Mistakes in Research

“Mistakes are the portals of discovery.” —James Joyce
Mistakes in Research

▶ “Mistakes are the portals of discovery.” —James Joyce
▶ “Science, my lad, is made up of mistakes, but they are mistakes which it is useful to make, because they lead little by little to the truth.” —Jules Verne
Mistakes in Research

Jeff Rouder and Mike Watson
Working in the Cloud? Best Practices for Sharing Data and Writing Collaboratively
Mistakes in Research

- **Bad Mistakes** are mistakes we should **Never** make:
Mistakes in Research

- **Bad Mistakes** are mistakes we should **Never** make:
  - Submitting the wrong version of draft
Mistakes in Research

- **Bad Mistakes** are mistakes we should **Never** make:
  - Submitting the wrong version of draft
  - Analyzing the wrong data set
Mistakes in Research

- **Bad Mistakes** are mistakes we should *Never* make:
  - Submitting the wrong version of draft
  - Analyzing the wrong data set
  - Incorporating the wrong figure

Jeff Rouder and Mike Watson

*Working in the Cloud? Best Practices for Sharing Data and Writing Collaboratively*
Mistakes in Research

- **Bad Mistakes** are mistakes we should **Never** make:
  - Submitting the wrong version of draft
  - Analyzing the wrong data set
  - Incorporating the wrong figure
- The goal is to eliminate the bad mistakes so we can go on making the good mistakes.
Research Is Collaborative

Jeff Rouder and Mike Watson

Working in the Cloud? Best Practices for Sharing Data and Writing Collaboratively
Research Is Collaborative

- Develop Experiments Jointly, Share Data
Research Is Collaborative

- Develop Experiments Jointly, Share Data
- Develop Analyses Jointly, Share Metadata
Research Is Collaborative

- Develop Experiments Jointly, Share Data
- Develop Analyses Jointly, Share Metadata
- Develop Papers Jointly, Share Documents
Research Is Collaborative

Collaborative Research Enhances The Chances of Bad Mistakes
Bad Mistakes
Bad Mistakes

- I Hate Bad Mistakes More Than Just About Anyone
I Hate Bad Mistakes More Than Just About Anyone
Left to My Own Devices, I Make Bad Mistakes More Than Just About Anyone
Bad Mistakes

- I Hate Bad Mistakes More Than Just About Anyone
- Left to My Own Devices, I Make Bad Mistakes More Than Just About Anyone
- Good Work Processes
Bad Mistakes

- I Hate Bad Mistakes More Than Just About Anyone
- Left to My Own Devices, I Make Bad Mistakes More Than Just About Anyone
- Good Work Processes
- Bad mistakes results from bad work processes, not from personal failings
Bad Mistakes

- I Hate Bad Mistakes More Than Just About Anyone
- Left to My Own Devices, I Make Bad Mistakes More Than Just About Anyone
- Good Work Processes
- Bad mistakes results from bad work processes, not from personal failings
- Bad mistakes should result in better work processes
What Should I Pay Attention To

Obviously,
What Should I Pay Attention To

Obviously,

▶ Backup
What Should I Pay Attention To

Obviously,

- Backup
- Security
What Should I Pay Attention To

Obviously,
- Backup
- Security
- Convenience in Usage

Jeff Rouder and Mike Watson
Working in the Cloud? Best Practices for Sharing Data and Writing Collaboratively
What Should I Pay Attention To

Obviously,

- Backup
- Security
- Convenience in Usage
- Easy to Set Up
What Should I Pay Attention To

Not So Obvious: **Institutional Sponsorship**
What Should I Pay Attention To

Not So Obvious: Institutional Sponsorship

- Input into policies
What Should I Pay Attention To

Not So Obvious: Institutional Sponsorship
- Input into policies
- Not responsible for security
What Should I Pay Attention To

Not So Obvious: **Institutional Sponsorship**

- Input into policies
- Not responsible for security
- Support
What Should I Pay Attention To

