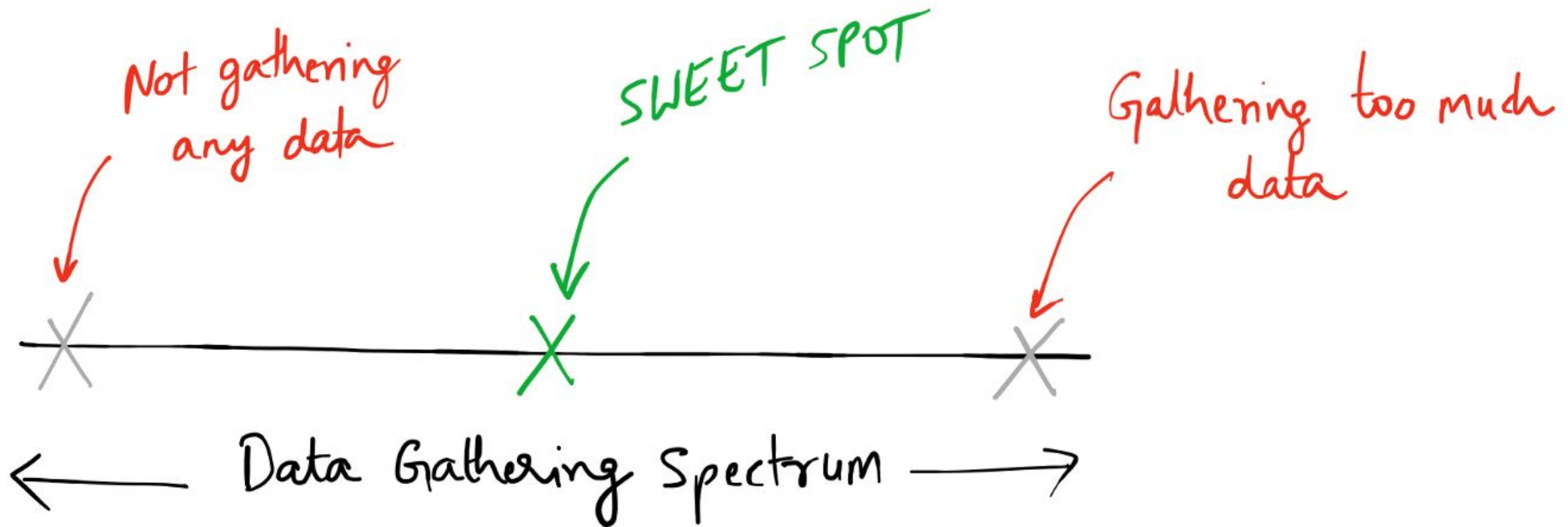


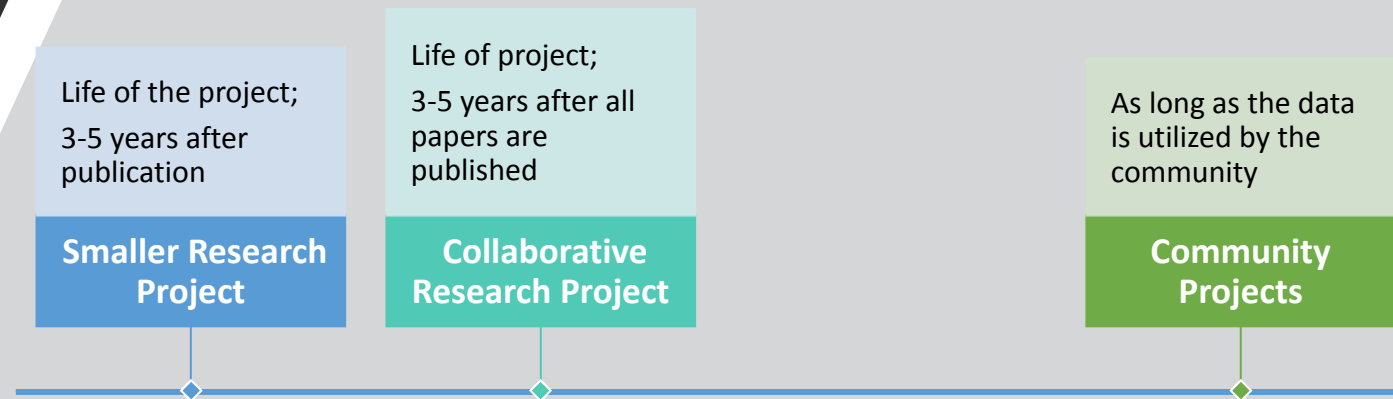
**Determining lifetime for  
simulation data, how to  
evaluate when data can be  
de-accessed from a  
repository?  
*A scientist perspective***

