Researcher Services: Proposal to Increase Distinctive Excellence

July 20, 2018

Tiffany Grant & Ted Baldwin (Co-Leaders, Researcher Services) with co-authors Amy Koshoffer (Science Informationist) and Kristen Burgess (Special Advisor to the Senior Associate Dean)

Introduction and Methods

A retreat focused on the Researcher Services Area of Distinctive Excellence (ADE) was held on March 14, 2018. Following the retreat, several future actions were identified to carry forward the momentum of the ADE and to assist in forming a proposal that increases distinctive excellence for Researcher Services. Two sub-groups were formed to conduct research and prepare recommendations. First, a "Services" sub-group prepared a service catalog detailing the following: 1) current scope of the Researcher Services team's skills and educational efforts, and 2) critical skills gaps that must be filled in order to deliver a comprehensive program of researcher support and enable us to pursue deeper partnerships. Second, a "Data Science" sub-group researched other institutions' models for advanced data analysis/data science support. In these models, we identified tiered engagement models (triage, amplify, partner) that provide recognition for increased levels of contribution.

With this proposal, we bring together findings of the two sub-groups, in the context of UCL's ADE framework, to recommend resources and strategies that are necessary to enhance Researcher Services' efforts and truly achieve distinctive excellence. We present this summary and recommendations along with four detailed appendices. We present an expanded mission and vision for Researcher Services ADE, a new name for the unit and its associated spaces, and a tiered engagement model to provide recognition for deeper partnerships. In addition, we request resources and actions related to staffing, professional development, and outreach/promotion that allow us to achieve a holistic support model and to offer data analysis/data science support and education.

Service Catalog of Current Work, Data Science Roles

The "Services" sub-group identified skills/activities in the UCL "Research Data Services" scope of work. The summary below is based on the three-tier engagement model. Complete findings for this sub-group are in Appendix A. Also, recommendations for proposed new roles in "Data Science" are described in Appendix B.

Staffing

Triage

- Data Management Consultation
- Data management plan construction
- ORCID profile assistance
- Altmetrics instruction
- Evaluate data collections for library collections
- GIS research guide
- Data management compliance with federal and university regulations

Amplify

- Education workshops and consultation
- Research services outreach
- Data management ethics consultation
- Research data modeling
- Bibliometric network analysis
- Data literacy education
- Extensive workshop series: Bioinformatics, REDCap, literature searching, R, ArcGIS, QGIS, MyNCBI, SciENcv, data management, LabArchives, Research Impact factor, data preservation
- Project design and planning
- Publishing assistance
- Systematic or integrative reviews

Partner

- Metadata creation
- Data storage and archive
- Data visualization
- Conference host/partner for informatics
- Host outside speakers on informatics topics
- Dataset creation
- Data cleaning/preparation
- Low level spatial analysis

Partnerships

Critical skills gaps we must full to achieve increased excellence:

- Statistical/analytical expertise (both quantitative and qualitative data analysis), including data cleaning, analyses, and interpretation/inference
- Knowledge of methods and best practices for research reproducibility
- Finding and extracting data from existing published datasets and record sets
- Programming/coding experience
- Knowledge of next practices in clinical data management
- Database design

Recommendations and Budget Request

- Re-name this Area of Distinctive Excellence as *Research and Data Services* or *RDS*
    - Adopt campus-specific space names: *RDS East*, *RDS West*
- Expand mission and vision statements to encompass a holistic view of support and partnership:
    - **Mission Statement**: The UCL Research Data Services group inspires the creation of knowledge and enhances research productivity across the UC research community through the development and implementation of interdisciplinary research data services that enables research and promotes synergistic collaborations between UCL and UC researchers.
    - **Vision Statement**: The UCL Research Data Services group will be recognized as innovative and effective research partners by engaging researchers in data services through collaboration, consultation, and training that aims to amplify the global UC research profile.
- Hire a new staff position, Data Analysis Specialist (see Appendix C for proposed Position Description). $ TBD
- Fund the Data Visualization Specialist position as a permanent position, to ensure retention and long-term contributions from this vital role.
- Increase student worker staffing on both East and West campuses by 30 hours/wk. $15,000
    - The additional student support opens time for librarians/informationists to partner more deeply with researchers on data consultations and education.
- Expand professional development funding to increase skills in Data Analysis/Data Science (see Appendix D for full list of identified opportunities). $15,000
    - Example development opportunities:
        - "Foundations of Data Sciences" onsite training - $5,000 plus travel for facilitator (District Data Labs)
        - Data Science Online Course- $2,895/person x 2 people $5,790 (Data Incubator)
- Build an assessment strategy, beginning with a targeted user survey distributed to those we have engaged through research data services' events, education, and consultations (RDS "master contact list"). Confirm that priorities and efforts align with users' educational and job needs.
- Update the UCL Research Data Services website to strengthen branding. Create an online service catalog to clearly inform and guide users and stakeholders.

