

Reproducibility in an Office World

*A Brief History of Failures
and the Odd Success*

Noam Ross,
R-Ladies NYC
2018-11-06
@noamross

First: Have you voted?

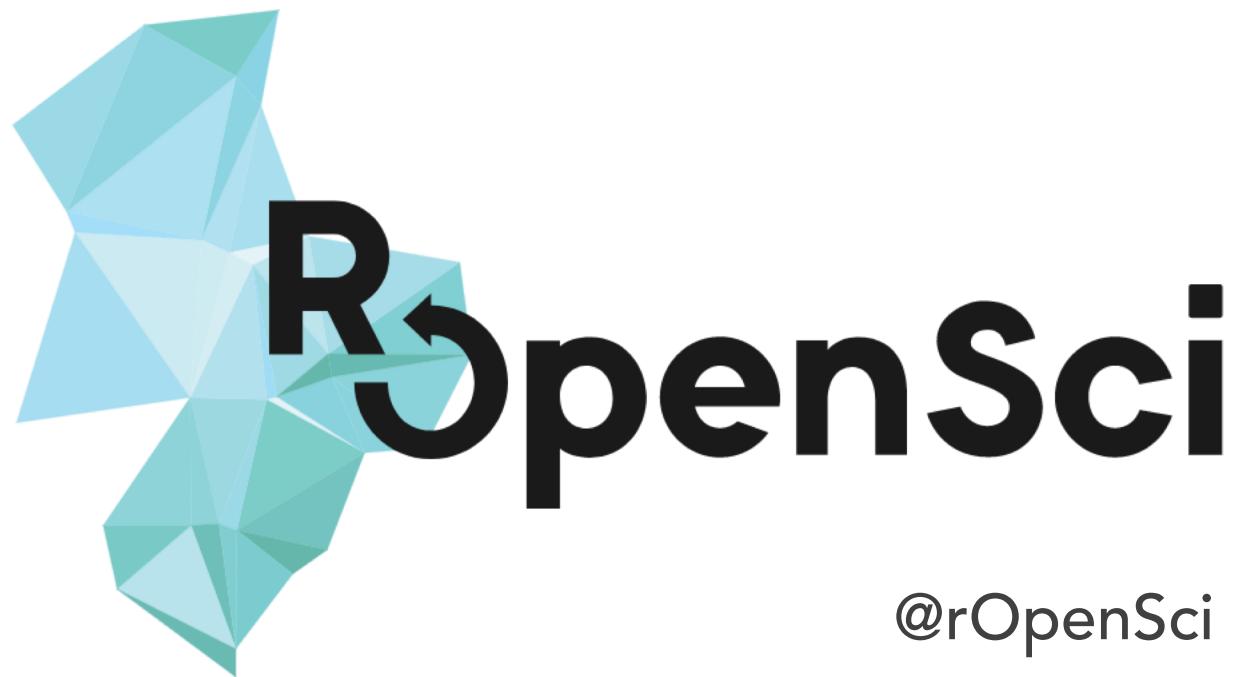
It's OK, you have two more hours.

Leave now.



