To:	Susan R. Wente, Provost and Vice Chancellor for Academic Affairs John M. Lutz, Vice Chancellor for Information Technology
From:	Research IT Special Project Working Group
Date:	May 31st, 2018
Subject:	Recommendations of the Research IT Special Project Working Group

Please see the attached report from the Research IT Special Project Working Group, convened during Fall 2016 in response to the report from the Faculty Advisory Committee on Research IT, June 2016.

Progress-to-Date

The working group has met monthly (2017-2018), considered the report of the previous committee, and developed proposals in response to the report presented in May 2017. The attached report focuses on strengthening Vanderbilt's ability to connect end-to-end our researchers and scholars to the research infrastructure through technical and consultative services, enhanced information sharing, and more effectively leveraging our current resources, particularly to support the computing, communication, and storage needs of the forthcoming Data Science Initiative.

This past year, the working group accomplished several items proposed in the May 2017 report (approved in the September 12, 2017 memo to the working group). Those items include creation of the **Research IT Resources Portal** and formation of sub-working groups to address and propose solutions for **Scalable Computing, Communication, and Storage**; **Research-Software-Developer-for-Hire and Consultation Service**; and **Research IT Training Modules**. The working group played a key role in hiring Lindsey Fox as the Senior Research IT Consultant (January 2018) and has established the **Vanderbilt University Research IT Service (VU-RITS)** office, which coordinates research IT services to meet the needs of researchers across campus with guidance from a faculty advisory committee, as described below.

The **Research IT Resources Portal** managed by VU-RITS was successfully launched at several Town Hall meetings conducted in April 2018. The portal helps researchers on campus learn about VU-RITS, schedule consultation services, provide services feedback, and find information specific to the proposed services. The portal also serves as the home for outreach efforts related to training on research IT-related topics. VU-RITS is also developing a disciplined-based resources portal containing information about research IT resources specific to broader disciplines, such as Geospatial resources for Peabody faculty, the Jean and Alexander Heard Library, and the Center for Digital Humanities resources for Humanities scholars.

Proposals for 2018-2019

To ensure effective enactment of services and resources overseen by VU-RITS, we propose that a standing **Research IT Faculty Advisory Committee** be established and charged with prioritizing communication, security, programming, and storage resource needs. Guidance from a faculty advisory group is key to establishing and triaging priorities when demand exceeds available resources. This committee will be comprised of faculty researchers and staff from Vanderbilt's schools and institutes and would operate under the auspices of the Office of Vice Provost for Research. We also propose that a sub-committee be formed to aid in the determination of training curricula and sustainment model.

We also seek approval for VU-RITS to advance the following during 2018-2019, through active collaborations with VUIT, the Jean and Alexander Heard Library, ACCRE, and other Vanderbilt University entities:

1. **Conduct a study with the help of a focused working group aimed at understanding scalable storage and processing needs** for Vanderbilt researchers and documenting findings in an actionable manner to provide solution especially for the data-intensive workflows related to scientific and scholarly research. This collaboration will help understand existing enterprise arrangements with key cloud providers, develop an automation and orchestration plan to respond quickly to researcher needs for bursts of computing and the response-time needed for processing data from new instruments, as well as archival collections, and devise a show-back/charge-back mechanism to include in grant proposals. We will also evaluate the ability to pool Amazon and other cloud provider credits for use in these scenarios.

2. **Fund a research software engineer to serve customized computer programming needs**. An initial FTE will enable VU-RITS to begin piloting the Research-Software-Developer-for-Hire service and generate interest. Additional shared research software developers may be hired using any excess funds once the FTE is fully paid through funds accumulated in the Core-services re-charge center. These new hires will possess skillsets that complement existing staff at ACCRE, as well as the Jean and Alexander Heard Library, while also providing a breadth of expertise to the research community.