Not So Obvious: **Institutional Sponsorship**
- Input into policies
- Not responsible for security
- Support
- Compliance with granting agencies
What Should I Pay Attention To

Without Institutional Sponsorship:

Real Facebook Post: “Apparently, Google drive doesn’t believe that my Bayes workshop is ‘appropriate’, and have removed the file(s) from Google drive. Frequentist conspiracy, or idiotic copyright-flagging spider?”
What Should I Pay Attention To

Not So Obvious: **Versioning**
What Should I Pay Attention To

Not So Obvious: **Versioning**

- Keep all changes as revisions
What Should I Pay Attention To

Not So Obvious: **Versioning**
- Keep all changes as revisions
- Method of identifying changes, merging changes, backing out of changes
What Should I Pay Attention To

Not So Obvious: **Versioning**

- Keep all changes as revisions
- Method of identifying changes, merging changes, backing out of changes
What Should I Pay Attention To

Not So Obvious: Versioning

- Keep all changes as revisions
- Method of identifying changes, merging changes, backing out of changes
- A record of who did what, where, when, maybe why (enforced logging)
What Should I Pay Attention To

Not So Obvious: **Versioning**

- Keep all changes as revisions
- Method of identifying changes, merging changes, backing out of changes
- A record of who did what, where, when, maybe why (enforced logging)
- Well-defined before-the-fact workflow
What Should I Pay Attention To

Not So Obvious: Versioning

▶ Keep all changes as revisions
▶ Method of identifying changes, merging changes, backing out of changes
▶ A record of who did what, where, when, maybe why (enforced logging)
▶ Well-defined before-the-fact workflow
▶ Ability to work asynchronously without communication
Email: The Good

Jeff Rouder and Mike Watson

Working in the Cloud? Best Practices for Sharing Data and Writing Collaboratively
Email: The Good

- Really, Really Easy
Email: The Good

- Really, Really Easy
- Institutional Support & Backup
Email: The Bad
Email: The Bad

- No organization, Proliferation of files, Idiosyncratic naming: “What Is That?”
Email: The Bad

- No organization, Proliferation of files, Idiosyncratic naming: “What Is That?”
- Hard to search: “Where Is That Attachment”
Email: The Bad

- No organization, Proliferation of files, Idiosyncratic naming: “What Is That?”
- Hard to search: “Where Is That Attachment”
- Not Necessarily Reliable: Sent to Junk, Read and Forgotten
Email: The Bad

- No organization, Proliferation of files, Idiosyncratic naming: “What Is That?”
- Hard to search: “Where Is That Attachment”
- Not Necessarily Reliable: Sent to Junk, Read and Forgotten
- Horrible for big files
Email: The Bad

- No organization, Proliferation of files, Idiosyncratic naming: “What Is That?”
- Hard to search: “Where Is That Attachment”
- Not Necessarily Reliable: Sent to Junk, Read and Forgotten
- Horrible for big files
- Fairly expensive from an institutional standpoint
Email: The Ugly

Email is Perfect For Making Bad Mistakes.
DropBox, Google Drive: The Good
DropBox, Google Drive: The Good

- Really Easy
DropBox, Google Drive: The Good

- Really Easy
- Snapshot backup
DropBox, Google Drive: The Good

- Really Easy
- Snapshot backup
- Organized as your filesystem
DropBox, Google Drive: The Good

- Really Easy
- Snapshot backup
- Organized as your filesystem
- Some minimal versioning capabilities
DropBox, Google Drive: The Bad

- Not enough versioning capabilities
DropBox, Google Drive: The Bad

- Not enough versioning capabilities
- No Institutional Umbrella
DropBox, Google Drive: The Bad

- Not enough versioning capabilities
- No Institutional Umbrella
- Not sure of backup, security, viruses, copyright spiders, etc.
DropBox, Google Drive: The Ugly
DropBox, Google Drive: The Ugly