Summary

With new staff positions, increased funding for student workers, and restructuring and development of our current faculty and staff, we will be able to reach new levels of excellence within Research Services. A new identity, mission and vision, combined with these additional resources will allow us to grow strategically and purposefully align with UCL's vision for increased distinction and scholarship.

Appendix A: Service Catalog and Gap Identification

A sub-group of the Researcher Services Team (Ted Baldwin, Tiffany Grant, Amy Koshoffer, Richard Johansen, and Kristen Burgess) was tasked with creating a "service catalog" of 1) skills that are currently operational and 2) skills that are not currently operational but are highly desired to have a holistic programs of support for UCL's and UC's research missions. The sub-group documented skills and activities across these categories: data science/analysis, visualization, data management, informatics/health informatics, GIS, data gathering, research data, and education. Each skill or activity was designated as operational or desired, and we defined the primary person responsible for performing the work (informationist, librarian, specialized student, staff specialist/technician, or a new position). We also identified the phase in the data lifecycle phase in which the task or skill primarily occurs (pre-planning, data creation, data processing, data analysis, reuse/preservation, research impact). Lastly, we noted the most relevant tier of engagement (triage, amplify, partnership).

Table 1 below lists highly-desired skills that are not currently operational, organized by tier of engagement. This list informs our needs for additional resources in staffing and professional development. Table 2 lists currently-operational skills and activities. This list is beneficial for promotion and outreach on available programs of support. Finally, Table 3 provides a single comprehensive view of operational and not currently operational skills and activities. This list is organized by tier of engagement. This list is also available as an Excel spreadsheet, for ease of analysis: https://uc.box.com/s/aj0ihvjiefrohzlmntv7gsr1h7u5yufk.

We are well-positioned to be leaders at UC in some areas of Researcher Services - data management, GIS, education. However, new positions and professional development investments are necessary for us to become leaders and key partners in other areas - research data services, data science/analysis, database design, data cleaning/mining, informatics/health informatics and ontology support. We currently lack staffing capacity and skills to adequately expand into these areas.

Table 1: Highly-Desired Skills Not Currently Operational

Tier	Category	Skill	Primary Person Responsible
Amplify	Research Data	Survey design consultation	New Position
Amplify	Research Data	Dataset Discovery	New Position
Amplify	Research Data	Field requests for data from ICPSR (social sciences)	New Position
Amplify	Research Data	Database design	New Position
Amplify	Data Science / Analysis	Research Reproducibility	Informationist
Amplify	Data Science / Analysis	High Performance Computing	New Position
Partnership	Data Science / Analysis	Data Analysis	Specialized Student
Partnership	Data Science / Analysis	Data Cleaning	New Position
Partnership	Data Science / Analysis	Data Mining	Staff Specialist/Technician
Partnership	Informatics / Health Informatics	Biomedical Big Data Analysis (RNAseq, GWAS, phenotypic analyses)	New Position
Partnership	Informatics / Health Informatics	Clinical Data Management	Informationist
Partnership	Informatics / Health Informatics	Ontology Support	New Position
Partnership	GIS	High level spatial analysis	New Position

