

GitHubで行う reproducible research

瓜生真也

横浜国立大学大学院 環境情報学府

GitHub Kaigi(@Shibuya), June 1, 2014

github.com/uribo



論文

論文

論文 書き方

論文 英語

論文 引用

論文 参考文献

論文 構成

論文 捏造

論文 コピペ

論文博士

論文とは

Google 検索

I'm Feeling Lucky



Reproducible Research (RR)

再現可能な研究

論文に書かれている方法、条件で行えば、
同様の結果が得られる

- ✓ 客観性
- ✓ 明瞭性
- ✓ 妥当性



開発者と科学者で共通するもの

Prlić, A., and J. B. Procter. 2012. PLOS Computational Biology 8

OPEN ACCESS Freely available online



Editorial

Ten Simple Rules for the Open Development of Scientific Software

Andreas Prlić^{1*}, James B. Procter²

¹ San Diego Supercomputer Center, University of California San Diego, La Jolla, California, United States of America, ² School of Life Sciences Research, College of Life Sciences, University of Dundee, Dundee, Scotland, United Kingdom

1. Don't Reinvent the Wheel
2. Code Well
3. Be Your Own User
4. Be Transparent
5. Be Simple
6. Don't Be a Perfectionist
7. Nurture and Grow Your Community
8. Promote Your Project
9. Find Sponsors
10. Science Counts

研究の流れ

統計解析と論文の執筆は鬼門



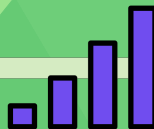
先行研究
を調べる



研究計画
を練る



データ
を得る



統計・解析
をする



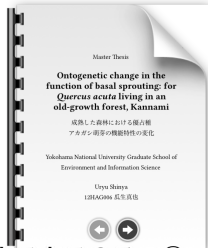
投稿論文
を書く



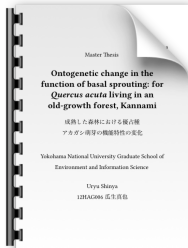
あるある？その①

解析のためのフォルダは
魑魅魍魎の世界...

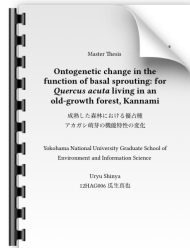




あるある？その②
Quac_draft_v0.pdf

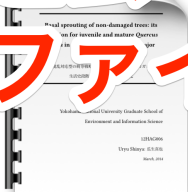


Quac_draft_v1.pdf



Quac_draft_v2.pdf

過去の（使わない）
ファイルが多量に存在



Quac_draft_v3.pdf



Quac_draft_v4.pdf



Quac_draft_v5.pdf

GitHub の特徴を活かした RR な研究

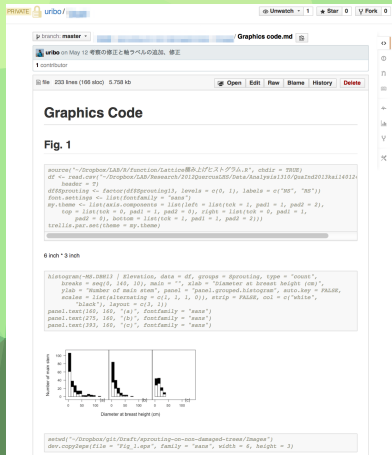
ソーシャルコーディングの概念を研究に...



- ▶ テキストファイルをもとにした進行
- ▶ 進捗状況に応じたブランチの切り替え
- ▶ 用途を記述したコミットコメント
- ▶ バックアップ、差分確認は気にしない

データ解析:

リポジトリ内のソースコード、データファイルを用いれば同様の結果が得ることができる

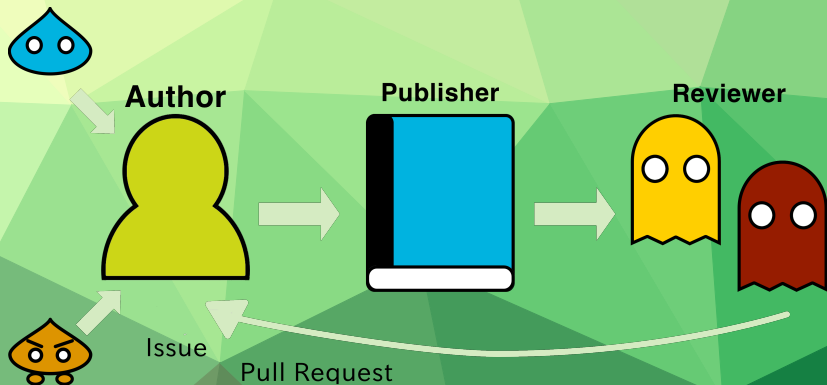


論文執筆: L^AT_EX

テキストファイルなので差分の確認が容易。
必要があればバックアップから復元

[illegible]

Pull Request, Issues を使った査読



今後の展望と課題

- ▶ Computational Science だけでなく、いかなる分野の研究にも reproducible research の概念が必要
- ▶ Data analysis と Writing に GitHub が活用できる
- ▶ ソーシャルコーディングの概念は研究の再現性だけでなく、客観性や明瞭性を向上させる

今後の展望と課題

- ▶ GitHub に生データを上げるのはどうなの？安全？
- ▶ 査読は匿名、というのが原則... やめちまえ
- ▶ ソーシャルコーディングとの矛盾...