue new opportunities with any new hires, system integrations, and existing processes for review and evaluation of librarians given added data responsibility needs, including:

- With the new *IR@UF* Coordinator, the DMCTF initiated discussions on *IR@UF* training that present the *IR@UF* within the context of an integrated service. For instance, training on resources from the Libraries for grants could include: grant databases for finding funding opportunities, research skills for developing the proposal, the DMPTool for data plans, responsible conduct in research and ethics, author rights and copyright for publications, the *IR@UF* for grant products (posters, publications, data, etc.), and more.
- The Application Engineers for SobekCM are developing additional data support<sup>52</sup> and will be working to integrate SobekCM's front-end interface with world-class computational power back-end from Research Computing. In the near future, the DMCTF should coordinate scheduling for presentations, facilitated discussions on user needs and design concerns, as well as trainings on data support in SobekCM, with these for internal and then external groups.
- The DMCTF should support librarians and supervisors in considering the role of data for current and expected near-future needs to determine if the annual assignment goals or position description should be revised to include the data-related roles.

### **Charge: Recommend the role of the Institutional Repository (IR@UF) and Research Computing in storing, finding, and accessing working and final data, and linking publications to supporting data**

**Status:** Research Computing is bringing up a new data system, with expected support at end of Q1 2014. The new system will have basic and functional controls, support for access controls and permissions (including supporting UF researchers and their collaborators outside of UF), and other core supports needed for data management.<sup>53</sup> The *IR@UF* will need to be connected to the Research Computing data

---

<sup>51</sup> Other events could include the *IR@UF* to focus on the need and availability to support posters from poster sessions and grey literature. Conference posters are often cited in published articles based on the abstracts alone with no access to the poster; researchers are cited less often with no access to their content from posters; etc. The *IR@UF* component would also include showing all presenters and attendees how to use the *IR@UF* specifically for their posters, offering to meet and assist those with many posters, and offering to visit departments to share on the *IR@UF* as a data management resource.

<sup>52</sup> <http://ufdc.ufl.edu/AA00017907/00001/pdf> ; <http://ufdc.ufl.edu/AA00019155/00001/pdf> ; and data support already in place: <http://ufdc.ufl.edu/AA00017119/00001/allvolumes>

<sup>53</sup> The DMCTF evaluated many Open Source software data packages. None emerged as a clear best or good solution for a majority of needs. SobekCM emerged as a best-fit for some of the identified needs, provided there was additional support for computation and a backend data store, which Research Computing is currently implementing. Additional work will be required to integrate these systems. To identify data needs, the DMCTF evaluated the Dataverse Network (DVN) data repository software with SURASERL. The DVN software is a poor match for current needs. While it has additional functionality, the functionality is implemented in complicated

store, integrating the systems as part of a connected infrastructure. The integrated system will solve many needs, and will inform how to best solve other needs. Already identified needs include enhanced support in SobekCM for provenance to provide the data collection and tracking needed for reproducible research.

Other UF systems also need to be integrated and connected. For instance, the *IR@UF* needs to be integrated with the systems for the Division of Sponsored Programs (DSP) at least for author disambiguation. The Libraries applied for an ORCID implementation grant to add ORCID's unique IDs for authors to the *IR@UF* along with a full authority structure for authors and other entities, which was not funded. The Libraries are now reworking the proposal to focus on the needed authority system work, with plans to submit the proposal as an internal emerging technologies grant in early 2014.

Integrating, connecting, and enhancing core and continuing UF systems is ideal, whenever possible and appropriate. This follows the recommendations for the rest of the report which focus on leveraging and building from existing expertise and capacity.