EcoHealth
Alliance

@EcoHealthNYC



@rOpenSci

The Two Realms of Workflows

Programmatic

Linear

Centralized

Automated

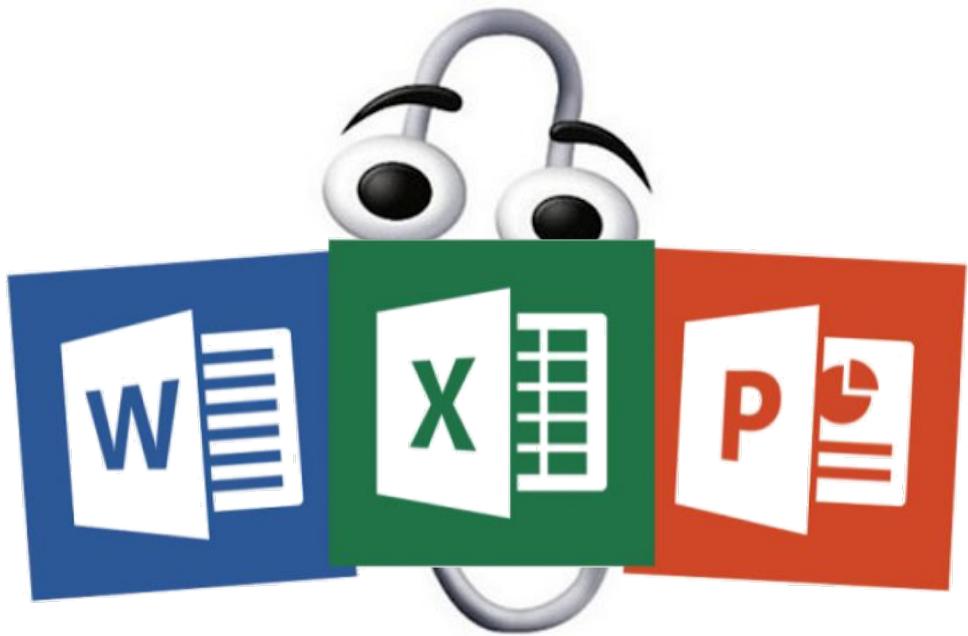
Interactive

Iterative

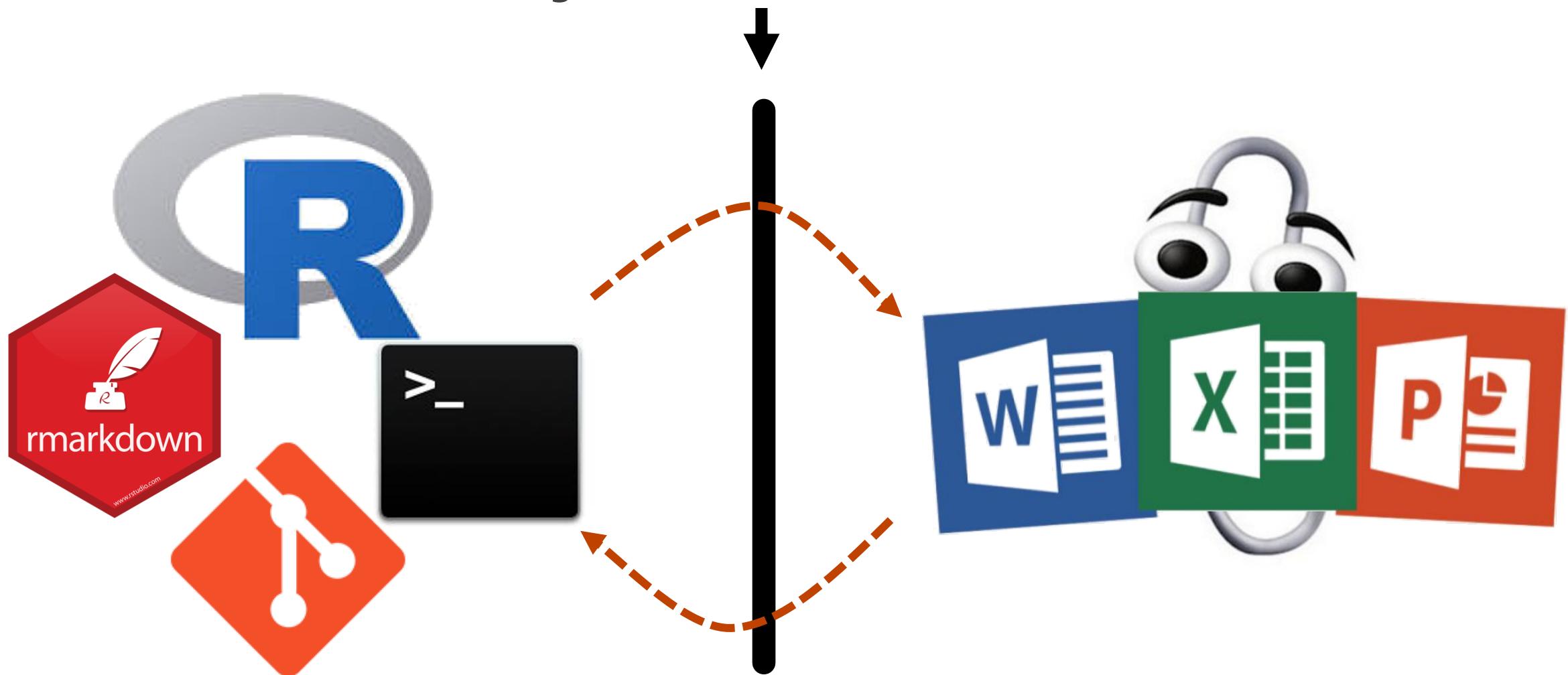
Decentralized

Manual

The Two Realms of Workflows



Valley of Heartbreak



The First Path: Warpath Across the Valley

“I know, I’ll teach everyone in my organization
git!”