Table 2: Skills Currently Operational

Tier	Category	Skill	Primary Person Responsible
Triage	Data Management	Data Management Consultation	Meeting with lab groups to discuss data management practices
Triage	Data Management	Data Management Plan Consultation/Construction	Provided Edits to UC internal Grant data management plan for School of Social Work Professor on
Triage	Research Data	ORCID Profile Assistance	Helping Students and Faculty register for ORCID
Triage	Research Data	Altmetrics Instruction	Research Impact workshops, Literature database workshops i.e. Scopus
Triage	Research Data	Evaluate Data Collections for Library Collections	Respond to researcher requests for data set purchases
Triage	GIS	Research Guide	https://guides.libraries.uc.edu/GISandData/Home
Triage	Data Management	Compliance with Federal and University Regulations	Coordinated with UC IP office to host talks on intellectual property
Amplify	Data Management	Education	All workshops and consultation sessions
Amplify	Data Management	Outreach	Outreach to new faculty on Researcher Services
Amplify	Data Management	Ethics Consultation	Handling Restricted data - promoting use of REDCap
Amplify	Research Data	Data Modeling	Discussions with faculty about data to be collected and how to organize for better management
Amplify	Research Data	Bibliometric Network Analysis	Systematic reviews for COM grants
Amplify	Research Data	Data Literacy Education	Spreadsheet best practice workshops, restricted data webinars
Amplify	Informatics / Health Informatics	HSL Workshops and Webinars (informational)	Bioinformatics workshops, REDCap instruction, literature searching series, etc.
Amplify	Informatics / Health Informatics	HSL Data Analysis Teaching	HSL students teaching and tutoring in R
Amplify	GIS	Introduction workshops on ArcGIS and QGIS	Current workshop on ArcGIS
Amplify	GIS	Project design &	Consultations to organize spatial data, address

		planning	spatial analysis
Amplify	GIS	Dataset Discovery	finding shapefiles, census data
Amplify	GIS	Publishing/Communication	Data submitted to Scholar/ ORCID created
Amplify	Education	Hands-On REDCap	Current workshop taught during school year
Amplify	Education	Literature Searching for Systematic or Integrative Reviews	Current workshop taught during school year
Amplify	Education	MyNCBI and SciENcv: A More Personalized NCBI	Current workshop taught during school year
Amplify	Education	Introduction to Programming in R	Current workshop taught during school year
Amplify	Education	Statistical Analyses with R	Current workshop taught during school year
Amplify	Education	Bioinformatics: Seeking and Finding Data	Current workshop taught during school year
Amplify	Education	NLM Biomedical Informatics: Bedside to Bench with NCBI	Current workshop taught during school year
Amplify	Education	Managing Data from Generation to Preservation	Current workshop taught during school year
Amplify	Education	Data Preservation	Current workshop taught during school year
Amplify	Education	Maximizing Your Research Impact Factor: It's Personal too	Current workshop taught during school year
Amplify	Education	Top 10 Data Management Tips	Current workshop taught during school year
Amplify	Education	Introduction to GIS	Current workshop taught during school year
Amplify	Education	Data Management Series	Current workshop taught during school year
Amplify	Education	Spreadsheet Best Practices	Current workshop taught during school year
Amplify	Education	LabArchives Electronic Lab Notebook Workshop	Current workshop taught during school year
Partnership	Data Management	Metadata Creation	Work of Metadata librarian - Scholar guidelines
Partnership	Data Management	Data Storage and Archive Options Box, LabArchives, OSF, REDCap, Scholar)	Provost Grant for LabArchives, Set up of OSF
Partnership	Data Science / Analysis	Data Visualization	Data visualization specialist hired April 2018
Partnership	Informatics / Health Informatics	Conference host/partner for informatics	Precision Medicine conference with CCHMC

Partnership	Informatics / Health Informatics	Host outside speakers on informatics topics	NLM Series
Partnership	GIS	Dataset Creation	data mining to compile data set for analysis
Partnership	GIS	Data Cleaning/Preparation	geocoding, clipping shapefiles, variable specifications
Partnership	GIS	Low level spatial analysis	Intermediate tasks - buffers, joining datasets, flow analysis, animations
Partnership	GIS	Data Visualization	Map created for journal article

Table 3: 2018 Comprehensive Service Catalog and Skill Gaps (Sorted by Tier - Triage / Amplify / Partnership

Tier	Category	Skill	Status	Primary Person Responsible	Example
Triage	Data Management	Data Management Consultation	Operational	Informationist	Meeting with lab groups to discuss data management practices
Triage	Data Management	Data Management Plan Consultation/ Construction	Operational	Informationist	Provided Edits to UC internal Grant data management plan for School of Social Work Professor
Triage	Research Data	ORCID Profile Assistance	Operational	Librarian	Helping Students and Faculty register for ORCID
Triage	Research Data	Altmetrics Instruction	Operational	Informationist	Research Impact workshops, Literature database workshops i.e. Scopus
Triage	Research Data	Evaluate Data Collections for Library Collections	Operational	Librarian	Respond to researcher requests for data set purchases
Triage	GIS	Research Guide	Operational	Specialized Student	https:/guides.libraries.uc.edu/GISandData/Home
Triage	Data Management	Compliance with Federal and University Regulations	Operational	Informationist	Partnered with UC Intellectual Property Office to host talks on intellectual property and patents
Amplify	Research Data	Survey design consultation	Desired	New Position	Data collection strategies and tools
Amplify	Research Data	Dataset Discovery	Desired	New Position	Locating ocean current data for plant biology lab to understand seed migration

