RESEARCH AND INFORMATICS (R&I) DIVISION
UNIVERSITY LIBRARIES
2014-2015 ANNUAL REPORT
(prepared by Julie Speer with content and input from members of the division)

Introduction

The research and informatics division of the Libraries has made great strides in accomplishing many of the strategic goals we set for ourselves in the last 18-month implementation plan period (see Appendix A). Our division-wide strategic goals align with the Libraries objectives to 1) connect library services to - and embed our expertise in - the university's research infrastructure, 2) provide services and expertise to support new research practices and processes, 3) advance university digital curation and preservation activities, 4) increase the volume of Virginia Tech open access intellectual output, and 5) design educational programs that offer researchers pathways for developing competencies in data curation and scholarly communication. Our hope is that the activities described below contribute to the transformation of the library as platform and partner in *Research and Curation*, *Collections Access*, and *Teaching, Learning, and Literacies* thematic areas.

Personnel and Organizational Changes

The research and informatics division of university libraries was created in June 2014 comprised of three new units: Scholarly Communication, Data Curation, and Technology Development. Eight new faculty and staff members joined the team. We transitioned responsibility for research and informatics Mac desktop support and implementation of a shared digital production network storage environment to the Libraries' IT Services department.

Research Environmental Assessment, Policy, and Innovation
(Julie Speer, Gail McMillan, Nathan Hall, Yi Shen, Natsuko Nicholls, Andi Ogier, Anita Walz, Philip Young, Ed Brooks, Shane Coleman, Kiri Goldbeck DeBose, Inga Haugen, Bruce Pencek, and Ginny Pannabecker)

Research environmental assessment was made a priority this year. To design and offer responsive digital data and scholarship services and infrastructures, we designed and completed three data-related assessment projects (one university level, one CNRE college level, and one geospatial data) and planning is underway for two additional projects with CALS and faculty in Public Health. The university level study, "Strategic Planning for a Data-Driven, Shared-Access Research Enterprise: Virginia Tech Research Data Assessment and Landscape Study" received a 25.8% response rate from university faculty and a 64.9% survey completion rate. As part of our CNRE assessment project, we interviewed 15 forestry faculty members about their data practices. The results of these studies have, and will continue to, inform the development of our consultative services and repository systems, as well as prospective new initiatives, through improved understanding of our researchers' datasets and data management practices.

Part of our ongoing effort to assess the university's research landscape involves understanding the factors influencing change. Fair Access to Science and Technology Research Act (FASTR) and the Shared Access Research Ecosystem (SHARE) are examples of responses to one important factor - the demand for public access to federally funded research. Members of the division are actively engaged in national communities like SHARE that seek to provide open, networked, and immediate access to data and information about research products and activities. Locally, and through our faculty members' university service, we've encouraged consideration of policy revisions to reflect changes to the global research environment, such as through the Commission on Research's review of Policy 13015 and VT Intellectual Property (VTIP) Committee's revisions to Policy 13000, and led conversations in the faculty senate about journal impact factors (read more on the Open@VT blog). Our outreach on open issues is contributing to the local research environment in meaningful as well; activities include an invitation to join an Office of the Vice President for Research (OVPR) leadership group meeting to discuss collaborative support of faculty impacted by the National Science Foundation public access policy and the many open knowledge outreach activities described below.

In support of student and faculty innovation, we also initiated a partnership with VTIP, inviting them to hold consulting hours in PORT: Research Commons. Descriptions of this partnership are available in the Roanoke Times, Ubiquitous Librarian blog, and VT News.

R&I Sponsored Research/Innovations. This year, we succeeded in securing $178,214 in external funding to support library research and innovations in digital preservation, digital resource development, and repository software development. We also secured resources in the form of storage allocations for digital library development through an XSEDE grant.

- "ETDPlus: Preserving & Curating ETD Research Data & Complex Digital Objects." (Institute of Museum and Library Services, National Leadership Grant for Libraries, $250,000) 2014-2016. Educopia Institute, University of North Texas, University of Tennessee, Penn State University, Purdue University, Virginia Tech. Principal Investigator, Katherine Skinner. Gail McMillan, Zhiwu Xie, Advisory Board. Julie Speer, Virginia Tech Principal Investigator for $126,214 Curation Workbench Subcontract.

- "Development of an Interactive Human Body Digital Reusable Learning Object (RLO) to Provide Whole Body Systems-based Learning in Vitamins and Minerals." (4VA, $22,000) 2015-2016. Co-Principal Investigators, Anita Walz, Lujean Baab, Deborah Good.

- "Archiving Transactions Towards Uninterruptible Web Service." (Columbia University Libraries/Information Services Mellon Grant for Collaborations in Web Archiving Web Incentive Award, $25,000) 2014-2015. Zhiwu Xie, Principal Investigator. Edward Fox, Co-Principal Investigator.

- Amazon Teaching Grant. (Amazon Web Services usage credit, $5,000) 2015-2016. Zhiwu Xie, Principal Investigator.

- XSEDE Startup Grant for the Goodwin Hall Live Lab project. (XSEDE: Initial allocation: TACC Stampede 25,000.0 SUs and TACC Ranch 500.0 GB) 2015-2016. Zhiwu Xie, Principal Investigator.

