

# Supporting open science through reproducible computing.

Bret Davidson | Eka Grguric

NCSU Libraries

[go.ncsu.edu/2017-c4l-se](https://go.ncsu.edu/2017-c4l-se)

# 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 with-  
out 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

Editors, <http://book.openingscience.org/>

- 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

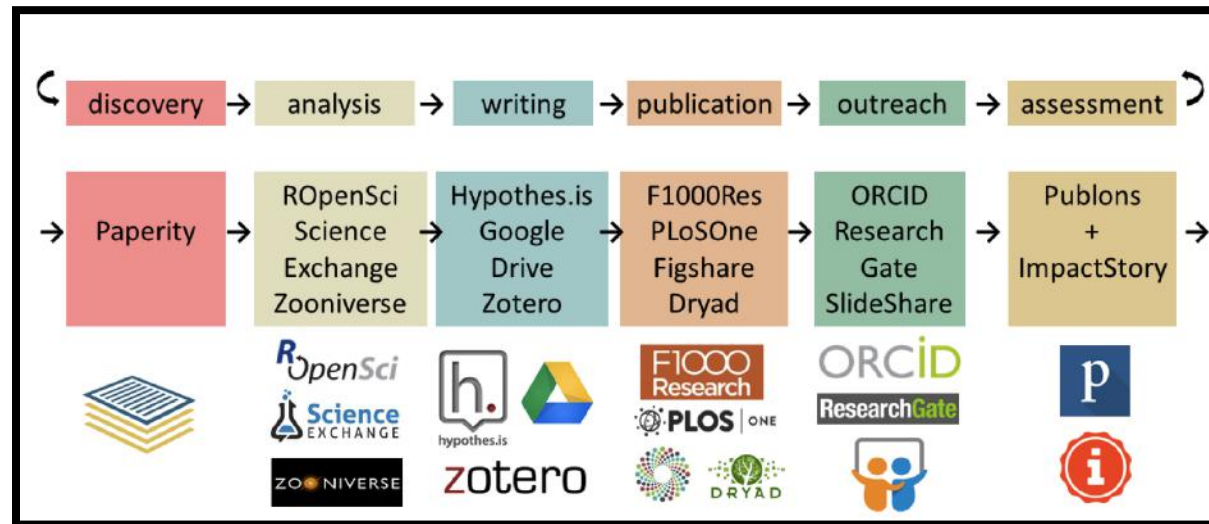
Academic Libraries

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

OECD publishing

Please cite this paper as:

OECD (2015), "Making Open Science a Reality", *OECD Science, Technology and Industry Policy Papers*, No. 25, OECD Publishing, Paris.  
<http://dx.doi.org/10.1787/5jrs2f963zs1-en>



OECD Science, Technology and Industry  
Policy Papers No. 25

**Making Open Science a  
Reality**

OECD

# Ecosystem of Support for Modern Research Practice at NCSU Libraries



# Research Support



## COLLECTIONS

Tripsaver, Special Collections, Digital Repository



## CITATION MANAGEMENT

RefWorks, Zotero, Mendeley, EndNote



## DATA AND GIS

Finding geospatial and numeric data, GIS software assistance



## DATA MANAGEMENT

DMP review, sharing & discovery, best practices



Need Help?



## 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 Essic



Karen DeWitt



Danica Lewis



Mira Waller

All subject specialists >



## WORKSHOPS

All Workshops

Digital Media Workshops

Makerspace Workshops

Ready. Set. Go!

Research Workshops

Visualization Workshops

View other Library  
events. →

« NOVEMBER »						
S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

Special

# Visualization Workshops

NOV

18

## Infographics: Visualizing Information

November 18  
10:00 AM to 12:00 PM

AT THE

ITTC Lab 2, D. H. Hill Library

DEC

2

## 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

DEC

2

## 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

DEC

5

## R for Absolute Beginners

December 5  
1:00 PM to 3:00 PM

AT THE

# Makerspace

## D. H. HILL LIBRARY

A D-I-Y creation and collaboration space



## JAMES B. HUNT JR. LIBRARY

3D Printing services and more





**WOLFPACK  
CITIZEN  
SCIENCE  
CHALLENGE**

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

[NC State Home](#)[RESOURCES](#)[search ncsu.edu](#)