Amplify	Research Data	Field requests for data from ICPSR (social sciences)	Desired	New Position	Includes requests for non-medical protected data sets and data sets requiring physical secured access (enclaves)
Amplify	Research Data	Database design	Desired	New Position	Method to manage data
Amplify	Data Science / Analysis	Research Reproducibility	Desired	Informationist	Code writing and validation techniques
Amplify	Data Science / Analysis	High Performance Computing	Desired	New Position	Collaborating with IT@UC on HPC educational efforts
Amplify	Data Management	Education	Operational	Informationist	All workshops and consultation sessions
Amplify	Data Management	Outreach	Operational	Informationist	Contact new faculty to explain value of Researcher Services
Amplify	Data Management	Ethics Consultation	Operational	Informationist	Handling Restricted data - promoting use of REDCap
Amplify	Research Data	Data Modeling	Operational	Informationist	Discussions with faculty about data to be collected and how to organize for better management
Amplify	Research Data	Bibliometric Network Analysis	Operational	Informationist	Systematic reviews for College of Medicine grants
Amplify	Research Data	Data Literacy Education	Operational	Informationist	Spreadsheet best practice workshops, restricted data webinars
Amplify	Informatics / Health Informatics	HSL Workshops and Webinars (informational)	Operational	Informationist	Bioinformatics workshops, REDCap instruction, literature searching series, etc.
Amplify	Informatics / Health Informatics	HSL Data Analysis Teaching	Operational	Specialized Student	HSL students teaching and tutoring in R
Amplify	GIS	Introduction workshops on ArcGIS and QGIS	Operational	Informationist	Current workshop on ArcGIS
Amplify	GIS	Project design & planning	Operational	Informationist	Consultations to organize spatial data, address spatial analysis
Amplify	GIS	Dataset Discovery	Operational	Specialized Student	Finding GIS shapefiles, census data

Amplify	GIS	Publishing/Communication	Operational	Librarian	Data submitted to Scholar, ORCID ID created
Amplify	Education	Hands-On REDCap	Operational	Informationist	Current workshop taught during school year
Amplify	Education	Literature Searching for Systematic or Integrative Reviews	Operational	Librarian	Current workshop taught during school year
Amplify	Education	MyNCBI and SciENcv: A More Personalized NCBI	Operational	Informationist	Current workshop taught during school year
Amplify	Education	Introduction to Programming in R	Operational	Specialized Student	Current workshop taught during school year
Amplify	Education	Statistical Analyses with R	Operational	Specialized Student	Current workshop taught during school year
Amplify	Education	Bioinformatics: Seeking and Finding Data	Operational	Informationist	Current workshop taught during school year
Amplify	Education	NLM Biomedical Informatics: Bedside to Bench with NCBI	Operational	Informationist	Current workshop taught during school year
Amplify	Education	Managing Data from Generation to Preservation	Operational	Informationist	Current workshop taught during school year
Amplify	Education	Data Preservation	Operational	Informationist	Current workshop taught during school year
Amplify	Education	Maximizing Your Research Impact Factor: It's Personal too	Operational	Informationist	Current workshop taught during school year
Amplify	Education	Top 10 Data Management Tips	Operational	Informationist	Current workshop taught during school year
Amplify	Education	Introduction	Operational	Informati	Current workshop taught during