[or R, or markdown, or...]



Ted Hart

@emhrt_

Writing a multi-author paper on github is a far bigger test of git skills than any coding I've done.

6:28 PM - 8 Jun 2016

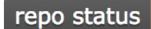
The Second Path: A Grand but Shaky Bridge

A web-based GUI editing interface

A continuous integration setup to compile

A *beautiful* R Markdown template

A Heroku app to host compiled docs behind a
Google OAuth

 repo status Suspended  build failing  build passing  codecov unknown

An R parser for ArchieML

This package is a wrapper for [archieml-js](#), a parser for the New York Times' [ArchieML](#) format. ArchieML is designed to include structured data in free-form documents.



“ Structured text,
for an unstructured world.”

Web and
programmatic outputs



Documents and further
iterated outputs



The Third Path: Attempting to Move the Earth

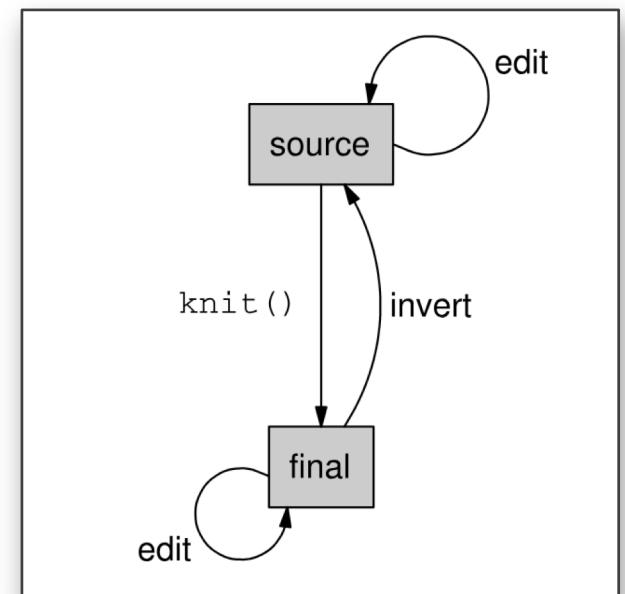
Invertible Reproducible Documents

Eric Lim jlim062@aucklanduni.ac.nz¹, Paul Murrell paul@stat.auckland.ac.nz¹, and Finlay Thompson finlay@dragonfly.co.nz².

¹[Department of Statistics, University of Auckland](#)

²[Dragonfly Science](#)

27 August 2014



 [ropensci / unconf16](#)

[Code](#) [Issues 30](#) [Pull requests 0](#) [Projects 0](#) [Wiki](#) [Insights](#) [Settings](#)

Google Docs/Drive

[! Open](#) noamross opened this issue on Mar 15 · 1 comment

Google Drive/Docs Rmarkdown

[Packages](#)

 **Louis** Mar '15

I just found [this repo](#) 87 which contains a script to convert Google Docs to markdown, and it's got me wondering whether Google Docs could be used somehow to go about Rmarkdown document editing differently, i.e. upstream of the existing RStudio/Rmarkdown framework.

 [ropensci / unconf18](#)

[Code](#) [Issues 64](#) [Pull requests 0](#) [Projects 0](#) [Wiki](#) [Insights](#) [Settings](#)

Incorporate word doc track changes back into R markdown

#76

[! Open](#) goldingn opened this issue on May 18 · 11 comments

 [ropensci / unconf18](#)

[Code](#) [Issues 64](#) [Pull requests 0](#) [Projects 0](#) [Wiki](#) [Insights](#) [Settings](#)

Workflow for publications using Rmarkdown with users that won't get passed Word/Google docs #42

[! Open](#) lauracion opened this issue on Apr 25 · 31 comments

This repository has been archived by the owner. It is now read-only.