[FIND](#)[GET HELP](#)[SERVICES](#)[LIBRARIES](#)[ABOUT](#)

[ASK US](#)[MY ACCOUNT](#)[HOURS](#)[FAQ](#)[LOG OUT](#)[CHAT NOW](#)

**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](http://go.ncsu.edu/opensci2016_registration)

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

**SCHEDULE**

- 9:00 - 10:00 Coffee and Mingling
- 10:00 - 11:00 Keynote
- 11:00 - 1:15 Breakout Planning, Lunch, and Voting
- 1:15 - 4:00 Two Rounds of Breakout Sessions



*Open Science Logo, Greg Emmerich, CC-BY-SA-2.0*

# Follow-up Informal Interviews

- Modern Research Skills Gap
- Insufficient Incentives



## EVENTS

All Events & Exhibits  
All Workshops  
Events Calendar  
Exhibits

## EVENT SERIES

Coffee & Viz  
**Summer of Open Science**  
Fabulous Faculty  
Amazing Alumni  
Stellar Students  
AV Geeks at the Hunt Library  
Read Smart  
Making Space

## WORKSHOPS

All Workshops  
Visualization Workshops  
Makerspace Workshops  
Research Workshops  
Digital Media Workshops

* MAY *						
S	M	T	W	T	F	S
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

# 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.



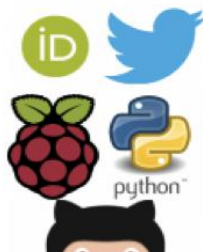
## INTRODUCTION TO THE COMMAND LINE INTERFACE

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  
10:00 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  
10:00 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 broad range of 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



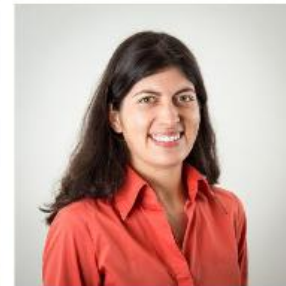
Brittany Johnson



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
- 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

- Inconsistent user environments
- Inconsistent course materials
- Provisioning is time consuming
- Difficult to collaborate

# Student Challenges

- Data types and structures
- Module system
- Control Structures
- Exception Handling
- Working with file system
- Retrieve a web page with Requests
- Parse content with BeautifulSoup
- Generate a word cloud with matplotlib

Computing Tasks  
vs.  
Computing Environments

# Rise of Scholarly Code

# Researcher Needs

- Consistency across lab environments
- Ability to see results of code
- Consistency across time
- Ease of collaboration

# Many Options

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

github.com/NCSU-Libraries/scholars-backpack

The screenshot shows the GitHub repository page for `NCSU-Libraries / scholars-backpack`. The repository has 7 watchers, 0 stars, and 0 forks. The main navigation bar includes links for Code, Issues (0), Pull requests (0), Projects (0), Wiki, Pulse, Graphs, and Settings. Below the navigation bar, there is a section for repository statistics: 1 commit, 1 branch, 0 releases, 1 contributor, and MIT license. A bar at the bottom of this section allows switching branches (currently master) and includes buttons for 'New pull request', 'Create new file', 'Upload files', 'Find file', and 'Clone or download'. The file list shows the initial commit by `bretdavidson` 23 seconds ago, including files: `ansible`, `code`, `.gitignore`, `LICENSE`, `README.md`, and `Vagrantfile`. The `README.md` file is selected, showing the title 'Scholar's Backpack' and a paragraph of text.

NCSU-Libraries / scholars-backpack

Unwatch 7 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Pulse Graphs Settings

No description or website provided. — Edit

1 commit 1 branch 0 releases 1 contributor MIT

Branch: master New pull request Create new file Upload files Find file Clone or download

bretdavidson initial commit Latest commit 2bb566f 23 seconds ago

ansible	initial commit	23 seconds ago
code	initial commit	23 seconds ago
.gitignore	initial commit	23 seconds ago
LICENSE	initial commit	23 seconds ago
README.md	initial commit	23 seconds ago
Vagrantfile	initial commit	23 seconds ago

README.md

## 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.

# Our Approach

- Vagrant for managing operating system
- Ansible for provisioning and configuration
- Anaconda for managing environments and packages
- Workshop specific 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!