#### **Recommendation on the role of the *IR@UF*:**

Recommended next steps include author disambiguation and visualization support:

- Submit the grant for name disambiguation support by integrating an authority system in SobekCM (for the *IR@UF* and all UF resources), with future plans for integration with the Division of Sponsored Programs.
- In addition to work to connect the *IR@UF* and Research Computing, and the *IR@UF* and DSP through ORCID IDs, the DMCTF recommends continuing to pursue work on the *IR@UF*, specifically for adding integration with simple visualization APIs where the data would be stored in the *IR@UF*, and users could also display the data as a graph or chart in the *IR@UF* (through the use of the API) without needing to download the data. This will support users for some needs, support users in thinking through how they want to store, access, and use stored data. The DMCTF also recommends, if possible and useful, for the IT for the *IR@UF* to be in contact with Research Computing, ICBR, and/or another central IT group during the simple visualization API connection to seed discussion for connecting the *IR@UF* repository storage with tools on the other central IT group systems to plan how to further expose data in the *IR@UF* for connection for visualization.

The DMCTF should continue ongoing research to identify applicable, tools and systems to connect, interlink, or support separately when appropriate. For this, the DMCTF may work from specific use cases, functional and non-functional requirements specifications, and other information as it applies to operational, planned, and research activities to ensure all recommended technologies are accompanied with as much support as possible for defining best practices, appropriateness or fit-ness for purpose,

---

ways and impacts usability. Further, running and supporting the DVN does not seem to be feasible for UF's planned uses without an additional dedicated person on the DVN alone, and this would not lend itself to leveraging or integration with other systems, and so would be an added cost/requirement.

and more to define when, where, what, and how the technology best supports the academic goals, and how the technology relates and connects to others. With each new planned or available technology, outreach, promotion, and usability testing will be needed.

### **Charge: Develop means to enhance and expand the librarian liaison model with the goal of making librarians partners in research activities**

**Status:** Librarians are unevenly and inconsistently included as partners in research activities. The DMCTF recommends adopting an overall “Library Liaison Team”<sup>54</sup> approach for data and digital scholarship project needs, in addition to integrating support for data workshops within existing liaison activities. For some areas, this may be an integrative approach that builds on and from existing areas of expertise. For instance, the GIS expertise in Libraries can be reframed as visualization (GIS and other) to expand the roles and opportunities for the Libraries in regards to visualization and data overall.

Additional work is needed for enhancing and expanding the librarian liaison model with the goal of making librarians partners in research activities, and further work is needed to ensure all librarians are fully supported for this work.

**Recommendations on Visualization and GIS:** The Libraries should build on existing data visualization expertise and activities in GIS for greater awareness and outreach that extends further into other visualization areas. For this, the GIS Librarian welcomes the opportunity for this expanded role which would include many areas, and those could initially be:

- Providing expert consultation, contributing narrative, and participating as an investigator on grants from across campus
- Developing a list of campus plotter services, for providing referrals with this as a service not intended for support within the Libraries
- Collaborating with the Libraries on the 3D printing service for referral support and any connected technical and research concerns
- Developing and maintaining a list of available visualization tools and supports for awareness and to then develop additional trainings, materials, and activities as based on the available and relevant tools for research needs
- Developing and providing support for additional visualization tools as they become available (researcher requests include technologies for HPC simulation/video rendering, remote sensing, etc.)
- Establishing a regular, open “Visualization Open House” or Visualization Lab” in Marston 308 (or other appropriate space), with set lab hours for specific days/times each week for at least one full semester, with the GIS Librarian holding these open hours (possibly with an expert from Research Computing and other visualization experts, when possible) to provide expertise, hands-on consultation, training, and support

---

<sup>54</sup> See: <http://ufdc.ufl.edu/AA00017119/00021/pdf>

- Promotion for the “Visualization Open House” could include focusing on it as an integrated and connected service. For instance, the Libraries and Research Computing could co-author a short news release explaining how this new service is deeply connected with Libraries and Research Computing for current and future work, with students (and those teaching and needing to support their students) can now access ArcGIS through UF Research Apps, and soon all faculty will have access to ArcGIS through UF Research Apps, which is coming soon, and which the Libraries and Research Computing look forward to supporting in many ways, like the “Visualization Open House.”
- For this, success will be measured in terms of success in promoting awareness on the expertise and support available for visualization and GIS. At the end of the first semester, the “Visualization Open House” or “Lab” will be assessed in relation to the level of promotion, awareness, and outreach. At the end of the first semester, there may be refinement or other methods may be selected for this work. Additionally, if both the GIS Librarian and an expert from Research Computing are available for the open labs, this pilot will be assessed on how well it facilitates and supports further collaboration and community building across the Libraries and Research Computing.