 [ropenscilabs / markdown-remix](#)

 [Code](#)  [Issues 1](#)  [Pull requests 0](#)  [Projects 0](#)  [Wiki](#)  [Insights](#)  [Settings](#)

 [\[ABANDONED\]](#)
[Manage topics](#)

This repository has been archived by the owner. It is now read-only.

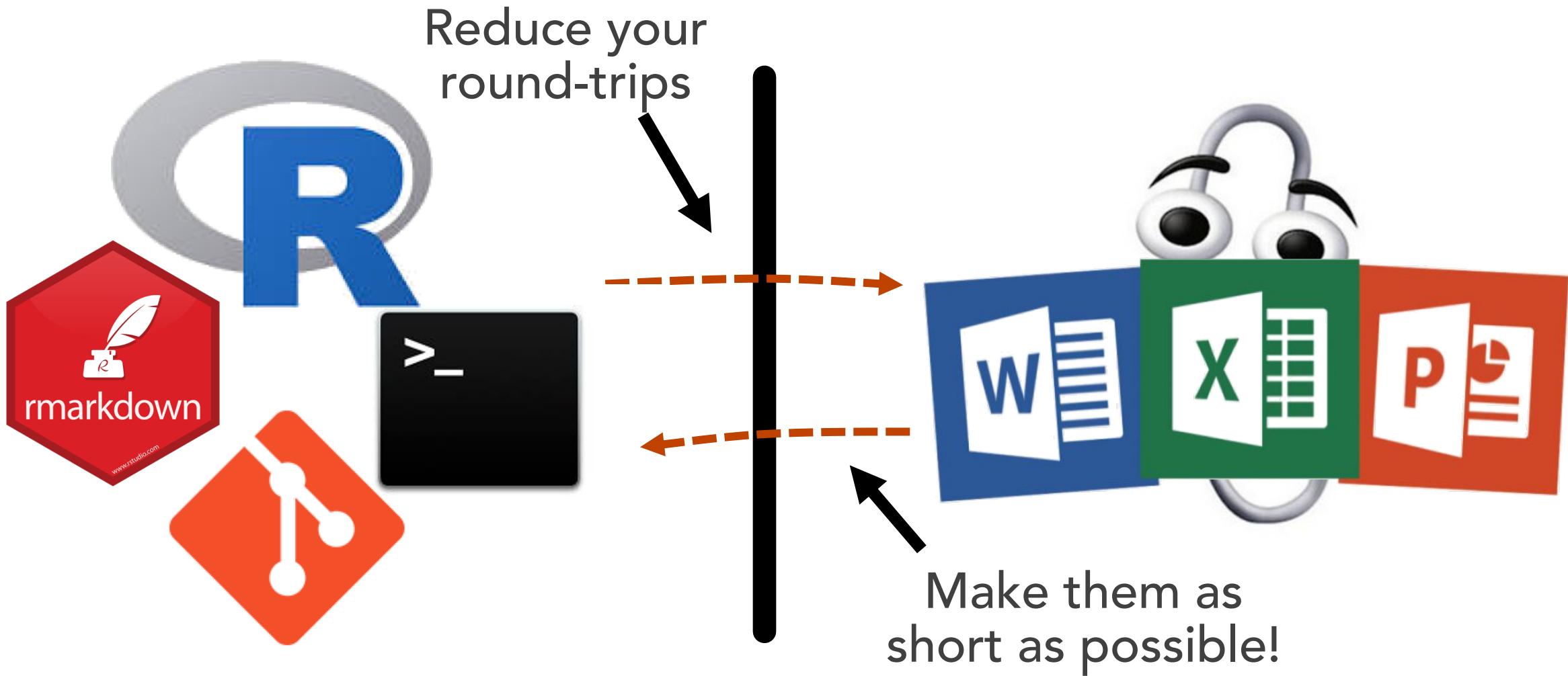
 [ropenscilabs / gdoc](#)

 [Code](#)  [Issues 0](#)  [Pull requests 0](#)  [Projects 0](#)  [Wiki](#)  [Insights](#)  [Settings](#)

 [\[ABANDONED\] Mixed Google/R workflows](#)
[unconf16](#) [r](#) [r-package](#) [rstats](#) [Manage topics](#)