3. **Allocate vouchers to support a research IT training module plan** directed by VU-RITS to host research IT training and periodic workshops throughout the year. **Vouchers are also requested to support the Research-Software-Developer-for-Hire service**. A detailed summary of these plans appear in the attached report. Vouchers will enable VU-RITS to gauge the efficacy of the proposed services and to make adjustments as needed. Such funds will also provide an incentive for using

VU-RITS, e.g., VU-RITS will pay for the first 20 hours of software development service or subsidize the cost for research software developers provided via shared Core services. Vouchers could also be used for networked storage and processing infrastructure.

Synergies with Data Science Visions

The recent Data Science Visions (DSV) report underscores the importance of the research IT infrastructure to advance data science initiatives at Vanderbilt. The proposals from the Research IT Special Project Working Group were therefore intentionally designed to maximize synergies with the DSV. In particular, the proposed Research-Software-Developer-for-Hire and Consultation Service and the Research IT Training Modules aim to support the emerging DSV efforts, while also creating transferable approaches to research IT by leveraging existing resources and those proposed by the DSV.

The remainder of the report contains proposals that provide the detailed rationale, approaches, and examples for the requests above. All proposals aim to (1) maximally leverage existing resources on campus, (2) tap into new resources, and (3) ensure that the solutions created are financially and administratively sustainable for the institution.

The Research IT Special Project Working Group Members:

Clifford Anderson
Associate University Librarian for Research & Learning

Will French
Director of Research Computing Operations ACCRE

Jens Meiler
Professor of Chemistry, Pharmacology, & Biomedical Informatics

Corey Brady
Assistant Professor of Teaching & Learning Department

Noam Lupu
Associate Professor of Political Science

Douglas C. Schmidt
Cornelius Vanderbilt Professor of Computer Science & Computer Engineering

Walter Chazin
Chancellor's Professor of Medicine, Professor of Biochemistry & Chemistry

Laurie Cutting
Patricia & Rodes Hart Professor of Special Education

John McCammon
Lead Relationship Manager VUIT

Jason Reusch
Senior Director IT Service Delivery, VUIT

Padma Raghavan
Vice Provost for Research Professor of Computer Science & Computer Eng.

Madeline Casad
Senior Lecturer Cinema & Media Arts

Carlos Lopez
Assistant Professor, Department of Biochemistry

Gayathri Narasimham
Associate Director of Vanderbilt Institute for Digital Learning

Shanmuga Sundaram
Assistant Vice Chancellor for IT

Research IT Sub-Working Groups and Members:

Scalable Computing, Communication, and Storage: Clifford Anderson, Walter Chazin, Will French Noam Lupu, John McCammon, Jens Meiler, and Shanmuga Sundaram

Research-Software-Developer-for-Hire: Corey Brady, Laurie Cutting, Will French, Noam Lupu, and Douglas C. Schmidt

Research IT Module Training and Resources: Madeline Casad, Gayathri Narasimham, and Jason Reusch

Scalable Computing, Communication, and Storage Proposal

Rationale

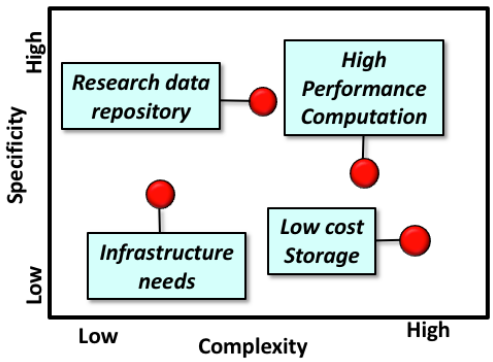
Vanderbilt researchers increasingly need support for scalable computing, communication, and storage to facilitate their discoveries and inventions. Some needs are addressed through locally provisioned services and other needs are addressed through institutional cores and shared resources, such as the Advanced Computing Center for Research and Education (ACCRE) and Distributed Online Research Storage (DORS). Based on our analysis and experience, however, many research IT needs remain unmet, especially in terms of storage, backup, and archiving. The Research IT Working Group has recently made inroads in understanding some of these unmet needs, which are addressed by this proposal.