**Recommendations on Grants:** The DMCTF is working to help build a culture of data management, and the grants program is working to build a culture of grantsmanship for the full project, from connecting to collaborators before the project idea through project completion and on to new phases. With many related activities, the DMCTF is in discussion with the Grants Committee for collaborative work on overall shared concerns and specifically for making librarians partners in research activities.

**Charge: Develop recommendations for the Libraries’ campus-level role in support of data management and curation**

**Charge: Propose a corresponding framework and resources for library support of the data life cycle**

**Charge: Propose a corresponding framework and resources for library support of the data life cycle**

**Status:** Faculty, students, and staff from all across UF are in need of data management and curation support now, and will have greater needs in the future. The Libraries must address immediate needs, build towards long-term support, and work towards growing the larger culture to increase the overall institutional capacity. The Libraries already have a campus-level role in data management and curation because the Libraries are campus-level leaders in collaboration, human infrastructure, technical infrastructure for curation, and so much more. The DMCTF has worked to grow and develop this collaborative-connector role in collaboration with Research Computing.

**Recommendation:** The DMCTF recommends that the Libraries continue to strengthen the collaboration with Research Computing, and work to cross-promote Research Computing as a core, library-related resource. Building on this collaboration, the DMCTF recommends growing this collaboration to connect with campus IT groups and campus IT professionals supporting data. For this, the DMCTF recommends targeting specific “best match” groups for shared data needs:<sup>55</sup>

- **Selected IT groups and individuals:** As information connectors, supporters, teachers, and collaborators, academic libraries have a close affinity with their campus counterparts in information technology, as is clearly seen with the collaboration with the Libraries and Research Computing. The collaboration needs to continue on to connect with other groups to grow the culture of and capacity for data management across the groups. This is needed for all groups and areas where data management is done.
- **Selected campus units/groups:** For instance, the Samuel Proctor Oral History Program (currently hiring for a Digital Humanities Academic Production Specialist), ICBR, and the Informatics Institute. Ideally, in selecting campus units/groups, the DMCTF will connect the new groups and also strengthen existing collaborations with other campus groups on data.

## Recommendations for Marketing and Outreach

The DMCTF has additional specific recommendations in regards to marketing and outreach:

### Recommendations:

- Request a link for Data to be added to the UF Smathers Libraries homepage, which is soon to update to the new design (*Requested November 2013*)
- Collaboratively write a brief (1 page) news article with Research Computing for UFIT news
- Collaborate with Research Computing on outreach for resources and services from Research Computing and HiPerGator
- Integrate data with existing events, trainings, and activities, including:
  - Establishing “Visualization Open House” or Visualization Lab” open hours
  - Developing *IR@UF* training that presents the *IR@UF* as an integrated service
  - Updating all LibGuides and appropriate websites to add a tab/page on data resources
  - Integrating data outreach and promotion into existing events (e.g., InfoCommons, GIS Day, Digital Humanities Day, etc.) for promotion and awareness for resources from the Libraries and Research Computing
  - Integrating data into existing and new courses taught by librarians

---

<sup>55</sup> The “best matches” must be involved with data, but also include those in areas of dramatic change who are positioned to enable radical collaboration. The age of Big Data blends previously separated areas for librarians (e.g., functional, technical, and subject expertise) in a parallel manner to the blended expertise necessary for research computing (e.g., systems administration, programming, database administration, experimental processes, concerns and needs specific to academic research as with understanding the difference in acceptable error rates in research versus commercial web searching, and how to support the rigorous requirements for research). The blurred roles and goals for the related “application engineer” roles are also useful in identifying potential “best matches” (see <http://llc.oxfordjournals.org/content/26/2/217.full>).

- Integrating data into reference support
- Integrating data into overall library services and resources promoted by Liaisons