Tools for a Reproducible Workflow

Garret Christensen[1]

[1]UC Berkeley:
Berkeley Initiative for Transparency in the Social Sciences
Berkeley Institute for Data Science

RT2, September 2017
Slides available online at
`https://github.com/BITSS/RT2London2017`

BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

Workflow

Christensen
Introduction
Workflow
Conclusion

"Reproducibility is just collaboration with people you don't know, including yourself next week"
—Philip Stark, UC Berkeley Statistics

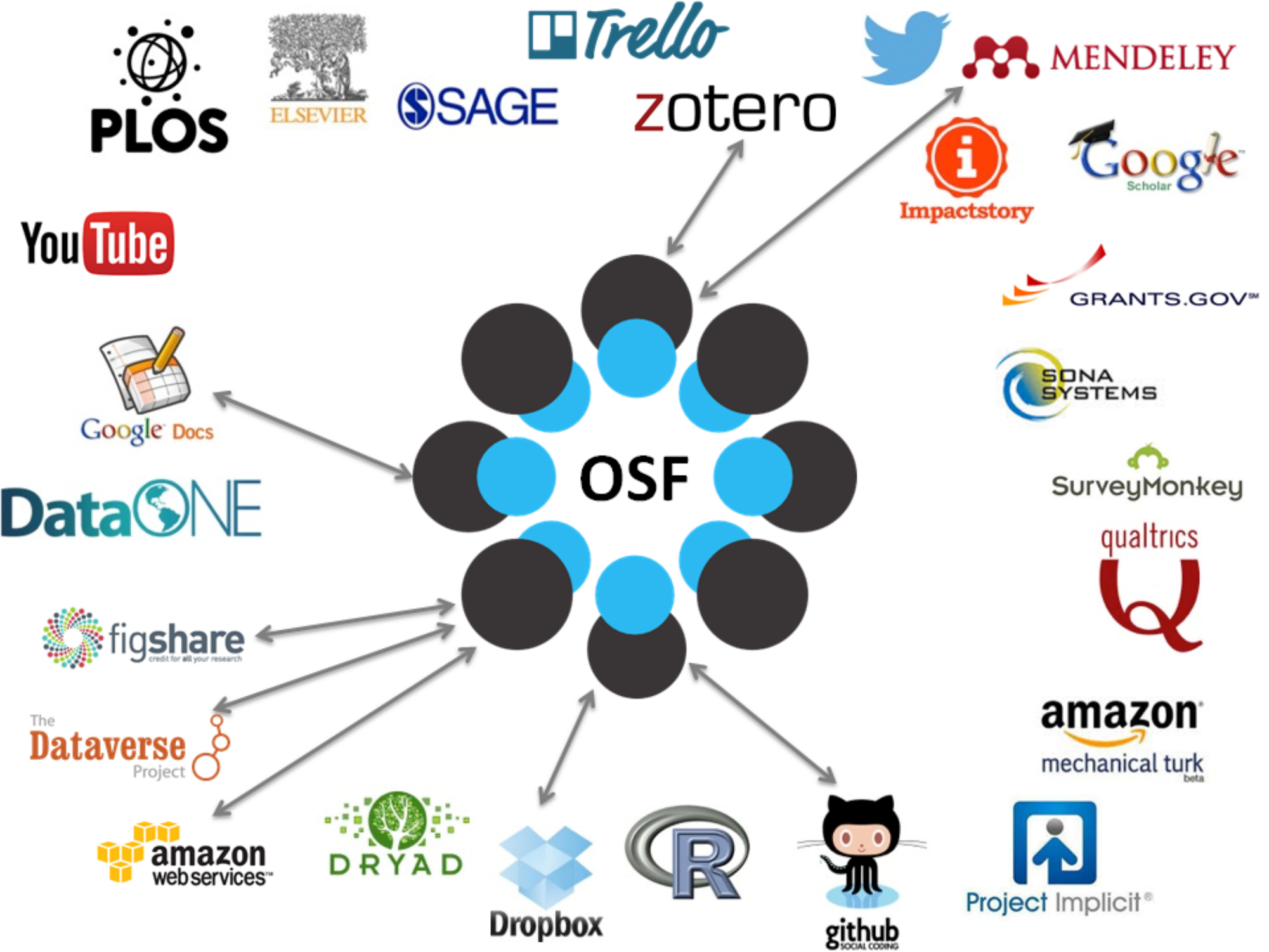
Workflow

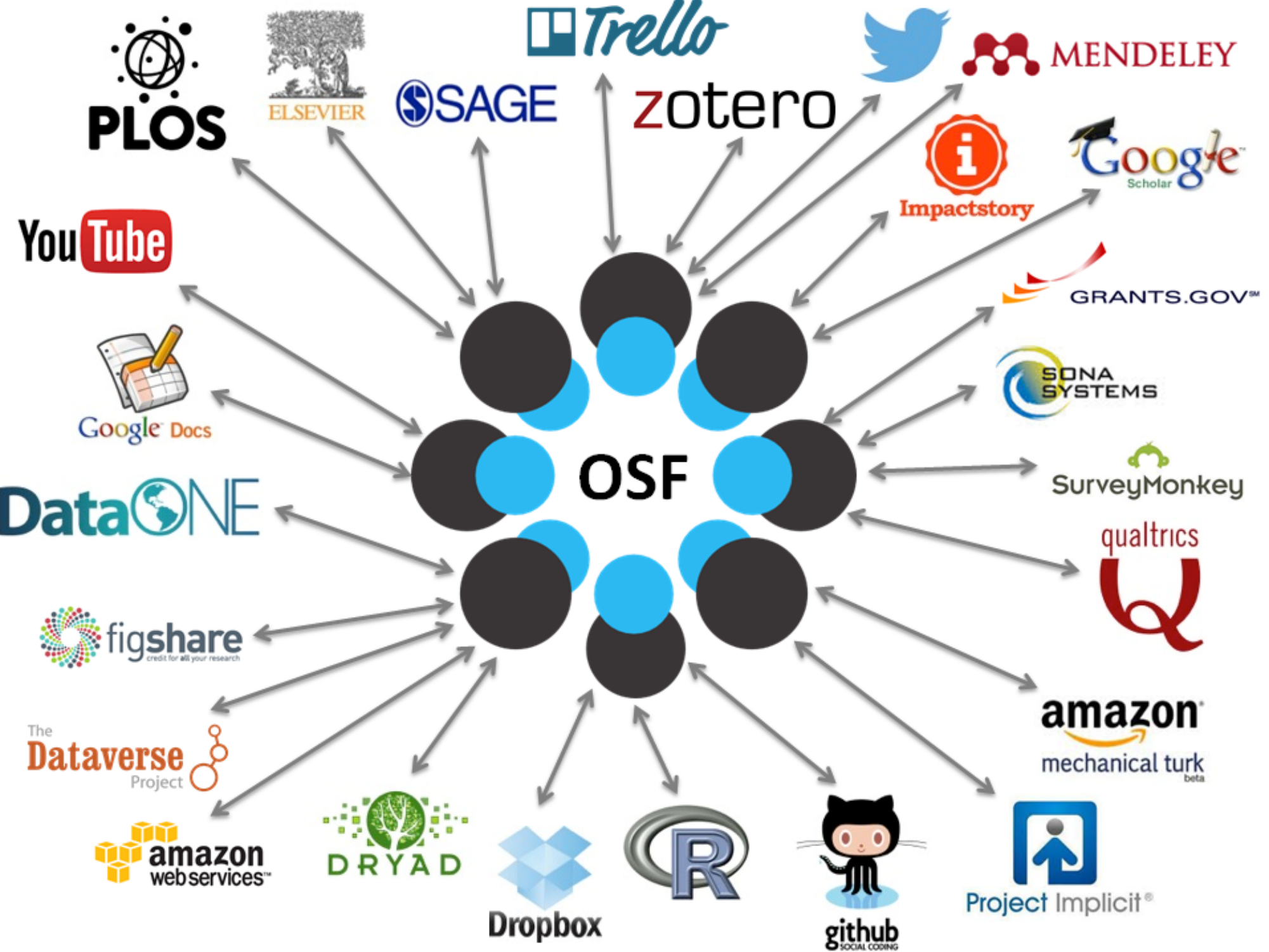
Christensen
Introduction
Workflow
Conclusion

- OSF
- Version Control
- Dynamic Documents

Put your work all in one place with the Open Science Framework Link

- Pre-Registration
- Data
  - Host
  - Link to Dataverse
- Version Control
- More to Come


PLOS
ELSEVIER
Trello
SAGE
zotero
MENDELEY
Impactstory
Google Scholar
YouTube
GRANTS.GOV
Google Docs
SONA SYSTEMS
OSF
DataONE
SurveyMonkey
qualtrics
figshare
The Dataverse Project
amazon mechanical turk beta
amazon web services
DRYAD
Dropbox
github
Project Implicit


PLOS
ELSEVIER
Trello
SAGE
zotero
MENDELEY
Impactstory
Google Scholar
YouTube
GRANTS.GOV
Google Docs
SONA SYSTEMS
OSF
DataONE
SurveyMonkey
qualtrics
figshare
credit for all your research
amazon mechanical turk beta
The Dataverse Project
amazon web services
DRYAD
Dropbox
github SOCIAL CODING
Project Implicit

Dynamic Documents

Write your code and your paper in the same file so you won't lose information or make copy and paste mistakes.