**Cindy Bruyère**  
([bruyerec@ucar.edu](mailto:bruyerec@ucar.edu))

# How long to keep data

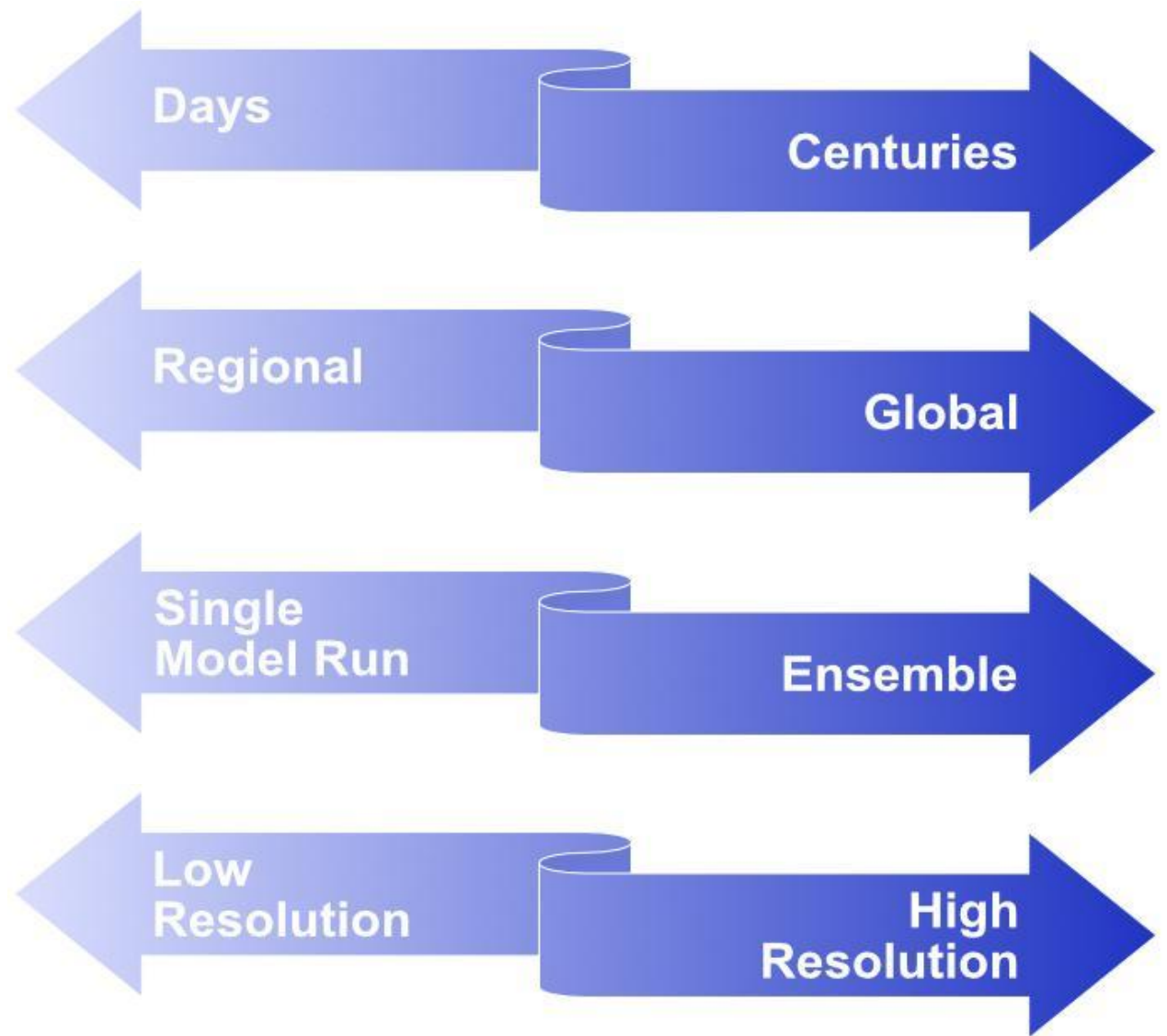
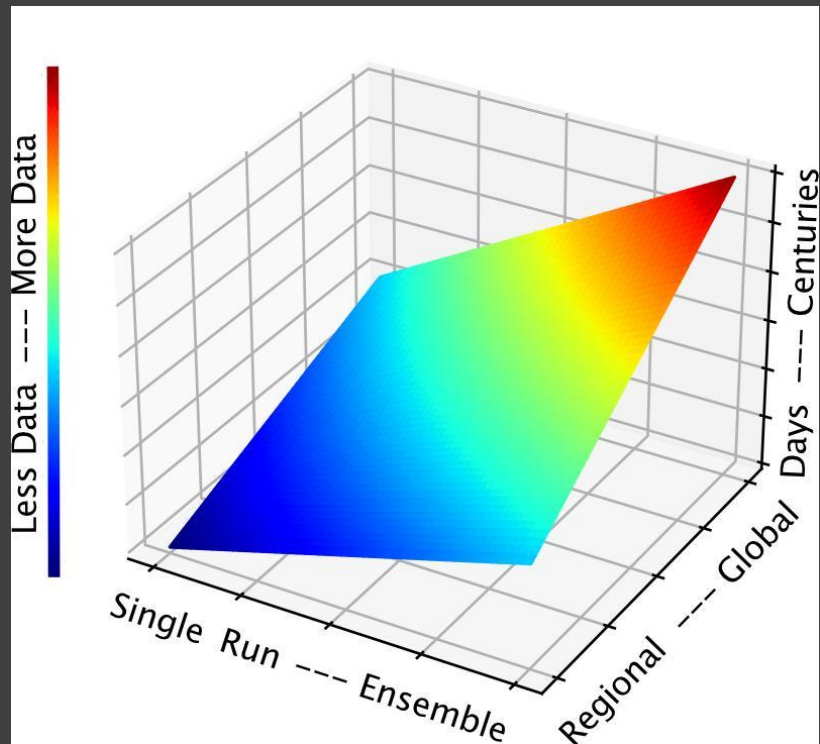


