The Scholar’s Backpack: Using virtual environments to support modern research practice.

Bret Davidson | Eka Grguric
NCSU Libraries

bretdavidson.github.io/cni-2016
Agenda

- Open science as problem space
- Open science as modern research practice
- Open science at NC State
- Scholar's Backpack
Open Science: what is it?

- Open Access
- Open Data
- Open Notebooks
- Open Source
Open Science is a return to first principles of scientific practice.
PHILOSOPHICAL TRANSACTIONS: GIVING SOME ACCOMPT OF THE PRESENT UNDERTAKINGS, STUDIES, AND LABOURS OF THE INGENIOUS IN MANY CONSIDERABLE PARTS OF THE WORLD.

Vol I.
For Anno 1665, and 1666.

In the SAVOY, Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane; Printers to the Royal Society.
Nullius in Verba

"Take nobody's word for it."
Open Science can increase reproducibility.
Five Schools of Thought
by Sönke Bartling & Sascha Friesike

- Infrastructure
- Public
- Measurement
- Democratic
- Pragmatic
Why Libraries?
Aligns with core library values

- information access
- open peer review
- community-based knowledge creation
- the preservation and dissemination of research
- libraries are champions of open (open source; open data)
Libraries are about supporting their users
Academic Libraries are about supporting research practice
Ongoing disruption by digital technologies in modern research practice
Hypothetical Open Science Workflow

101 Innovations in Scholarly Communication,
https://innoscholcomm.silk.co/
Policy Shifts in support of open
Ecosystem of Support for Modern Research Practice at NCSU Libraries
Research Support

COLLECTIONS
Triposaver, Special Collections, Digital Repository

CITATION MANAGEMENT
RefWorks, Zotero, Mendeley, EndNote

DATA AND GIS
Finding geopolitical and numeric data, GIS software, assistance

DATA MANAGEMENT
DMP review, sharing, & discovery, best practices

GRANTS AND FUNDING
Opportunities, proposal writing, sponsored research compliance

MEASURING RESEARCH IMPACT
Citation analysis, author networks and metrics, benchmarking

PUBLISHING AND COPYRIGHT
Copyright agreements, Open Access, fair use

VISUALIZATION
High-tech spaces, video walls, data visualization

TEXT AND DATA MINING
Datasets, tools and tutorials

CHANCELLOR'S FACULTY EXCELLENCE PROGRAM
Libraries' support for faculty cluster programs

SUBJECT SPECIALISTS

Jeff Edric
Karen DeWitt

Darcia Lewis
Mira Weller

All subject specialists ➤
Visualization Workshops

Infographics: Visualizing Information
November 18
10:00 AM to 12:00 PM
AT THE
ITTC Lab 2, D. H. Hill Library

Getting Started with Data Visualization: Tools for Research!
December 2
11:00 AM to 12:00 PM
AT THE
ITTC Lab 2, D. H. Hill Library

Getting Started with Data Visualization II: Elements of Design
December 2
12:00 PM to 1:00 PM
AT THE
ITTC Lab 2, D. H. Hill Library

R for Absolute Beginners
December 5
1:00 PM to 3:00 PM
AT THE
Makerspace

D. H. HILL LIBRARY
A DIY creation and collaboration space

JAMES B. HUNT JR. LIBRARY
3D Printing services and more
The NCSU Libraries’ Open Science Initiative
Goals

- explore open science practice at NCSU
- better understand researcher needs in context
Take a non-prescriptive user-centered approach.
Create opportunities for communication.
Open Science Unconference 2016

When
Tuesday, March 22, 9:00 a.m. - 5:00 p.m.

Where
Duke Energy Hall 2nd Floor, James B. Hunt Jr. Library

About
NCSU Libraries will be hosting an Open Science Unconference in the Duke Energy Hall at the James B. Hunt Jr. Library on March 22nd, 2016. The unconference will be an informal, participant-driven, event for researchers across NCSU who are interested in open science to meet, discuss, and discover opportunities for collaboration. The unconference structure is loose and collaborative; participants are welcome to help set the agenda, define the outcomes, and develop the deliverables. Hosted by the NCSU Libraries, the unconference will feature breakout sessions and a keynote presentation by Dr. Marcus Hanwell.

Attendance is free but space is limited. To register please fill out this form: go.ncsu.edu/opensci2016_registration

To keep in touch and hear about future events, please join our google group: group-openscience@ncsu.edu

SCHEDULE

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 - 10:00</td>
<td>Coffee and Minging</td>
</tr>
<tr>
<td>10:00 - 11:00</td>
<td>Keynote</td>
</tr>
<tr>
<td>11:00 - 1:15</td>
<td>Breakout Planning, Lunch, and Voting</td>
</tr>
<tr>
<td>1:15 - 4:00</td>
<td>Two Rounds of Breakout Sessions</td>
</tr>
</tbody>
</table>
Follow-up Informal Interviews

- Modern Research Skills Gap
- Insufficient Incentives
Summer of Open Science Event Series

The NCSU Libraries Summer of Open Science is a series of workshops and meetups that support modern research practice through hands on skill building.

Researchers are increasingly using digital tools in a complex, increasingly open, scholarly ecosystem. This has created a technical skills gap for experienced and novice researchers alike. The Summer of Open Science is designed to address this skills gap.