- Several members of the division were involved in the pursuit of external funding for collaborative research projects as principal investigator, co-investigator, senior investigator, advisor, or key staff/personnel. Two library-led proposals were submitted to IMLS (one unsuccessful, one still under review), and we contributed to a successful NEH CLAHS digital humanities proposal, a successful NHPRC CLAHS digital history proposal, and an unsuccessful CALS proposal.

Virginia Tech's Research Information Management Network
(Julie Speer, Andi Ogier, Keith Battleson, Peggy Layne, Daniel Hung, Ken McCrery, Randy Crockett, Beth Tranter, Matthew Swift, Zhiwu Xie, Paul Mather, Chreston Miller, John Borwick)

Through a partnership with the Office of the Senior Vice President and Provost, Institutional Research, OVPR, and IT, we contributed to development of an integrated faculty activity reporting, repository, and researcher profile service infrastructure. These and other IT partnerships described below document related activities this year.

Symplectic Elements Implementation. Elements will soon be the new faculty activity reporting system replacing the existing Digital Measures (DM) system. We served on the implementation team, provided general application and technical support, and migrated legacy DM data into Elements. The next phase of implementation is Elements integration with VTechWorks and VIVO. The Provost Office will invite 3 colleges/departments to use the system in fall 2015.

VIVO Implementations. The university's VIVO pilot project is 100% complete and we are actively developing a project plan for implementation. We offered university stakeholders a demonstration of the VIVO system; with their support, we began exploring VIVO as VT's researcher networking platform. We successfully installed the Elements VIVO Harvester and tested performance with positive results. Through our university service and in partnership with OVPR, we also began participating as one of 5 institutions involved in a VIVO-based Southeastern Universities Research Association federated research expertise database project.

Networked Research IT Infrastructure. We contributed to the development of an Advanced Research Computing NSF Campus Cyberinfrastructure grant proposal, "CC*DNI Networking Infrastructure: A Campus Research Network and Distributed Science DMZ" that, if funded, would offer advanced infrastructures to facilitate data transfer into library repository and curation environments.

ORCID Service. In spring, we signed an ORCID agreement to create, and upon an individual's approval, establish ORCID profiles for VT researchers. We made progress developing a

campus communication plan; although an email service issue has delayed progress with technical implementation.

Outreach and Communication on Data and Scholarly Communication
(Julie Speer, Gail McMillan, Natsuko Nicholls, Yi Shen, Andi Ogier, Anita Walz, Ginny Pannabecker, Charla Gilbert, Debbie Cash, Inga Haugen, Philip Young, Nathan Hall, Patrick Tomlin, Kimberli Weeks, Chreston Miller, and liaison librarians)

Outreach about our data and scholarly communication services, while a core activity, was made a priority this year to address a general lack of awareness among researchers about recently introduced, or augmented, data, repository, and publishing services. We created three new libguides on Research Data Management (an almost completely revised DM guide), currently with over 420 pageviews (an increase from 8 views in just 15 weeks), a National Science Foundation requirements guide, with over 630 pageviews (an increase from 10 views in 15 weeks), and a National Institutes of Health requirements guide, with over 180 pageviews (an increase from 7 views in 15 weeks), and further development of our Open Educational Resources guide to provide links to discipline-specific resources. These curated guides helped instructors across the libraries supplement face-to-face workshops, NLI sessions, as well as course-embedded instructional sessions.

We also created promotional materials to help individuals with outreach responsibilities advertise our services and expertise. This includes a data services handout distributed to OVPR Proposal Development Institute's (PDI) participants and by the Office of Sponsored Programs Pre-Award unit to faculty seeking NSF CAREER Grants, as well as a data curation and a scholarly communication handouts to assist liaisons with the scholarly communication component of their responsibilities. To further support liaisons and share division-wide updates, we established a monthly research and informatics and liaisons meeting.

Through a partnership with Planning and Branch Operations, we presented on research and informatics services and projects to Sanjay Raman (Associate Vice President for the National Capital Region) and Nicholas Stone (Director, NCR Operations) during a visit to the university's Northern Virginia Research Center in Arlington; this activity resulted in several new opportunities for the Libraries, including increased virtual participation of NCR in data literacy educational sessions.

Our research environmental assessment efforts led to opportunities to raise awareness of our data services and repository planning among the campus community, including a discussion with Srinath Ekkad (Associate Vice President for Research Programs) and a project update presented to the Commission on Research.

To raise awareness among the College of Engineering of our new data and informatics services, we met with department heads in Engineering, attended a Computer Science department meeting, and reached out to Mining and Mineral Engineering, Mechanical

Engineering, Civil Engineering, Materials Science and Engineering, and Biological Systems Engineering.

Promotion among the campus community of new open access services and open publishing opportunities also broadened this year through R&I public programming activities. Research and informatics members led and continued to serve on the Libraries' Open Knowledge Committee, which serves as the coordinating body for Open Data Day, Open Education Week, and Open Access Week programming activities for the university community.