Avoiding the Valley of Heartbreak



The Officeverse

officer

mschart

rvg

rmarkdown

flextable

pandoc

or, the gohelverse:
David Gohel (@DavidGohel)



Reduce your Round-trips:

**Tasks are trapped
on your side of the
valley**

*“Can you please
make the points
blue?”*

Reduce your Round-trips:

Tasks are trapped
on your side of the
valley

*“Can you please
make the points
blue?”*

Empower your
audience

Use `rvg` for plots,
`flextable` for
tables

Shorten the Journey

**Tasks repeat after
each iteration**

*“Please update
the formatting on
the new version”*

Shorten the Journey

Tasks repeat after
each iteration

*“Please update
the formatting on
the new version”*

Reach further
across the valley

Use `rmarkdown`
and `officer` for
more refined
reports

[ecohealthalliance / ehastyle](#)

Code Issues 1 Pull requests 0 Projects 0

R Markdown templates for EHA documents

[eha-modeling-analytics](#) [eha-infrastructure](#) [Manage topics](#)

71 commits 1 branch

EMERGING DISEASE INSIGHTS

Research from the PREDICT Modeling & Analytics team

Identifying the Next Zika: An Analysis of Understudied Flaviviruses

The ongoing Zika virus epidemic was declared a public health emergency of international concern by WHO in February 2016, due to its rapid spread and link to birth defects. Despite discovery in 1947, evidence of human infection, and close relation to high-impact human viruses like Dengue and Yellow Fever, the study of Zika remained limited until the Yap Island, Micronesia outbreak in 2007¹ (Figure 1). The pandemic spread of Zika emphasizes the need to better understand the distribution, host range, and epidemic potential of other understudied flaviviruses².

The PREDICT-2 Modeling & Analytics team compiled a database of all known natural host and vector species for the 53 ICTV-recognized viruses³ in the genus Flavivirus. In total, 1,768 flavivirus-host associations (including 317 unique mammal species and 743 bird species) and 309 flavivirus-vector associations (149 unique insect vector species) were identified. We then calculated the taxonomic breadth and geographic range of each flavivirus in order to inform both current and potential risks of flavivirus spillover and spread.

April 16, 2018

For details on methods or analysis contact:
PREDICTmodeling@
ecohealthalliance.org

Figure 1: Publication effort 1900-2016 of each flavivirus (n = 53). Publication counts per virus retrieved from Web of Science on 10/27/16. Virus color based on human impact: nonhuman viruses = green, subclinical viruses (detected only via serology) = blue, rare viruses (<20 pathogenic cases) = purple, epidemic viruses (>20 pathogenic cases) = red. Zika is colored black to emphasize the sharp rise in publications after its human impact changed to epidemic in 2007.

In order to control for the effect of research effort and understand the relative importance of host and vector breadth in the zoonotic potential of Flaviviruses, we developed a generalized linear model of likelihood a virus is zoonotic. We simplify host and vector breadth to the lowest taxonomic group that encompasses all known natural host or vector species. Both non-human host breadth ($p = 0.02$) and vector breadth ($p = 0.03$) were significant in determining if a virus is classified as zoonotic, so these variables were used to prioritize viruses for more research (Table 1). Using the IUCN mammalian range and BirdLife spatial files, we created a unique vertebrate host distribution layer for each flavivirus, and combine these to produce a global map of known flavivirus diversity in wild mammals and birds (Figure 2).

Of the 53 flaviviruses, 37 (70%) have been detected in humans, and 25 of these human-

davidgohel / worded

Watch ▾ 8 Star 35 Fork 1

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights

Branch: master worded / README.md Find file Copy path

 davidgohel requires pandoc V2 35534f7 on Mar 27

1 contributor

137 lines (91 sloc) | 5.09 KB Raw Blame History

build passing  build failing lifecycle experimental CRAN not published

Pimp your R markdown documents... to produce Word documents.

The package is to be used when you want to use R Markdown documents to produce Microsoft Word documents but also want options for landscape orientation, with narrow margins, with formatted text, when some paragraphs have to be centered.

So...

Each realm or workflow has its place

Know your audience and output

Don't (re)-invent grand frameworks

Smooth the paths between realms



Thank you!

@noamross
@EcoHealthNYC
@rOpenSci