Proposal Overview

The Vanderbilt Research IT Working Group is proposing several initiatives and services to meet the computing, communication, and storage infrastructure needs of faculty, including (but not limited to) those outlined here and described further below:

1. **Accommodate ACCRE +AWS** as an extension to cloud services
2. **Conduct a study** to understand needs and available options (both on-premises and cloud) for research

The diagram shows the relationship between these various high-performance computing and storage approaches in terms of the type of project oversight and the project specificity.[1]



Scalable Computing, Communication, and Storage Approaches

We anticipate the need for a range of scalable computing, communication, and storage approaches, such as those described below for illustrative purposes (this is not an exhaustive list):

- **ACCRE+AWS**. During the last decade, cloud service providers, such as Amazon (AWS), Microsoft (Azure), Oracle, and Google (Google Cloud), have changed the way that many IT-based applications are deployed, both in industry and academia. For example, Vanderbilt has migrated some enterprise services exclusively to the cloud (e.g., Oracle cloud ERP, website hosting, etc.). Likewise, research needs are increasingly being addressed through cloud offerings. For example, the BIOLUCIDA application for Dr. Suzana Herculano-Houzel, ORSEE software for Dr. Gregory Leo, backup storage for Dr. Kari Hoffman, web framework for Dr. Corey Brady, compute cores to handle short bursts for Dr. Carlos Lopez's biochemistry research, etc. To accommodate the increasing need—and to assure a safe path (i.e., a dedicated route assuring Vanderbilt only traffic)—VUIT has also established an AWS Direct Connect link to extend its internal network into the Amazon cloud, including redundant one Gbps connections. Similarly, VUIT has established a one Gbps dedicated VPN using Internet2 to Microsoft Azure. These bandwidths can expand concomitantly with the need over the coming years.

  The economics of transitioning from on-premises-based infrastructure to cloud-based infrastructure is different when comparing traditional enterprise IT with research IT. In particular, research IT requires a level of computing, communication, and storage resources that is unique to its needs, be it more clustered GPUs or fast spinning disks. For this reason, research IT is often more economical with infrastructure that is managed on-premises. However, there are scenarios (e.g., if a researcher needs extreme bursting up to 10k CPU cores for short periods of time) where local resources are insufficient, unavailable, or lack the flexibility to support some of the more demanding research projects. These scenarios are already occurring at Vanderbilt and are becoming more common in the future. To avoid duplicating effort—and to leverage Vanderbilt-wide contract negotiations—a logical next step for Vanderbilt is to support integration of its existing research computing, communication, and storage infrastructure and support staff with a cloud-based infrastructure, such as AWS, Azure, Google, etc.

  ACCRE has started exploring this space, with the goal of integrating their high-performance computing and big data environments into AWS. From the computing side, ACCRE has started meeting with VUIT to better understand how to leverage the existing AWS Direct Connect link between Vanderbilt and AWS. ACCRE has also focused on better automating its provisioning of on-premises resources to provide equivalent capabilities in an automated manner

[1] We will update this diagram as VU-RITS identifies and analyzes new models.

within the cloud. The long-term vision is to develop an "ACCRE-in-a-box" service that allows researchers to move storage and computations seamlessly between on-premises resources and cloud-based resources. Finally, ACCRE plans to add Amazon S3 interfaces to each of its data storage platforms (e.g., GPFS, L-Store, HDFS, and tape archive), and thereby further enhance interoperability between ACCRE resources and AWS.

- **Study to understand research storage needs and available options.** Some options currently exist at Vanderbilt to address the scalable storage needs of some faculty for their research pursuits. Many faculty, however, are not fully aware of the options at their disposal. Moreover, the current options do not readily meet the diverse storage needs of all researchers, e.g., the offered price point in many cases is deemed as too expensive.

  To better address these needs, we request approval to conduct a trade study aimed at understanding available scalable storage options, both on-premises and cloud-based, with the goal of evaluating their merits vis-à-vis the needs of Vanderbilt researchers. To facilitate this effort, we will conduct the following activities

  - Analyze on-premises storage from ACCRE, DORS, and central IT, as well as *ad hoc* solutions deployed at various departments and units to meet their localized needs,
  - Meet with the appropriate Associate Deans at all the schools and colleges at Vanderbilt to identify researchers who have storage needs and understand their requirements and timelines and hold follow-up meetings with these researchers,
  - Assess commercial cloud services and other options, such as L-Store, AuriStor, and Zenodo, to determine if/how they might meet our needs,
  - Poll peer institutions to glean emerging trends and best-practices.