- Keynote speaker for this year's Open Access Week (Oct. 20-24) celebration was Brian Nosek (co-founder of the Center for Open Science and professor of psychology at UVa). We offered eight events during the week, including a panel discussion on open access by Virginia Tech faculty and graduate students, a workshop on reproducible research, and sessions on visual literacy and copyright, VT's Open Access Subvention Fund, Creative Commons licenses, author rights, and trends in scholarly publishing. All events were available for NLI credit.

- The Libraries hosted a kick-off event as part of the first International Open Data Day at Virginia Tech featuring open government data advocate Waldo Jaquith. Open Data Day, a collaboration with the local Code for NRV, was well attended, with approximately 15-20 people present in one or more sessions. As part of the celebration, we hosted several roundtable sessions on open data topics including mapping data (led by VT's Center for Geospatial Information Technology), public policy data (led by a member of the Blacksburg Town Council), journalism data (led by representatives from Educational Media Company at Virginia Tech, Inc. and the Collegiate Times). A report on the Open Data Day celebration is available on the Open@VT blog and one of the hackathon projects was featured in a news article.

- In partnership with Technology-enhanced Learning and Online Strategies, Center for Instructional Development and Educational Research (CIDER), and the Student Government Association, we also coordinated an Open Education Week event February 23-27 to raise awareness of Open Education and open licensing concepts. Events included a 'Get Creative (and stay legal): Copyright Compliance with Creative Commons and OER' session, an undergraduate student forum and panel discussion (led by Student Government Association's Academic Affairs Committee), a faculty and graduate student panel discussion, an OER for librarians session, an OER for instructional designers session, and a workshop led by Kristi Jensen and David Ernst from University of Minnesota on faculty open textbook adoption. Panelists and speakers included faculty and students at Virginia Tech, Executive Director of CIDER, and speakers from Virginia Military Institute. Participants included undergraduate students, graduate students, faculty, and OER authors and adopters. A report on the Open Education Week program is available on the Open@VT Blog.

- In partnership with the Graduate School, we offered two student travel scholarships to attend OpenCon, an early career researcher conference on open access, open data, and open educational resources. A master's student in public health and a doctoral student in computer science were selected to attend OpenCon, which took place in Washington, D.C. November 15-17. A report is available on the Open@VT blog.

- Our outreach was also visible through Open@VT, the OpenVT email list, and @openatvt, where we promoted open knowledge events, documented our activities, and helped to create a culture of openness at Virginia Tech. The Open@VT blog has been visited over 3,500 times, with a daily high of 405 views. We've made over 1,400 @openatvt tweets and gained 190 followers, including VT President Sands.

New Literacies Development and Instruction on Data and Scholarly Communication
(Gail McMillan, Andi Ogier, Kiri Goldbeck DeBose, Inga Haugen, Ginny Pannabecker, Natsuko Nicholls, Anita Walz, Philip Young, Shane Coleman, Ed Brooks, Chreston Miller, Zhiwu Xie, Nathan Hall, Anne Lawrence, and liaison librarians)

Our instructional sessions increased this year (see Appendix B) as we gained personnel with teaching responsibilities. We offered 18 NLI sessions on various scholarly communication, data curation and digital libraries topics, 12 course-embedded lectures on various data and scholarly communication topics, 11 workshop sessions, 10 invited lectures, and coordinated 3 workshops (including a scientific writing workshop created in partnership with 15 faculty and offered to over 100 graduate students, faculty, staff, post-docs, research technicians, and visiting scientists).

Data curation and scholarly communication drafted tiered curricula, designing educational programs that offer introductory, intermediate, and advanced pathways for developing competencies in data and scholarly communication-related issues; the goal is to create content that can be embedded in various disciplines, majors, and courses and offered to either faculty, staff, undergraduate, or graduate students. Making use of these curricula is already happening in many cases, but we hope to refine them in partnership with other library units and explore more systematic adoption of these guiding materials in our course/session design and delivery in the coming year.

We also hosted the 3rd annual Data Management Bootcamp in partnership with the University of Virginia, Old Dominion University, James Madison University, George Mason University, Virginia Commonwealth University, and the College of William and Mary. Participants were offered sessions on the following topics: introduction to data management, role of data management in research, archiving and preserving research data, Creative Commons and fair use, finding data, documentation and metadata, intellectual property, data wrangling, organizing files, version control and cloud storage, and database design and querying. For the first time this year, a room in the Falls Church Center was added as a satellite location.

At the suggestion of the Associate Vice President for Research Programs, we also began developing a content outline for web-based data management training modules for graduate

students. We hope to turn this content into video recordings for publication on the OVPR CITI training site as optional learning content for graduate students.

Digital Library Development
(Zhiwu Xie, Tingting Jiang, Yinlin Chen, Collin Brittle, Andi Ogier, Shane Coleman, Natsuko Nicholls, Chris Klein, Kimberli Weeks, Paul Mather, Chreston Miller, Kiri Goldbeck DeBose)

A priority for us this year has been reinforcing our core digital library infrastructure to support digital curation and documentation of more complex digital objects and their relationships, as well as developing our research partnerships through informatics projects. Our informatics innovations and core infrastructure developments have largely centered on Hydra+Fedora digital libraries and data extraction activities for Smart Infrastructure Laboratory (VT-SIL) projects such as the Building Occupant Classification project.