- Include tables by linking to a file, instead of a static image.
- Include number by linking to a value calculated by an analysis file, instead of a static number typed manually.
- Automatically update tables and numbers.
- Produce entire paper with one or two clicks.

Dynamic Documents

Possible in Python, R, and to a lesser extent, Stata

- Jupyter—several (many?) languages
- R—use R Studio to manage projects with built-in version control, and R Markdown/knitr for publication-quality dynamic documents.
- Stata–combine with LaTeX for two click workflow
- Stata–use ‘markdoc’ ado for some dynamic ability.

R
Studio®

jupyter

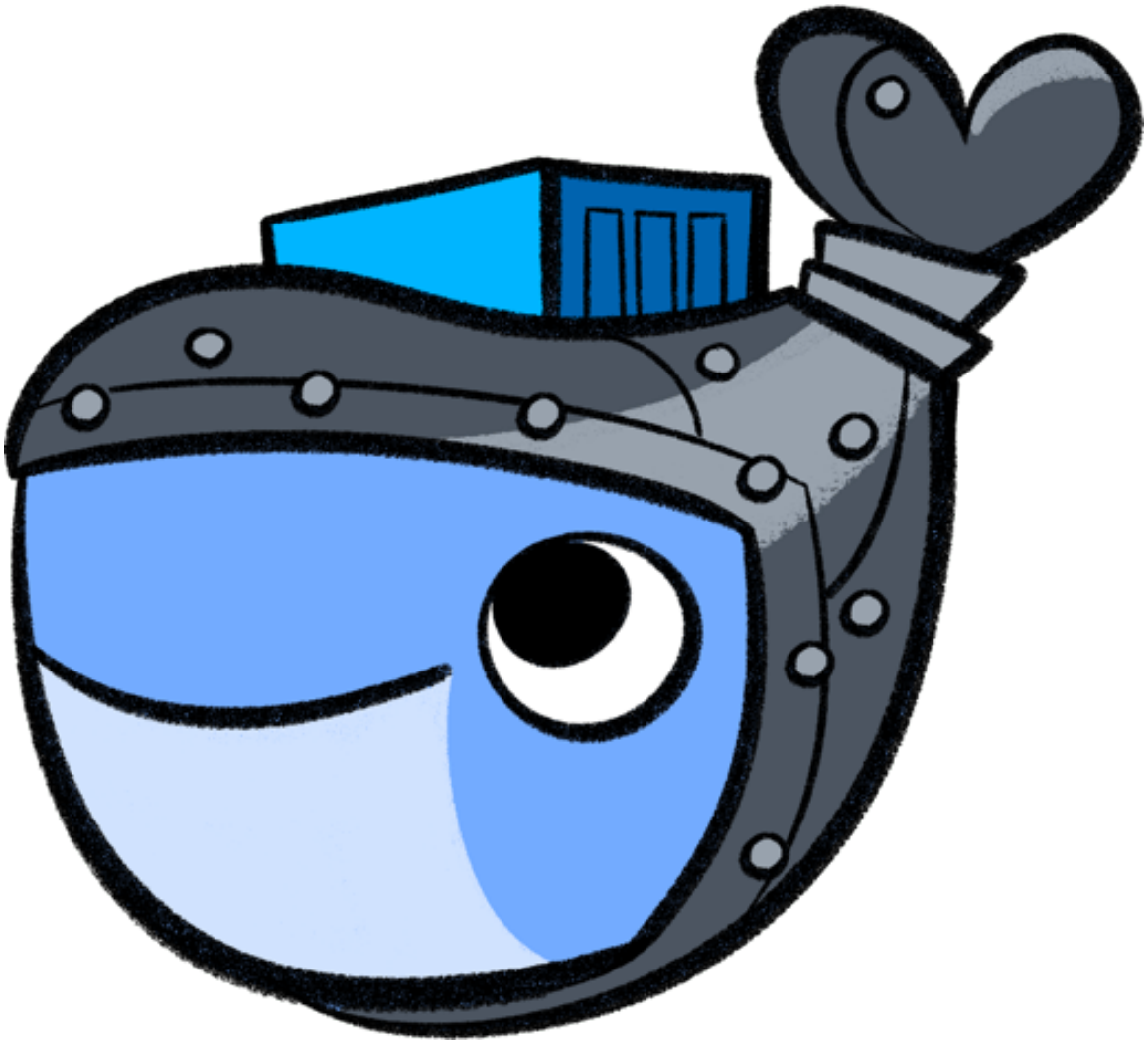
Christensen
Introduction
Workflow
Conclusion

Try them online:

- Jupyter
- R

For the hardcore

Christensen
Introduction
Workflow
Conclusion



Conclusion

Christensen
Introduction
Workflow
Conclusion

OK, I'm convinced. How do I learn more?
Seriously, how do I learn more?!
I need to know.

- Work through my demos. ▸ Link
- Software Carpentry's tutorials ▸ Link