  These activities will enable us to clearly document the research community's needs, identify the unmet gaps, understand the current state of scalable storage services at Vanderbilt, develop a comprehensive total cost of ownership for each option, and recommend a course of action that meets the immediate and future scalable storage needs for researchers.

Request

The Research IT Working Group is requesting support/approval from the Provost to pursue the above proposals, as follows:

- The ACCRE + AWS proposal requests approval to work closely with VUIT to understand and leverage existing enterprise arrangements with key cloud providers, develop an automation and orchestration plan to respond quickly to researcher needs for bursts of computing, and develop a show-back/charge-back mechanism for inclusion in grant proposals.
- Approval to conduct a study aimed at understanding scalable storage needs for Vanderbilt researchers and documenting findings.

Examples of Scalable Storage Approaches

The following list summarizes several examples of scalable storage solutions that may be relevant for researchers at Vanderbilt. This list is not exhaustive, it simply provides some examples of potential approaches that we propose to include in our trade study.

- **Implementing Zenodo for research data management and storage**. There is a growing need on campus for a research data repository. Researchers prefer an easy way to archive and provide access to data sets to share with peers, satisfy federal mandates, and foster the reproducibility of their results. At present, researchers make do with a variety of systems, including web servers, the library's institutional repository, and third-party sites like Inter-university Consortium for Political and Social Research (ICPSR). The lack of a shared data repository at Vanderbilt, however makes it hard to track research data sets across campus and has led to a wide range of data management practices, some of which fall short of evolving norms and best practices for archiving research data.

  We propose to implement Zenodo[2], which is open-source software developed by CERN for research data management. Zenodo provides key features for archiving and sharing research data, including the ability to issue Document

[2] https://github.com/zenodo/zenodo

Object Identifiers (DOIs) for versions of datasets, to select appropriate licenses for reuse, and to integrate with GitHub to preserve code along with data. Rather than using CERN's infrastructure as our backend storage system, we propose to host the datasets in Zenodo on ACCRE. Zenodo would provide a straightforward interface to ACCRE for researchers from disciplines (like the humanities and some social sciences) who are unaccustomed to working with the tools of high performance computing.

- **L-Store**. Storing, moving, and analyzing research data is a major challenge in university settings. While solutions exist at Vanderbilt through ACCRE, DORS, and VUIT, the cost of these services (e.g., backed-up storage on ACCRE or DORS is $204/terabyte/year) can be cost prohibitive for researchers with massive (e.g., >100 terabytes) storage capacity needs. ACCRE has been developing a storage platform called Logistical Storage (L-Store[3]) in-house with a special emphasis on making high-performance bulk storage more affordable and accessible.

  ACCRE has developed and "battle tested" L-Store over the past decade. Vanderbilt participates in the high-energy physics project called Compact Muon Solenoid (CMS) that is based in Switzerland at CERN, which is a European organization for nuclear research. For the last seven years Vanderbilt has served as a Tier2 CMS institution and currently hosts over 4 petabytes of storage (i.e., 4,000 terabytes) in support of this project alone. All this data resides in L-Store, along with over 300 terabytes of video data from the Vanderbilt Television News Archive (TVNA) located in the Jean and Alexander Heard Library. In addition, Dr. Bennett Landman, who is an Associate Professor of Electrical Engineering and Chair of the ACCRE Faculty Advisory Board, is the PI on a submitted Discovery Grant that proposes L-Store pilot projects revolving around two of the most data intensive workflows on campus (the eXtensible Neuroimaging Archive Toolkit and Next-Generation Sequencing/Genotyping data generated through the VANTAGE core).

  We envision L-Store being applied in the research data storage ecosystem at Vanderbilt to support large-scale storage that must be accessed and analyzed from a high-performance computing environment. Although L-Store is not designed for small files or desktop access, it is well-suited to meet the needs of many investigators, especially with the support of the university behind it.