- A goal this year was to develop a data repository prototype, which we developed once a development environment was created and production versions of Fedora 4 and Sufia (Hydra) were released. We set up a Github code repository and provided stakeholder demonstrations for phase 1 of prototype development - automatic DOI assignment, CAS authentication and authorization integration, automatic ingest of DataCite metadata, and ORCID service integration. Stakeholders helped to identify gaps in service and supplied deposit workflow considerations and deposit agreement language. We based the 'big data management system' prototype on the VT-SIL Goodwin Hall project and completed performance testing with positive results. We anticipate meeting our target of testing deposit of 100 datasets by Jan 2016.

- Our graduate students contributed significantly to development projects this year. One completed the initial coding for the FishTraits Geospatial feature development (see Github code repository) (our software engineers refactored the code and rebuilt the web site) and helped to refine the VTWise search and discovery tool to include more legacy digital collections content. A second completed a working model for the Uninterruptible Web Service project and continues to address performance issues and debug, helping us to reach our goals for the project.

- We developed data collection software for the VT-SIL Goodwin Hall project, enabling the research team to retrieve raw sensor data from building data acquisition units. We maintain a Github code repository for the software.

- Our software development IT infrastructure improved this year through 1) secure storage and compute resources from the National Data Services' (NDS) National Center for Supercomputing Applications (NCSA) and Texas Advanced Computing Center (TACC), 2) VT's Advanced Research Computing (ARC) as part of the VT-SIL project, and 3) cloud computing resources through an Amazon Education Grant.

- This year we set out to automate server deployment and software development environments as much as possible. As technologies change and we seek to align our environments with those preferred for production, we transitioned from using Puppet and Chef for server configuration management to Ansible. We're also learning more about incorporating Docker for application deployment into our systems engineering environments. We began using Jira as a ticketing system for development and Confluence for managing communications with project stakeholders. We'll begin phasing out use of the locally-hosted Redmine system for new projects in favor of these university IT-supported solutions.

- We actively participated in the Fedora open source software development community through contributions to the Fedora 4 production release and maintenance and through Fedora committee membership.

Consultative and Digital Curation Services

Open Access Services. In addition to our efforts to raise awareness of open access, open data, and open education issues and resources, we offer services that support faculty, staff, and students interested in making their digital scholarship freely available online. (*Gail McMillan, Philip Young)*

- With support from Office of the Senior Vice President and Provost and Collections and Technical Services, we were able to award $49,711 to subsidize open access fees for 48 approved scholarly, peer reviewed open access articles as part of the library-managed Open Access Subvention Fund (OASF) service. The funds were used to support 109 unique VT authors in CALS, COE, COS, CLAHS, CNRE, VetMed, and Fralin, VBI, and VTCRI research institutes. We celebrated OASF recipients during VT Authors Day, preparing slides for recognition during the event and contributed to decisions to fund OA initiatives that include Open Library of the Humanities and PeerJ. Our Creative Services team published an article June 18, in VT News, which included a video and was picked up by ARL on June 29.

Publishing Services. Our journal and conference proceedings publishing services expanded this year as we drafted of our first Journal Publishing Policies, welcomed partnerships with Continuing and Professional Education, and developed software to enable integration with external services such as EZID. *(Gail McMillan, Chase Dooley, Tingting Jiang, Jane Wills)*

- Our legacy journal publishing operations included 34 new journal issues and 520 articles posted. We reached out to current journal editors of 6 legacy journals to inform them of new publishing services.

- We made technical developments that improve the user experience and provide persistent links to digital content: 1) developing a theme plugin for OJS and making journal websites responsive to different devices; 2) developing a new workflow

leveraging the use of Jekyll, a static HTML generator built in Ruby, in combination with Python scripts and a Python GUI built with Kivy to reduce the time spent and errors during journal markup; and 3) developing an EZID DOI plugin for OJS ([Github code repository](#)), and contributing the code back to the OJS open source software community.

- We began publishing a new journal: [SPECTRA](#) (migrated from Wordpress); and were approached by Vinodh Venkatesh, CLAHS, for a new Spanish language and literatures journal, by *The Azalean* for the Azalea Society of America, and by Ruth Waalkes, Center for the Arts, to bring the journal [PUBLIC](#) to Virginia Tech from Syracuse.

- We also explored new publishing service partnerships this year, as we met with Erich Sawyer (Continuing and Professional Education) and presented Open Conference Systems (OCS) software to CPE as well as their clients. We supported 4 new conferences: ASPECT Graduate Conference: CLAHS March 20-21, Service Research & Innovation Institute, College of Business, June 7-10, International Conference on Services Management, CLAHS, Nov. 28-30, and Innovative Library Classroom, Libraries.

Repository Services. Central to our digital curation activities is the library-managed institutional repository service, VTechWorks. *(Gail McMillan, Nathan Hall, Melissa Lohrey, Anne Lawrence, Keith Gilbertson, Aaron Hunnewell, Jana Doyle, Mary Finn, Shane Coleman, Chris Klein, Amanda French, Jane Wills)*

- This year, the repository grew by 3,800 to 41,631 items. New repository collections include: Center for Refugee Studies Occasional Paper Series, Online Masters of Agriculture and Life Sciences' masters projects, articles by VT authors harvested programmatically from Sherpa Romeo and Web of Science, and several archived conference proceedings (North American Wind Energy Academy 2015 Symposium, 7th Symposium on Pavement Surface Characteristics – SURF 2012, Marine Energy Technology Symposium, and Veterans in Society: Changing the Discourse).