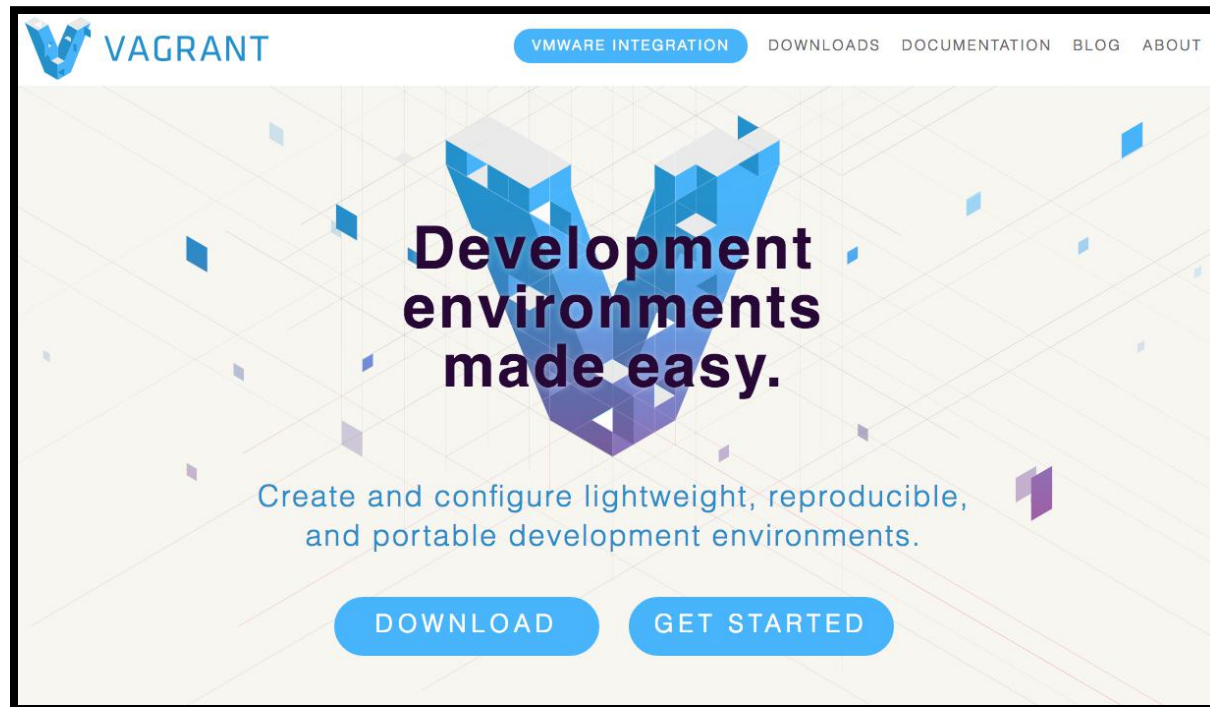
# Benefits

- Consistent environment user to user
- Single target for course materials
- Faster provisioning for new workshops
- Repeatable course to course

# Features

- Python
- R and R Studio
- Jupyter Notebook Server
- Example Notebooks
- Accessible from web browser

# 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

"Installation!"

# Configuration Management

"Establish and maintain **consistency** of an environment."

# Provisioning

- Anaconda
- Python & R
- Software packages
- Jupyter Notebooks
- etc.


# Configuration

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

# Anaconda

**ANACONDA**  
Leading Open Data Science Platform  
Powered by Python

**ANACONDA MAKES...**



**DATA SCIENCE TEAMS  
HAPPIER**



That means better and more results


# Python Packages












astropy, beautifulsoup4, conda, flask, jupyter, matplotlib, numpy, nltk, pandas, pillow, pip, pytest, qt, requests, scipy, scikit-learn, seaborn, sqlite, etc.

# R Packages

r, essentials, formatr, ggplot2, irkernel, knitr, kernsmooth, maps, markdown, mass, matrix, nnet, rbokeh, recommended, spatial, tidyr, etc.

 **jupyter** **Getting-Started** Last Checkpoint: Last Friday at 8:12 AM (autosaved) 

File Edit View Insert Cell Kernel Widgets Help Python 3 

         Markdown   CellToolbar

## Welcome to the Scholar's Backpack!

This notebook is intended to orient you to the environment and help you get started working with it.