- **AuriStor**. After more than a decade of explosive growth in the volume and velocity of data generation in nearly every field of research, the pervasive balkanization of storage across the institution is making problems of data logistics (i.e., "ensuring that data are in the right place at the right time and accessible by the right users") increasingly intractable. Researchers need the ability to (1) make large amounts of storage suitably proximate to the equipment generating the data flows, (2) move that data to other locations for analysis, and (3) share the data with remote collaborators in a secure manner.

  The AuriStor[4] file system can satisfy these needs. AuriStor provides a global namespace that allows secure, platform independent, and high-performance access to data in a transparent manner. AuriStor is an enterprise grade extension of the Andrew File System (AFS), but with modern security, performance, scalability, and functionality. The AuriStor product development began in 2008 under a Department of Energy SBIR to create a high performance global file system to meet modern performance, security, management, and operational requirements. As with the original AFS, AuriStor is designed for the Wide Area Network (WAN) to transparently and uniformly access data, regardless of where it is physically or geographically stored with zero local configuration. By providing a global namespace, institutions can break down data silos by having an easily deployed mechanism for independent laboratories or departments to expose their data and storage for access and collaboration in a secure, high-performance, and auditable manner.

  AuriStor file servers can utilize the same storage platform as L-Store with the hardware and software stack centrally managed. Likewise, each department or center can have a local administrator responsible for managing user access permissions and storage quotas. In this way, the AuriStor appliance can be centrally managed, but locally administered, thereby helping to reduce friction between researchers and administrators. We envision AuriStor fulfilling the role of an edge storage device connecting into the larger campus and national ecosystem to facilitate data movement to other locations for analysis.

[3] http://www.lstore.org
[4] https://www.auristor.com/

Research-Software-Engineer-for-Hire and Consultation Service Proposal

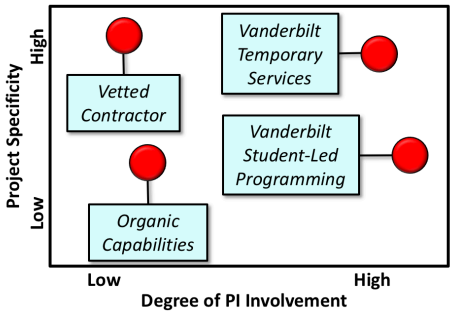
Rationale

Faculty often have research accounts or grant funds, but may be unwilling or unable to hire full-time staff to support their research software development needs. Such faculty would benefit from some type of "liquid workforce" service that provides them with on-demand access to shared technology expertise. This service, which would be managed by the Vanderbilt University Research IT Service (VU-RITS) Office, aims to help faculty develop research IT solutions, especially with data-intensive workflows, while also enabling shared technology experts to add value to multiple research programs throughout the university.

Proposal Synopsis

The Vanderbilt Research IT Working Group is proposing a new service aimed at providing researchers with consultation and software development resources by leveraging various liquid workforce models, including (but not limited to) those outlined here and described further below:

1. **Highly-specific, short-term project support with PI oversight**, e.g., short term hires provided through Vanderbilt Temporary Services.
2. **Specialized web development support for custom data gateways and digital collections with PI oversight**, e.g., HireADore , Vanderbilt student-led programming services, etc.
3. **General project support with various degrees of built-in project management**, e.g., Vanderbilt organic capabilities found within research Cores and groups, the library, and the Center for Digital Humanities.



4. **Highly specialized project needs with various degrees of built-in project management**, e.g., Vanderbilt approved contractors.

The diagram shows the relationship between these various liquid workforce models in terms of the degree of researcher involvement and the project specificity.[1] The VU-RITS office will connect researchers with consultative services and software development solutions that are customized to their project needs. In this proposed approach, a researcher would approach VU-RITS to convey their needs. VU-RITS would then identify an appropriate solution using one of the proposed liquid workforce models together with the level of project oversight deemed necessary. VU-RITS will work directly with faculty to develop a project plan and best-practices model for a variety of needs, including (but not limited to) coding for data analyses, simulations and computation in various programming languages (Python, R, Matlab, C, apps for mobile devices, etc.), various types of application development (such as web apps), and high-performance computing and storage solutions.