- Statistics demonstrate a 220% increase in usage (over 5.8 million item downloads) compared to last year. We received 6,425,226 bitstream views, 2,400,081 item views, and 1,210,252 searches were performed March 31, 2014 - May 22, 2015.

- We drafted the first set of VTechWorks policy and procedures guidelines and guidelines for descriptive metadata for student assistants involved in metadata creation. Sharing the [VTechWorks collection policy](#) online this year was also an important step towards transparency of repository practice and service.

- We provided guidance and support to other library units producing content for repository deposit: 1) helping to improve video production workflows to facilitate ease of repository deposit for our Event Capture team; and 2) helping to coordinate digitization operations to improve access to and preservation of bound theses and dissertations (BTDs).

- System functionality improvements were made through a DSpace software upgrade to v.4, which allowed new customizations to the interface and controlled vocabularies for certain fields. We continue to use [Github as a code repository](#) for VTechWorks development.

- Virginia Tech University Libraries partnered with Indiana University Libraries and University of Illinois Urbana-Champaign Libraries for the 10th International Conference on Open Repositories. Members of the division and the VTechWorks team served as volunteers and on the hosting organizing committee. We also hosted the [website](#) for the conference with support from IT Services and the Creative Services team.

Data Services. Our active data management and data consultative services and infrastructures expanded this year. Several library and university cross-departmental working groups and committees contributed to this effort, including the Geospatial Metadata Working Group, the Purchased Data Working Group, and the Data Literacy and Consulting Working Group *(Andi Ogier, Natsuko Nicholls, Ed Brooks, Shane Coleman, Ellen Krupar, Ed Lener, Chreston Miller, Seth Perry (OGIS), Jason Shelton (Facilities), John McGee (Forestry Resources and Environmental Conservation), Bruce Pencek, Robert Sebek, Annette Bailey, Ladd Brown, Melissa Lohrey, Kiri Goldbeck DeBose)*

- Data consults (30 total) included 8 data management plan consults, 2 returned/revised data management plan consults, 5 data management workflow consults, 9 engineering data/informatics consults with faculty/undergraduate/graduate students, and 6 DOI/deposit consults. Our engineering data support also involved interactions with the Discovery Analytics Center and labs in the department of computer science.

- We assisted faculty with active data management, including best practices for implementing data management plans and depositing datasets in VTechWorks (15 total); we're also currently exploring support for approximately 17TB of geospatial data from CNRE. Responsibility for the EZID DOI service transferred to the new Data Curation unit. Data curation and scholarly communication units also collaborated to develop required VTechWorks metadata for scientific datasets.

- Geospatial data consults included 100 unique consultation requests from faculty, staff, students, and members of state and national communities on data discovery (84 consults), accessing data (7 consults), and training on GIS services (14 consults). We acquired and served 344GB of geospatial data to the VT community. In close collaboration with Enterprise GIS, we led a proof-of-concept Hydra-based GeoBlacklight portal development project to design a search and discovery system for distributed collections of Virginia Tech geospatial datasets. As part of this effort, the Geospatial Metadata Working Group explored the creation of a VT geospatial metadata profile based on geographic metadata standards ISO 19115 and ISO/TS 19139.

<u>Digital Preservation Services</u>. Our digital preservation service activities this year included library-wide preservation service planning and digital curation/preservation infrastructure development and management. *(Gail McMillan, Nathan Hall, Adrienne Serra, Kimberli Weeks, Paul Mather, Keith Battleson, Collin Brittle)*

- We continued to manage MetaArchive production and test servers (we host 486 archival units (AUs) at 910GB, including 140 VT AUs at 166GB) and the Lots of Copies Keep Stuff Safe (LOCKSS) server (added 4000 AUs for LOCKSS, bringing the total to 44,000 AUs at 6TB). Our collaboration with LOCKSS led to development of a plugin for filtering DSpace to ingest new items. MetaArchive offers a distributed preservation strategy for content locally curated by University Libraries.

- Our involvement in the Academic Preservation Trust continued through service on planning-phase Content Advisory and Technical Groups and systems engineering support for the Amazon Web Services software development environment. We created an initial server environment for Hydra+Fedora for APTrust developers and worked closely with the APTrust Technical lead to prepare the staging area for partners to deposit content as part of initial testing. We also refactored APTrust's authentication and authorization code, paving the way for our own Sufia (Hydra) authentication and authorization.

- As part of the Libraries' cyberinfrastructure planning efforts, we continued with Dark Archive implementation. The purpose of the Dark Archive is to host files solely for preservation purposes rather than for storage or access to user available files. These preservation--ready files will be preserved through the MetaArchive Cooperative. A first draft of policies and procedures has been developed.

- Recognizing the need for shared storage for digital curation, the division purchased a server for 'working files' management for R&I and Special Collections digital production activities. We solicited stakeholder requirements for the environment and setup and configured a test server to meet user needs. IT Services assumed responsibility for implementation in spring 2015.