		to GIS		onist	school year
Amplify	Education	Data Management Series	Operational	Informati onist	Current workshop taught during school year
Amplify	Education	Spreadsheet Best Practices	Operational	Informati onist	Current workshop taught during school year
Amplify	Education	LabArchives Electronic Lab Notebook Workshop	Operational	Informati onist	Current workshop taught during school year
Partnership	Data Science / Analysis	Data Analysis	Desired	Specializ ed Student	Current R and Python Workshops and Data analysis assistance at HSL ((Statistical Support for numeric and quantitative data))
Partnership	Data Science / Analysis	Data Cleaning	Desired	New Position	Partitioning of Education Data to create usable subsets
Partnership	Data Science / Analysis	Data Mining	Desired	Staff Specialist /Technici an	Assistance with databases and finding appropriate data
Partnership	Informatics / Health Informatics	Biomedical Big Data Analysis (RNAseq, GWAS, phenotypic analyses)	Desired	New Position	
Partnership	Informatics / Health Informatics	Clinical Data Management	Desired	Informati onist	
Partnership	Informatics / Health Informatics	Ontology Support	Desired	New Position	Example from OHSU- http://www.ohsu.edu/xd/education/library/about/departments/ontology/
Partnership	GIS	High level spatial analysis	Desired	New Position	Spatial statistical analysis
Partnership	Data Management	Metadata Creation	Operational	Librarian	Work of Metadata librarian - Scholar guidelines
Partnership	Data Management	Data Storage and Archive Options Box, LabArchives, OSF, REDCap, Scholar)	Operational	Informati onist	Provostal grant for LabArchives pilot, set up of Open Science Framework

Partnership	Data Science / Analysis	Data Visualization	Operational	Staff Specialist /Technician	Data Visualization Specialist hired in April 2018
Partnership	Informatics / Health Informatics	Conference host/partner for informatics	Operational	Informatonist	Precision Medicine conference with CCHMC
Partnership	Informatics / Health Informatics	Host outside speakers on informatics topics	Operational	Informatonist	National Library of Medicine (NLM) Series
Partnership	GIS	Dataset Creation	Operational	Informatonist	Data mining to compile data set for analysis
Partnership	GIS	Data Cleaning/Preparation	Operational	Informatonist	Geocoding, clipping shapefiles, variable specifications
Partnership	GIS	Low level spatial analysis	Operational	Informatonist	Intermediate tasks - buffers, joining datasets, flow analysis, animations
Partnership	GIS	Data Visualization	Operational	Informatonist	Map created for journal article

Appendix B: Data Science Roles for UCL

Goal

The UCL Research Data Services group would like to provide interdisciplinary data science consulting and collaboration services involving data analyses and visualization, and statistical training and assistance.

Methodology

Four members of the UCL Research Data Services Group (Bill McMillan, Tiffany Grant, Emily Kean and Richard Johansen) met to discuss data science roles. Using information provided by Melissa Previtera, the group was tasked with providing an environmental scan of data science roles and practices in other institutions. The group was also tasked with seeking relevant job descriptions to fill a potential "data science like" role in UCL. An initial concern was how would UCL define Data Science given the many definitions and interpretations, and where might we focus our efforts as there are many relevant knowledge domains within the field.

Findings and Noted Gaps in Knowledge and Expertise

Working definition for Data Science at UCL

- Data Science refers to an emerging area of work concerned with the collection, preparation, analysis, visualization, management, and preservation of large collections of information.
    - Definition taken from: *An Introduction to Data Science*. A book developed by Jeffrey Stanton for the Certificate of Data Science program at Syracuse University's School of Information. https://storage2.ischool.syr.edu/media.ischool.syr.edu/oldmedia/documents/2012/3/DataScienceBook1_1.pdf
    - This definition was selected because it encompasses many of the data science associated domains and much of what we already do at UCL.

How do we determine where we put our focus and what the needs are?

- As we begin to consider the provision of data science services, we need to ensure that we are providing services that are desired by our research community. UCL in connection with IT@UC instituted a survey 2 years ago, but new data is now required to determine the exact needs of the UC research community.

Noted Gaps in Knowledge and Expertise

- Staff/faculty statistical/analytical expertise
    - How to extract, transform, clean, summarize and make inferences from data
- Working knowledge of methods to generate reproducible research using tools and pipelines that aid in increasing research transparency
- Finding and accessing appropriate datasets and records

Data Science role models

- *University of Virginia Library StatLab*: *Data Analytics*- https://data.library.virginia.edu/statlab/
    - We offer consultations, collaborations, and training in support of data science, applied statistics, and scientific computing, including data wrangling and cleaning, analysis and

visualization, statistical inference and computational methods, reproducibility and open science.