Liquid Workforce Models

We anticipate the need for a range of liquid workforce models, such as those described below for illustrative purposes (this is not an exhaustive list):

- **Highly-specific, short-term project support with PI oversight**. This model would involve Vanderbilt Temporary Services (VTS), who would work with VU-RITS and Vanderbilt faculty to identify a pool of research software developers (with a mixture of general-purpose and specialized IT skills). VTS would pre-vet this pool of developers, who would remain "deactivated" until needed for a specific project. VTS can activate and deactivate temporary staff as needed and charge out to the appropriate project for up to 6 months. Any projects that go beyond this period would fall under a "term" contract, which usually lasts a maximum on one year, but could go for longer periods of time if

[1] We will update this diagram as VU-RITS identifies and analyzes new models.

sufficient justification is provided (e.g., to support the SkyVU roll-out). Project oversight would be managed by a principal investigator (PI).

- **Specialized web development support for custom data gateways and digital collections with PI oversight**. This model would leverage HireADore, as well as student-led software development groups, such as "VandyApps" (which has existed for a decade) or the newly formed Vanderbilt AI club. We foresee that these student-led groups would have the ability to fulfill web development and programming needs. This model would also require PI oversight.

- **General project support with various degrees of built-in project management.** This model uses various existing research groups and institutes (such as the Institute for Software Integrated Systems and the Institute for Space and Defense Electronics), as well as core services at Vanderbilt (such as the Advanced Computing Center for Research and Education, Creative Data Solutions, and Vanderbilt Institute for Digital Learning) to cover a wide range of programming capabilities. With proper coordination from VU-RITS, it may be possible for these core services to provide more advanced levels of software development and project management support for researchers. VU-RITS will identify pockets of talent within these core services and connect them with the PIs. We anticipate that these core services would provide some form of project management, which will be built into the cost structure.

- **Highly specialized project needs with various degrees of built-in project management**. This model is intended to support faculty whose project needs cannot be easily met by any of the models above, e.g., projects where highly specialized technical solution may be required, such as natural language processing, concept mapping, or text mining applications using XML. To support these use-cases, VU-RITS will identify a manageable number of external consulting firms that can be pre-vetted to provide a range of highly specialized software development solutions and project management oversight.

Request

The Research IT Working Group is requesting Provost support to implement the liquid workforce models described in this proposal. Support may come in several forms, e.g., as outlined in the following two scenarios:

A. **An initial FTE**. In this scenario, an initial FTE will enable VU-RITS to begin piloting the service and generate interest. As demand grows, effort could be shifted from the Provost's Office to researcher grant dollars until the service is sustainable, which we expect to occur within one year. Additional shared research software engineers may be hired once the FTE is paid through and funds accumulate in the core-services re-charge center. These new hires will possess skillsets that complement existing staff at ACCRE, Libraries, etc. and provide a breadth of expertise to the research community.

B. **Start-up funds to kickstart the program**. In this scenario, seed funding in the form of vouchers will be applied to provide research software development support. Once demand is high enough to achieve self-sustainment, we expect that support from the Provost's Office will no longer be needed. These vouchers will also give VU-RITS the ability to gauge the efficacy of these support models and to make adjustments as needed. Such funds would go to providing an incentive for using VU-RITS, e.g., VU-RITS will pay for the first 20 hours of software development service or subsidize the cost for shared Core service software developers.

The advantage of scenario A is that it will result in a more specialized, advanced workforce to draw from for researchers, e.g., to serve data analyses needs. We foresee the need growing beyond the initial FTE request. Ideally, we would ramp up to support another FTE longer term.

The advantage of scenario B is that we expect service rates will be lower through Vanderbilt Temporary Services (although perhaps not external consulting agencies). Another advantage of scenario B is potentially a more diverse set of expertise. For example, Temporary Services has a larger pool of software developers to draw upon and thus may offer support for a wider range of software applications, platforms, languages, and tools.