The Github repository for the Scholar's Backpack is located here, along with a useful README for setting up future virtual environments: <https://github.com/NCSU-Libraries/Scholars-Backpack> .

## Getting Started

If you are reading this, your virtual machine is set up and you are able to interact with it. That means that you are looking at a Graphical User Interface (GUI) for the Centos 7 operating system, a Linux environment.

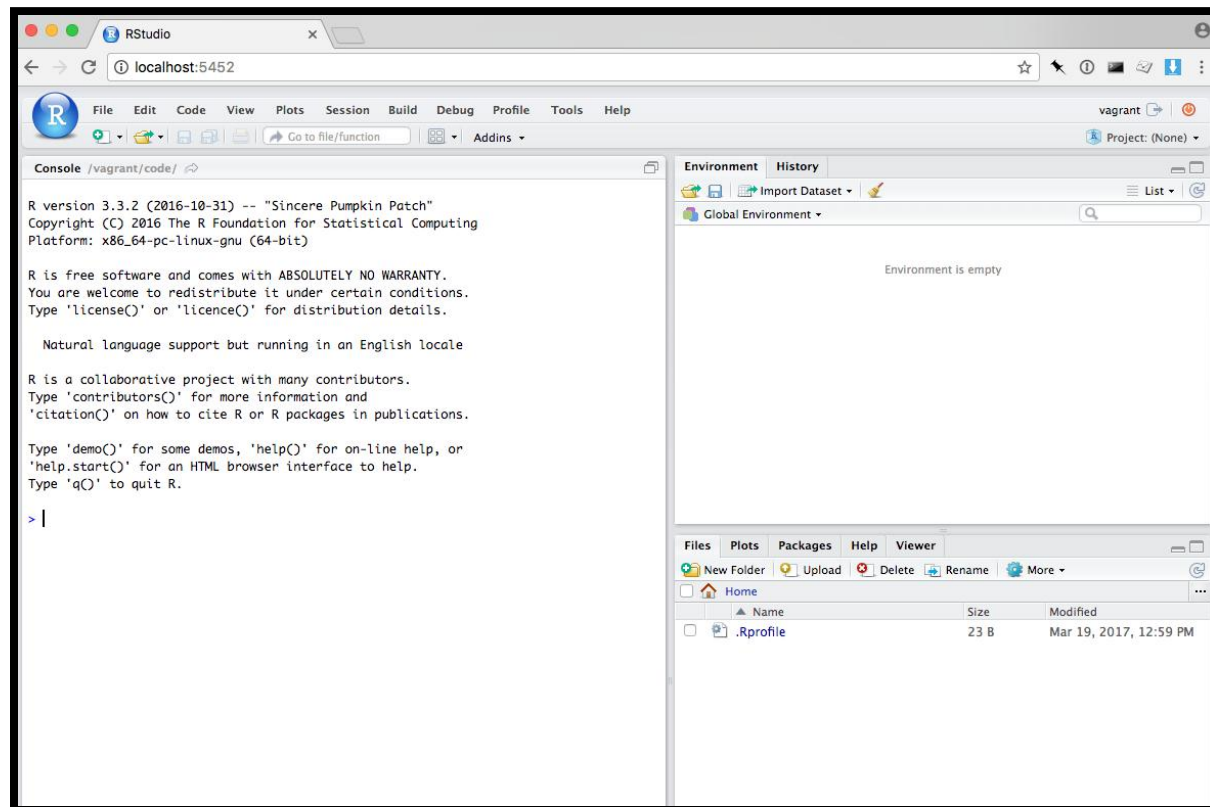
On the desktop there are two links to take note of:

- Getting Started (this notebook)
- Jupyter Notebooks (select Jupyter notebooks that orient you to different tools that are present in the environment; you can make your own from here)

There are also quick links to:

- RStudio Desktop
- The "code" directory (your working folder on start-up)





# Future Work

- Docker containers for portability
- Embedded use in curriculum
- Additional open source contributions

# 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/NCSSU-Libraries/scholars-backpack](https://github.com/NCSSU-Libraries/scholars-backpack)

[go.ncsu.edu/2017-c4l-se](https://go.ncsu.edu/2017-c4l-se)