- Staff expertise also encompasses open source programming languages like R and Python; statistical environments like Stata, SPSS, SAS; scientific computing technologies like Unix, GitHub, and ShareLaTeX; visualization software (Tableau), and software for qualitative data (Dedoose). We can help with data collection via web scraping and APIs in R and Python and working with structured and unstructured (e.g., text) data.

- *University of Michigan CSCAR Consulting-* http://cscar.research.umich.edu/consulting/
  - CSCAR consultants have advanced degrees and years of experience in a variety of areas relating to data analysis and computing with data.
  - SCAR consulting can address a variety of issues arising in the research process, including study design, question formulation, obtaining and managing data, statistical power, sources of bias and uncertainty, data processing and storage, optimization and verification of code, analytic methodologies, and interpretation and presentation of results.

- *John Little, Data Analysis Librarian/Consultant, Data & Visualization Services Department @ Duke University-* https://www.johnlittle.info/
  - I help Duke University students and researchers navigate practical data science challenges. To that end I provide a series of workshops on R and OpenRefine; offer walk-in and by-appointment consultations; and host the R we having fun yet learning series (Rfun).
  - My workshops and presentations include twitter stream gathering, web scraping, data parsing, and data cleaning. Each workshop is designed as a hands-on experience where you can download practice data, presentation slides, and workbook guides. Recorded video streaming is often available.

- *Columbia University Libraries Research Data Services Software and Tools* https://library.columbia.edu/services/research-data-services/software.html
  - The page contains a list of software and tools either available on the machines in the Research Data Service, or tools available online with the level of support available for each.

Recommendations

- Deploy survey regarding data management and data science needs to UC research community
  - Use many of the previous questions but include relevant data science inquires
  - The survey would be deployed using common means, but also a more target approach will be instituted to get the survey in the hands of those who have shown interest in our services.
- Open a *StatLab East* and *StatLab West* to offer consultations, collaborations and training in statistics, data analyses, and visualization. The StatLabs would be staffed by graduate student workers.
- Hire full time data analysis librarian (see job description)
- Provide data science professional development training to researcher services members

Appendix C: Proposed Position Description

POSITION TITLE: Data Analysis Specialist

GENERAL DESCRIPTION: As a member of the Research Data Services Team, the Data Analysis Specialist focuses on high-value research data services for the University of Cincinnati (UC) research community. The Specialist provides leadership and expertise in the areas of quantitative and qualitative data analysis and plays a lead role in developing and executing a research reproducibility support program across the disciplines. This position will offer data analysis and research reproducibility support in the forms of in-depth consultations, workshops, courses, and other targeted outreach and development initiatives. This is a full-time Administrative & Professional (A&P) staff position.

RELATIONSHIP AND AUTHORITY: Reports to _____.

DUTIES AND RESPONSIBILITIES:

- Leads library services related to the discovery, evaluation, and manipulation of data and the use of statistical or other analytical tools for quantitative and qualitative data.
- Support and expand library-based data services that promote the discovery, analysis, and visualization of data.
- Provides interdisciplinary statistical and reproducible research assistance to members of the UC research community.
- Advocate for open access of research data and assist patrons in the discovery and use of existing datasets.
- Develops and codes software programs, algorithms and automated processes to cleanse, integrate and evaluate large datasets from multiple disparate sources.
- Identifies meaningful insights from large data and metadata sources.
- Provides guidance and creates best practices for ethical data creation, production and dissemination.
- Conducts instruction sessions for a variety of users. Assists in developing instructional materials, utilizing pedagogical best practices, and in designing evaluation instruments.

REQUIRED QUALIFICATIONS:

- Bachelor's degree, with three (3) years' experience; -OR- Associate's degree and five (5) years' experience; -OR- seven (7) years' experience; degree must be in a related field

PREFERRED QUALIFICATIONS:

- Master's degree or other advanced degree in computer/library/information science or other relevant field technology, information/computer science, data science, or related field.
- Demonstrated understanding of critical issues regarding reproducible research best practices, research data lifecycle, and scholarly communication issues.
- Knowledge of research related database concepts such as relational databases, database design, and data types.
- Working knowledge of R, SAS, SPSS, STATA, MatLab or other statistical analysis software applications.
- Experience finding, acquiring, manipulating and analyzing quantitative data in a research environment.
- Experience utilizing data analysis and computation tools and visualization methods to create dashboards, infographics, etc.
- Demonstrated time management and organizational skills
- Ability to work independently and on teams in a flexible work environment.
- Familiarity with emerging trends in data management for scholarly communication.