Example Applications of Liquid Workforce Models

The following table summarizes several examples of applying the various VU-RITS liquid workforce models described above. This table is not exhaustive, it simply provides some examples of the research-software-developer-for-hire models.

Need	Solution
Researchers in neuroscience produce very large data sets requiring high-performance computational analysis and programming of GPUs in a cluster. While the research team includes a graduate student who also serves as a sys admin, this type of work exceeds their skillset and the time they have available to learn and practice advanced coding. As a result, the group often develops applications that are inefficient, error-prone, and expensive/time-consuming to maintain.	VU-RITS develops a project plan and consults on best practices for the required high-performance computational analysis. VU-RITS identifies a software developer for hire through ACCRE (**a Vanderbilt Core Service**) and arranges to split this staff's effort between ACCRE and the researchers' grant funds, to develop more efficient, easier-to-maintain applications in coordination with systems experts at ACCRE. VU-RITS devises a cost model and coordinates with the researchers as further development of the applications is needed and new projects come online.
Researchers in the Peabody School of Education develop specialized educational software and tools that require more advanced JavaScript development skillsets than their existing staff. The program needs updating frequently and requires a specialized need for familiarity with the program.	VU-RITS identifies a software developer for hire through **Vanderbilt Temporary Services**. The hired developer spends 6 months developing the software and remains on call for future debugging and code updates. While VTS maintains HR records and pay systems, the PI on the project manages personnel. Once completed with the project, the staff member assumes an inactive status until needed again.
Researchers in the department of medicine (epidemiology) are studying the prevalence and types of lung diseases found in the USA. They found these cases are correlated with spatial and environmental parameters. To understand the relationship between the environment and disease prevalence, the research team needs a geographic information specialist. They have the funds but do not have the time or personnel to implement the methods before their funds run out.	VU-RITS works with the researcher to map out the needs of their projects and identify expertise and resources that can be leveraged to complete this project. Since the time frame for the project is relatively short, the VU-RITS determines that a **pre-vetted private consulting firm** would be the best mechanism for the completion of this project. The VU-RITS analyst identifies a contractor within the firm and develops a project plan for the researcher in consultation with the contractor. The project is completed within the timeframe needed.
Researchers in the social sciences whose grant requires that they create and maintain a website detailing the project over its lifetime. Digital collections must be made available to other scholars around the world. The research group does not have the time or the background to create more than a basic site to meet the requirements of the grant. As a result, their research is not showcased, and the group misses out on a powerful tool for communicating within the project, as well as with study participants, program officers, policy makers, and the general public.	After working with the researcher on the project plan, an analyst from VU-RITS leverages the expertise and resources within the **Vanderbilt student-led programming services**. The analyst coordinates meetings with the students and at with the Vanderbilt Library Digital Scholarship team, which has expertise in archival practices, and can advise on the creation of a science gateway that showcases the research and serves as a portal for dissemination of the work. Web Communications maintains the website, and VU-RITS refines a model for establishing science gateways for other Vanderbilt researchers.

Research IT Training Module Proposal

Rationale

Faculty often have research questions or needs that can be met with services already available on campus. In many cases, addressing these questions/needs are simply a matter of connecting faculty with those resources. The Research IT Training program aims to connect faculty based on their IT needs or by their discipline. A basic level of training on resources will also help ease communication barriers between researchers and their unmet research IT needs. For example, incoming research faculty could benefit from workshops connecting them with commonly used research IT resources on campus, thereby strengthening the network of incoming faculty and reinforcing the use of current resources. Likewise, faculty can also benefit from knowing about resources that enable more effective research in their disciplines or that could yield fruitful cross-disciplinary collaborations. The Research IT Training program could also supplement existing award programs, such as TIPS, Discovery grants, and others.