<u>Digital Project Collaborations and Support</u>. We collaborated on, and supported, several digital projects this year through team-based consultations in digital scholarship, copyright, metadata, active data management, and informatics. *(Andi Ogier, Shane Coleman, Ed Brooks, Steve Tatum, Bruce Pencek, Gail McMillan, Kimberli Weeks, Ross Edmonds, Kiri Goldbeck DeBose, Zhiwu Xie)*

- Consults and technical support for faculty research projects included 1) advising on database development and web design for Tom Ewing's undergraduate tuberculosis citizen science project (later became an NEH proposal); 2) advising on project sustainability and repository considerations for Philip Radke's TreeData project; 3) consulting on repository options for Virginia Tech Transportation Institute's National

Transportation Safety Board dataset; 4) transforming data for Celia Silverstein's insect species project; 5) setting up eXist-db and PyBossa implementations and supporting current version of Lord Byron and His Times for David Radcliffe; 6) providing Omeka support, copyright and data expertise for Paul Quigley's Mapping the 4th project; 7) participating in project planning and contributing to repository systems analysis (VITAL, Omeka) for Bob Leonard's VTArtWorks project; and 8) coordinating transfer of the Center for Digital Discourse and Culture's April 16 Omeka Archive to University Libraries.

Conclusion

This year we achieved many of our July 2014-Dec 2015 18-month implementation plan goals. Summarized below is a reflection on related outcomes:

- Research environmental assessment: Assessing the research landscape through 3 studies resulted in a better understanding of faculty data management practices and data characteristics. We came very close to meeting our target of a 70% survey completion rate (64.9%) and exceeded our target of interviewing 10 faculty members. Documentation of a subset of our plans is available in the repository and repository requirements elicited from the studies were communicated to the data repository development team.

- Outreach and communication of data curation and scholarly communication services: Libguides development and an increase in instructional sessions and public programs resulted in greater awareness of our services and consultative expertise, as demonstrated by growth in Libguides usage, consultation requests, and data and digital scholarship collaborations.

- Advancing data curation services: Service policies development and improved documentation of data consults and data management plan (DMP) workflows helped to advance the Libraries' data curation services this year. We also created and are incorporating data deposit workflows into the new repository system and pursued strengthening of our geospatial data support through a collaborative data discovery project. Our hope is to make our service and archiving policies available online in the coming year.

- Advancing repository and publishing services: The expected outcome of VTechWorks activities was increased number of materials in the repository. We achieved this target generally and may still reach a 20% increase by 2016. We supported 4 conferences with OCS and expect to communicate archived proceedings in the next reporting period. Preliminary dark archive policies were developed and DSpace was installed for development purposes. We published one new OJS-based journal and communicated the opportunities presented by migration to an OJS platform to existing journal editors.

- Digital library development: Our efforts this year resulted in prototype development of a data management Hydra head over a Fedora repository system. The data repository team solicited feedback from internal stakeholders and incorporated these recommendations into repository design. Our plan is to involve faculty members as external stakeholders in the fall, when testing of deposit workflows will begin. Documentation is available in several Github repositories.

- Data and scholarly communication curricula development: We achieved our measured target of creating curricula for new data and scholarly communication literacies. This involved developing learning outcomes for tiered educational opportunities for faculty, staff, and students. We increased our instruction and educational programs on these topics and plan to put these curricula into action in the coming year.

- Advancing VT's research information management network: We achieved our target of completing the VIVO pilot and exceeded our target of 50 profiles, though we hope to involve more faculty stakeholders in discussions about policy and usage next year. Exposing the data to the national network will occur when the system is in production. Elements implementation is approximately 60% complete with publications, teaching and grants data now in the system. Repository and VIVO integration are planned activities; meeting the Jan 2016 deadline is expected. User training will begin in fall 2015. The Elements VIVO Harvester has been tested and the production Elements UI will be shared by Jan 2016.

Appendix A

RESEARCH AND INFORMATICS
18-MONTH IMPLEMENTATION PLAN GOALS
July 2014-December 2015

Goal: Research Environmental Assessment

    Theme 2, Objective 1

    <u>Outcome</u>: Services and infrastructures developed and refined based on research environmental assessment (data survey, focus groups and interviews with faculty and researchers)
    Measure: Number of departments, colleges, research institutes surveyed; number of researchers surveyed/interviewed
    Target: 70% survey completion rate among VT researchers; one research center studied; 10 researchers interviewed
    Documentation: Assessment plan, documentation of outcomes and research areas represented, requirements gathered for data services to support research faculty

Goal: Outreach and Communication of Data Curation and Scholarly Communication Services

    Theme 2, Objective 1 and 2

    <u>Outcome</u>: Greater awareness among the campus community of data curation and scholarly communication services through developed/refined libguides and web pages and increased number of instructional sessions and workshops
    Measure: Increased consults, instruction participants, libguides usage, VTechWorks dataset deposits, data-related research collaborations, digital scholarship collaborations, workshops, presentations, Open Access Subvention Fund (OASF) requests
    Target: 10% increase in metrics
    Documentation: library web pages, published libguides, instruction and workshop documentation, consults and collaborations documentation