- Too easy to overwrite yours or your colleague’s work.
DropBox, Google Drive: The Ugly

- Too easy to overwrite yours or your colleague’s work.
- Too hard to detect these conflicts and too hard to fix them if detected
DropBox, Google Drive: The Ugly

- Too easy to overwrite yours or your colleague’s work.
- Too hard to detect these conflicts and too hard to fix them if detected
- Bad mistakes won’t happen as often as with email, but it is not a truly safe method
Meet box.com

- MU is flirting with Box (box.com), a Dropbox clone.
Meet box.com

- MU is flirting with Box (box.com), a Dropbox clone.
- Convenience of Dropbox
Meet box.com

- MU is flirting with Box (box.com), a Dropbox clone.
- Convenience of Dropbox
- Institutional Umbrella
Meet box.com

- MU is flirting with Box (box.com), a Dropbox clone.
- Convenience of Dropbox
- Institutional Umbrella
- Better versioning than Dropbox or Google Drive, but still not ideal
Version Control Systems: The Good

SVN, Bazaar, git, Mercurial
Version Control Systems: The Good

SVN, Bazaar, git, Mercurial

- Designed For Versioning, Extensive Versioning Capabilities
Version Control Systems: The Good

SVN, Bazaar, git, Mercurial
- Designed For Versioning, Extensive Versioning Capabilities
- Used for the development of programming code distributed across tens or hundreds of team members
Version Control Systems: The Good

SVN, Bazaar, git, Mercurial

▶ Designed For Versioning, Extensive Versioning Capabilities
▶ Used for the development of programming code distributed across tens or hundreds of team members
▶ Adapt to the research setting for avoiding bad mistakes.
Version Control Systems: The Bad
Version Control Systems: The Bad

- Hard to use
Version Control Systems: The Bad

- Hard to use
- Hard to get a colleague to use (though not as hard as getting a colleague to use LaTeX).
Version Control Systems: The Bad

- Hard to use
- Hard to get a colleague to use (though not as hard as getting a colleague to use LaTeX).
- Need a server
Version Control Systems: The Bad

- Hard to use
- Hard to get a colleague to use (though not as hard as getting a colleague to use LaTeX).
- Need a server
- Need someone to administer the server
Version Control Systems: The Ugly

- Not for everyone, not even for most of us...yet.
Meet collaborate.missouri.edu

- Research-dedicated git server
Meet collaborate.missouri.edu

- Research-dedicated git server
- Web-based graphical interface for server and clients
Meet collaborate.missouri.edu

- Research-dedicated git server
- Web-based graphical interface for server and clients
- Institutional Umbrella

Jeff Rouder and Mike Watson
Working in the Cloud? Best Practices for Sharing Data and Writing Collaboratively
Meet collaborate.missouri.edu

- Research-dedicated git server
- Web-based graphical interface for server and clients
- Institutional Umbrella
- Not easy
Challenges
Challenges

- No good answer yet, we can't have our cake and will need to make some trade-offs
Challenges

- No good answer yet, we can't have our cake and will need to make some trade-offs
- Two reasonable solutions coming: box.com under MU and collaborate.missouri.edu git server
Challenges

- No good answer yet, we can’t have our cake and will need to make some trade-offs
- Two reasonable solutions coming: box.com under MU and collaborate.missouri.edu git server
- Efforts needed into making easier methods safer and harder methods easier.

Jeff Rouder and Mike Watson
Working in the Cloud? Best Practices for Sharing Data and Writing Collaboratively
How Do You Get Help With All This

- http://doit.missouri.edu/it-pro/
- Local IT/graduate/post doc support
- http://doit.missouri.edu/research/
Possible Support Models

- http://adn.missouri.edu/
- http://ircf.rnet.missouri.edu:8000/

Jeff Rouder and Mike Watson
Working in the Cloud? Best Practices for Sharing Data and Writing Collaboratively
Discussion

- What will work best for you?
- How can we collaborate to increase your productivity?
- What is missing in our current support structure?