Proposal Overview

The Vanderbilt Research IT Working Group is proposing a new service aimed at providing research faculty and staff with a training series open to current and incoming faculty and staff researchers, as well as graduate students. This training series will include various learning modalities, such as **seminars, online workshops, training videos, in-person workshops, and bootcamp-style multi-day workshops** for basic research-intensive tools and data workflows. Research IT training efforts can be taught by various leaders in specific areas of the Vanderbilt community. These short workshops and seminars would cover a range of basic information necessary to help researchers be productive at Vanderbilt. Ideally, each short course would be free of a disciplinary focus and general enough that anyone could attend (from novice researcher to advanced). While broader training would be free of such focus, customized training could be available for specific research groups needing to onboard new members to hone new skills for existing members. This training series could range in content from topics including (but not limited to):

- Introduction to Data and Databases
- Introduction to Git and Version Control
- Introduction to High-Performance Computation
- Parallel Computing
- Introduction to Data Storage on Campus
- Introduction to the Command Line (Unix/Linux)
- Introduction to Python
- Introduction to R
- Introductory Software Carpentry
- Introduction to Data Science at Vanderbilt

It may be possible to monetize these workshops in some manner so they could ultimately be self-sustaining, similar to how the Center for Quantitative Science runs their Summer Institute. We also propose that a sub-committee of the Research IT Faculty Advisory Committee be formed to determine training curricula and appropriate formats.

Request

The Research IT Working Group is requesting Provost support for implementation of the services described above. Support could be provided in the form of the following scenarios:

A. New faculty orientation time (half an hour) for pitching VU-RITS resources and training component.

B. Start-up funds for training component. In this scenario, seed funding will be directed to VU-RITS to host research IT training and periodic workshops throughout the semester. Funds would go to the following:

   a. Guest speakers (honorariums, travel funds where applicable) [cost: approximately $10-15,000]

   b. Services like W3Schools, Lynda.com, DataCamp, provide basic level training in a number of areas and this content can service a general campus population (basic programming, technical courses, and data science). [Cost: An initial subscription of 50 licenses to Lynda or DataCamp can range, depending on the number of seats, up to $15,000 /year]

   c. Faculty and graduate incentives for learning (badging) [cost: approx. $2000]

d. Student assistants to help with video production for resources content (short courses). [cost: $10-20,000 depending on graduate and/or undergraduate assistants]

e. Miscellaneous (posters, flyers, "lunch and learn") [cost: approx. $1000]

The advantage of these scenario is that the training can be hosted by various personnel on campus and attendance can be incentivized where necessary by providing software licenses and digital badging. To provide longevity to the training material, and solutions in scenarios like in the examples above, these could be captured in simple videos (with the help of VIDL and their "self-serve" studio) and made available online via the portal.

Examples of Research IT Training Approaches

The following table contains examples of applying the various research IT training approaches presented above. Again, this table is not exhaustive.

Need	Solution
A group of researchers has several incoming staff and graduate students that will be working on data-intensive projects. The team does not have the organic capacity or expertise in the fields that is necessary for their project, including geospatial analysis, and high-performance computation. The group has had problems in the past with general organization, and records of file names. The team begins to research new IT possibilities but is impeded by inefficient or convoluted procedures for accessing support.	The group contacts VU-RITS and the office identifies key areas that the research team will need to complete their projects. VU-RITS organizes a half-day workshop aimed at orienting both new and more advanced research staff. VU-RITS organizes basic training for HPC, Python, GIS, and general research data workflows. The group leaves training with a customized strategy for storing their research files on a shared workspace, with standardized file naming structures. The group is able to easily adapt their new workflow for use with ACCRE and other entities on campus. The boot camp courses are filmed at VIDL for others to access remotely.
A researcher in the humanities whose work involves capture, editing, and analyzing video data needs resources to help train the project team on the technical background needed to execute this methodology for completion of successful research. The team also needs help with identifying a procedure for assigning DOIs and storing of their datasets. Processes exist on campus, but the group is unsure of which one is appropriate.	The research group partners with VU-RITS to develop a project plan. VU-RITS coordinates with on-campus entities (such as VIDL) to provide training for the video capture and editing services needed and helps develop an appropriate storage solution. The researchers are connected with ACCRE to get an instance of Zenodo for requesting DOIs and public facing storage. VU-RITS identifies that this researcher's problem is a common one. VU-RITS documents the solution designed here as a "best practice" that becomes known through the online Resource Portal.