Goal: Advancing Data Curation Services

    Theme 2, Objectives 2 and 3

    <u>Outcome</u>: Defined data curation service model
    Measure: Defined scope of service and documented data curation services, policies, and workflows; campus partnerships explored
    Target: Improved documentation of Data Management Plan support, library geospatial data services, repository data deposit processes, data curation processes, and data

transformation requests; customized DMPTool; published service model framework and research data archiving policies
Documentation: library web pages, research data archiving policies, DMPTool, VTechWorks, service activity documentation

Goal: Advancing Repository and Publishing Services

Theme 2, Objectives 2, 3, and 4; Theme 4, Objective 2

Outcome 1: Increased number of materials archived in the repository
Measure: Increased number of archived conference proceedings, technical reports, OASF funded articles, digitized theses and dissertations
Target: 20% increase in open works in the repository (to 50,000 items), 3000 digitized theses and dissertations
Documentation: VTechWorks, service activity documentation

Outcome 2: Operationalized Dark Archive repository
Measure: Preliminary policies and procedures drafted and circulated for input to internal library stakeholders, DSpace installed and maintained, Special Collections master files and Event Capture master files deposited, initial harvest into the MetaArchive Cooperative
Target: Functional Dark Archive repository, 200 files deposited
Documentation: Dark Archive repository UI, policies, service activity documentation

Outcome 3: Expanded publishing services
Measure: Monograph publishing explored, increased number of journals using Open Journal Systems (OJS) platform; increased number of conferences supported by Open Conference Systems (OCS) platform
Target: 200 existing journal issues migrated to OJS, 2 new OJS-based journals
Documentation: journal and conference websites, service activity documentation, VTechWorks

Goal: Digital Library Development

Theme 2, Objectives 1, 2, and 3

Outcome 1: Prototype data management repository developed based on the hydra+fedora framework (also based on use cases Signature Engineering Building project and/or OIRED project)
Measure: Internal and external stakeholder requirements solicited (research environmental assessment, service model planning/development), incorporated, prototype tested, sample research datasets deposited, preliminary data archiving policies developed
Target: Functional and pilot tested hydra+fedora repository framework, 100 datasets deposited
Documentation: Github code repository, repository UI, research data archiving policies

Outcome 2: Prototype web archiving digital library developed that integrates SiteStory with Apache web server to act as a server crashing prevention tool
Measure: Satisfied granting agency requirements, software developed, tested, openly released, and results disseminated
Target: Functional prototype, feasibility testing of large web operation complete
Documentation: Github code repository, user manual, research paper/white paper, final report

Goal: Data and Scholarly Communication Curriculum Development

Theme 2, Objective 4; Theme 3, Objective 2

Outcome 1: Developed and improved scholarly communication literacy instruction
Measure: curriculum developed and learning outcomes created; increased number of intellectual property, copyright and scholarly communication instructional sessions, workshops and consults
Target: 10% increase in number of instructional sessions, workshops, and consults
Documentation: scholarly communication curriculum, service activity documentation

Outcome 2: Developed and improved data literacy instruction
Measure: scaffolded framework for data literacy developed and learning outcomes created; toolkit of data literacy instructional content developed; increased number of workshops and partnerships
Target: 10% increase in number of workshops; 5% increase in number of partnerships
Documentation: toolkit, data literacy curriculum, service activity documentation

Goal: Advancing VT's Research Information Management Network

Theme 2, Objective 1

Outcome 1: Completed VIVO pilot project
Measure: Sample researcher profiles in test instance of VIVO, data exposed to the national network, feedback solicited from stakeholders
Target: 50 profiles created for researchers in arts, humanities, science, and engineering disciplines
Documentation: VIVO UI, assessment plan, documentation of outcomes and research areas represented

Outcome 2: Completed Symplectic Elements implementation
Measure: Synchronizer for publications, grants, and teaching data feeds working, repository ingest tested, VIVO harvester tested, user training initiated
Target: profiles created for all faculty members, 50 papers deposited in VTechWorks, 50 profiles harvested into VIVO, training completed for 10 departments
Documentation: Elements UI, VIVO UI