INTRODUCTION TO THE COMMAND LINE INTERFACE
May 19, 2016
2:00 PM to 4:00 PM
At the D.H. Hill Makerspace, D.H. Hill Library

WEB SCRAPING WITH PYTHON
May 25, 2016
2:00 PM to 5:00 PM
At the D.H. Hill Makerspace, D.H. Hill Library

UNDERSTAND AND BUILD YOUR SCHOLARLY IDENTITY
Jun 2, 2016
1:00 AM to 12:00 PM
At the Multimedia Seminar Center, D.H. Hill Library

SCIENTIFIC COMPUTING WITH PYTHON AND RASPBERRY PI
Jun 7, 2016
2:00 PM to 5:00 PM
At the D.H. Hill Makerspace, D.H. Hill Library

BUILD YOUR SCHOLARLY WEBSITE THE EASY WAY
Jun 10, 2016
1:00 AM to 12:00 PM
Goals

- Hands on skill building
- Provide networking opportunities
- Increase visibility of library spaces & services
Skills

- Scholarly identity creation
- Scientific computing
- Building a website
- Data harvesting
- Code collaboration
The Planning Team

Representation from both technical and non-technical departments.

Ekatarina [Eka] Grguric (Project Lead)
NCSU Libraries Fellow, Digital Libraries Initiatives / User Experience

Lauren Di Monte (Project Manager)
NCSU Libraries Fellow, User Experience / Administration

Alison Blaine (Content Development)
NCSU Libraries Fellow, Digital Libraries Initiatives / Research & Information Services

Bret Davidson (Technical Lead)
Digital Technologies Development Librarian, Digital Libraries Initiatives

Jennifer Garrett (Community Development)
Research Librarian for Mgmt, Education, and Social Sciences, Research & Information Services
Summer of Open Science

• Workshops
  ▪ Intro to the Command Line Interface
  ▪ Web Scraping with Python
  ▪ Understand and Build Your Scholarly Identity
  ▪ Scientific Computing with Python & Raspberry Pi
  ▪ Build Your Scholarly Website the Easy Way

• Events
  ▪ Meetups
  ▪ End-of-Summer Showcase
Instructors

Brittany Johnson  Eka Grguric  Lauren DiMonte  Alison Blaine

Madison Sullivan  Will Cross  Todd Stoffer
Scientific Computing with Python & Raspberry Pi

40 person waiting list
Interdisciplinary Need:
over 40 departments across ~16 colleges
Takeaways

• Libraries are well positioned to fill gaps in the curriculum
• "Open Science" attracted a range of disciplines
• High demand for introductory skill training, particularly coding skills (Python)
• Interest in interdisciplinary research sharing
• Summer presents interesting opportunities and challenges
Virtual Environments for Reproducible Computing
Technical workshops are ripe for disaster.
What could go wrong?

- Images reset overnight
- Improper permissions
- Network connectivity issues
- Language Versions
- Missing packages
Instructor Challenges

- Consistency across user environments
- Consistency of course materials
- Time to provision computing environments
- Ease of collaboration
Student Challenges

- Basic data types and structures
- Module system
- Retrieve a web page with Requests
- Parse content with Beautiful Soup
- Generate a word cloud with matplotlib
- Control Structures
- Exception Handling
- Working with file system
Computing Tasks

VS.

Computing Environments
Many Options

- Custom Operating System Images
- Custom Distributions, e.g. Anaconda
- Interactive Environments, e.g. Jupyter
Our Approach

- Vagrant for managing operating system
- Ansible for provisioning and configuration
- Course or lab specific packages and resources
Easy!

1. Install Vagrant
2. Install VirtualBox
3. Clone project repo
4. `vagrant up`
5. `vagrant ssh`
6. Execute code!
This is reproducible computing!
Benefits

- Consistent environment user to user
- Single target for course materials
- Faster provisioning for new workshops
- Repeatable course to course
Rise of Scholarly Code
Researcher Challenges

- Consistency across lab environments
- Ability to see results of code
- Consistency across time
- Ease of collaboration
Scholar's Backpack

Modern research practice asks researchers to engage with information in new ways through the use of a rapidly changing array of digital technologies. The Scholar's Backpack will bring together a sampler of commonly used digital tools that support the research lifecycle in one virtual machine, both decreasing the overhead of locating, installing, and learning how to use new tools and improving the reproducibility of scientific computing environments.
Features

- Python
- R and R Studio
- Jupyter Notebook Server
- Example Notebooks
Vagrant
Create and configure lightweight, reproducible, and portable development environments.
Usage

- Easy installation through binary package
- Flexible configuration via text-based configuration file
- Single command: `vagrant up`
Ansible

"Automation engine" for provisioning and configuration management.
Provisioning

"To make something available."

Installation!
Configuration Management

"Establish and maintain consistency of an environment."
Provisioning

- Text editor
- Python & R
- Git
- Web Browser
- etc.
Configuration

- Start Jupyter notebook server
- Set environment variables
- Set default login directory
Benefits

- Improved consistency
- Ability to see results of code
- Ease of collaboration
Future Work
Richer Environment

- Broader scientific computing
- Improved adherance to best practices
- Docker containers for portability
Embedded Use

- Curricular use
- Laboratory use
Summary
Open Science represents a new framework for research and provides an opportunity for libraries to engage researchers in new ways.
NCSU Libraries has done workshops and outreach around this framework and there is evidence of strong interest across disciplines.
We are redeploying existing technical resources and cutting edge technology in ways that used to be difficult or impossible.
This approach has helped us identify a new leadership role for libraries in open research support.
Thanks!

bret_davidson@ncsu.edu
eka_grguric@ncsu.edu | @egrguric
github.com/NCSU-Libraries/scholars-backpack
bretdavidson.github.io/cni-2016