Appendix D: Data Science Training Opportunities

District Data Labs
https://www.districtdatalabs.com/data-science-training/

We can help equip your employees with the analytical and technical skills they need. Our data science training will enhance your team's abilities, as we teach them how to use the latest tools and technologies to work more efficiently, produce better results, and arrive at better data-driven decisions.

Offered Courses:

- Intro to Data Science with Python
- Data Acquisition and Ingestion
- Data Wrangling and Storage
- Statistical Data Analysis
- AI & Machine Learning
- Big Data Machine Learning & Analytics
- Natural Language Processing & Text Analytics
- Network Graph Analytics
- **Foundations of Data Science (Full day course $5,000 plus instructor travel costs) - 20-30 people**
    - **Course Content**
        - What is Data Science?
        - Practitioners of Data Science
        - The Rise of Data Products
        - The Data Science Pipeline
        - Agile Data Science
        - Skills Required to Become a Data Scientist
        - Data Science Tools & Additional Resources

Data Incubator
https://www.thedataincubator.com/fellowship.html

FREE Data Science Fellowship

- The Data Incubator is a Cornell-funded data science training organization. We run a free advanced 8-week fellowship (think data science bootcamp) for PhDs looking to enter industry. A variety of innovative companies partner with The Data Incubator for their hiring and training needs, including LinkedIn, Genentech, Capital One, Pfizer, and many others. The program is free for admitted Fellows. Fellows have the option to participate in the program either in person in New York, San Francisco Bay Area, Boston, Washington DC, or online.

- **Data Science Foundations Online Course** https://www.thedataincubator.com/foundations.html
    - The Data Incubator's Foundations of Data Science online training course is an introductory 8-week, part-time bootcamp geared towards giving ambitious college and graduate-level students, recent college graduates, and working professionals an

immersive hands-on experience with foundational data science techniques. Class sessions are LIVE online presentations, twice each week. Our Foundations of Data Science online training course curriculum has been developed with feedback from our hundreds of industry partners, using the same rigorous methodology as our Fellowship program, to transform data amateurs into data professionals.

- 2 days/week 2 hours/day for 8 weeks
- Will offer a day course starting on September 11th 2-4pm
- Cost $3495/person
  - Early registration - $2895/person

- **Training Curriculum**
  - Taught as 3-5 day workshops, but the content and length are customizable. https://www.thedataincubator.com/curriculum.html
  - Cost $8K/day plus travel expenses
- Custom Tracks https://www.thedataincubator.com/curriculum.html
  - Python Data Science Stack
  - Machine Learning
  - Distributed Computing

Coursera Data Science Specialization
https://www.coursera.org/specializations/jhu-data-science

This Specialization covers the concepts and tools you'll need throughout the entire data science pipeline, from asking the right kinds of questions to making inferences and publishing results. In the final Capstone Project, you'll apply the skills learned by building a data product using real-world data. At completion, students will have a portfolio demonstrating their mastery of the material. Created by Johns Hopkins.

Beginner Specialization – No prior experience required

10 Courses
- The Data Scientist's Toolbox
- R Programming
- Getting and Cleaning Data
- Exploratory Data Analysis
- Reproducible Research
- Statistical Inference
- Regression Models
- Practical Machine Learning
- Developing Data Products
- Data Science Capstone

Data Science and Visualization Institute
https://www.lib.ncsu.edu/data-science-and-visualization-institute

The Data Science and Visualization Institute for Librarians is a week-long course for librarians who are passionate about research and scholarship. Develop knowledge, skills, and confidence to communicate effectively with faculty and student researchers about their data and be able to provide initial consultancy on course topics.

- Immersive learning about data science and visualization in collaboration with academic peers
- Interactive sessions led by expert instructors and focused on mastery of core concepts
- Hands-on exposure to selective open source and highly used commercial tools
- Sharing of practices and experiences across institutions is encouraged

Sessions address topics such as:

- Data Exploration and Analysis
- Data Visualization
- Data Cleaning and Preparation
- Web Scraping
- Parsing HTML & JSON, Orchestrating APIs, and Gathering Twitter Streams
- Bibliometric Network Analysis
- Data Description, Sharing, and Reuse

$2,500 per person. Costs include instruction, breakfasts and lunches each day, and one evening reception. Costs do not include travel and lodging expenses.