Appendix B

Selected Data and Scholarly Communication Instructional Activities

- 18 sessions to faculty, staff, and students through the Networked Learning Initiatives (NLI):
    - Data Management for Engineering (2 sessions) (NLI): Nicholls, C. Miller. Feb. 2015 and April 2015.
    - National Science Foundation (NSF) Data Management Plan Workshop (2 sessions) (NLI): Ogier, Nicholls. Sept. 2014 and Mar. 2015.
    - How to Create a National Institutes of Health (NIH) Data Sharing Plan (2 sessions) (NLI): Nicholls, Pannabecker. Oct. 2014 and Mar. 2015.
    - Is it a Fair Use? A Hands-On Discussion (NLI): Pannabecker, Walz. Mar. 2015.
    - Journal Impact Factors (NLI): Young, Mar. 2015.
    - GIS: Institutional Support Resources for Geospatial Teaching and Research at Virginia Tech (Seth Perry): Coleman, Brooks (guest lecturers). Mar. 2015.
    - Scholarly Publishing Trends (2 sessions) (NLI): Young, Oct. 2014 and Feb. 2015.
    - Receive $1500 for Publishing Your Next Article: VT's Open Access Subvention Fund (NLI): McMillan. Oct. 2014.
    - Data Visualization (2 sessions) (NLI): Ogier, C. Miller. Fall 2014.
    - Optimize Access to Your Published Research: Author Rights for Life & Health Sciences (NLI): Pannabecker. Fall 2014.
    - Data Management: Deep Dive (NLI): Xie. Fall 2014.
    - Introduction to Cloud Computing (NLI): Xie. Fall 2014.
    - Transforming Research, Teaching, & Learning through the New University Libraries (NLI): McMillan, Pressley, Miller, Ogier; June 16-17 2014.
- Over 12 course-embedded sessions address research, scholarly communication and data related topics:
    - Engineering Education ENGE 5704 (Dr. Stephanie Adams): Copyright in Research and Scholarship: McMillan. May 2015.
    - Sustainable Biomaterials graduate seminar: IP, Patents and Copyright: Walz. April 2015.
    - English 1106: How to successfully include fieldwork data in research paper (Dr. Natalie Richoux): Nicholls. Mar. 2015.
    - Environmental informatics (Dr. Randy Wynne): Intro to Data Management for Remote Sensing: Ogier, DeBose. Feb. 2015.
    - CNRE, Forest Resources and Environmental Conservation Department, Urban Forestry/Urban Horticulture graduate course (Dr. Susan Day): Data in Scholarship: Nicholls, DeBose. Jan. 2015.
    - Environmental informatics (Dr. Evan Brooks): World of Data; Data Modeling, multi-session: Ogier, DeBose, Ed Brooks. Sept./Oct. 2014.
    - Environmental research design (Dr. Elizabeth Grant): Copyright, creative commons, and fair use. Ogier, Walz. Sept. 2014.

- ○ English Graduate Assistants' seminar (Dr. Charlene Eska): Copyright and Fair Use: McMillan. Sept. 2014.
  - ○ Searching strategies, critical analysis of information, and information ethics sessions taught to various classes: statistics FYE. Ogier, Walz. Sept. 2014.
  - ○ AAEC 5104, BCHM 1014 (FYE), BIOL 1004, BIOL 4404, BIOL 4764, BIOL 5174, Life Sciences Learning Community, PSYC 2094, PSYC 4984, Public Health Masters Program: Pannabecker.
  - ○ CS 4624: Software development using regular expressions to parse a bibliography and transform it into a file that could be uploaded to VTechWorks: Hall, Lawrence.
- ● 10 invited lectures:
  - ○ Undergraduate Research Program (Keri Swaby): Understanding data basics and ethical use of data. Nicholls. June 2015.
  - ○ English Department instructor forum: Open Textbooks and Beyond: Access, Affordability, and Academic Success: Walz. April 2015.
  - ○ Undergraduate Research Program (Keri Swaby): Big Data. Ogier. June 2014.
  - ○ Copyright, Open Access, and Ethical Use of Data for College of Business Graduate Assistants, McMillan organized presentations by Walz, Ogier, Young. Aug. 2014.
  - ○ TLOS VT faculty session (Aaron Bond): Finding Resources for Your Flipped Classroom: Walz. May 2014.
  - ○ TLOS VT faculty session for Northern Virginia teaching cohort (Lujean Baab): Finding Resources for Your Flipped Classroom (online): Walz. May 2014.
  - ○ TLOS VT faculty Summer Scholar Series (D. Thacker): Copyright and Online Instruction: Walz. May 2014.
  - ○ TLOS VT faculty orientation (Shelli Fowler): Get Creative (and stay legal): Copyright Compliance with Creative Commons and Open Educational Resources: Aug. 2014.
- ● 11 workshop sessions:
  - ○ Advanced Research Skills Certificate Program: Becoming a Researcher, co-instructors: Pannabecker, Swaby. Jan. 21, 2015; Understanding Data, Nicholls. Feb. 4, 2015; Writing successful conference proposals and research abstracts, DeBose. Mar. 4, 2015.
  - ○ Virginia Data Management Bootcamp: Introduction to Data Management. Ogier, Nicholls. Jan. 7.; The Important Role of Data Management in Research. Ogier, Nicholls. Jan. 7.; Archiving and Preservation of Research Data. Ogier, Nicholls. Jan. 9.; Scholarly Communication: Creative Commons, Fair Use Data in Context. Data Management Bootcamp. Pannabecker, DeBose. Jan. 9.;
  - ○ OVPR Proposal Development Institute: Data Management: Plans, policies, tools, and resources for addressing federal research mandates. Ogier, Nicholls. April 2015.
  - ○ VA Cooperative Extension Winter Conference, 3 sessions, co-instructors: Haugen, DeBose. March 4.
- ● 3 workshops coordinated:

- - Scientific Writing Workshops: Haugen Jan 15-16, May 19-20.
  - Graduate Student Mini-Data Management Bootcamp: C. Miller, DeBose, Mar. 2015
  - VT's Virginia Data Management Bootcamp: Ogier, Nicholls. Jan. 2015
- We also contributed to course redesign for GRAD 5124 and the Advanced Research Skills certificate program: Pannabecker, Walz, DeBose.