# Who is dependent on the data



Archiving data vs documenting the method and making code available

# Complexity / Time to Create Data



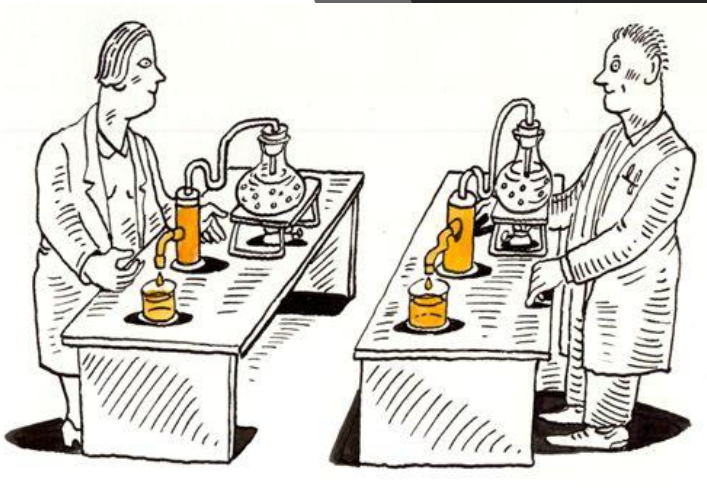
Less complex data is easier to store  
Which fields/periods to keep

# Reproducibility?

---

**Repeatability** – you can reliably repeat your own computation

---



**Replicability** - someone else can obtain the same results using the same setup as the author

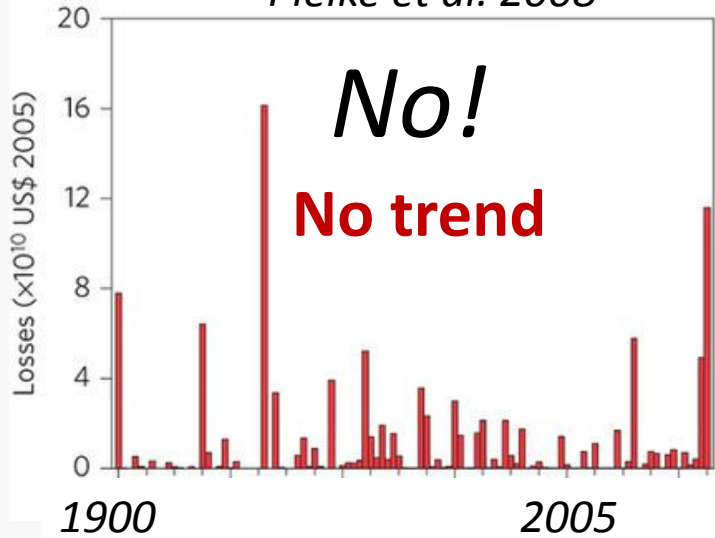
---

**Reproducibility** - someone else can obtain the same result using independently developed code or the same code on a different machine

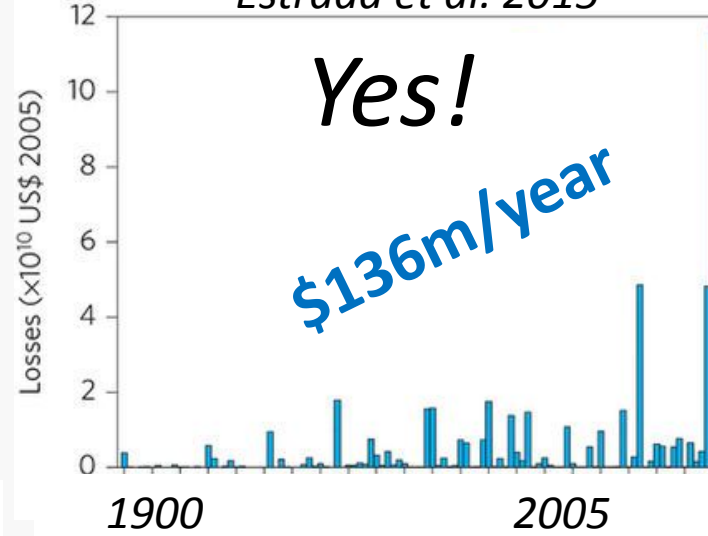
---

**Should data be 100% reproducible?**

*Pielke et al. 2008*



*Estrada et al. 2015*

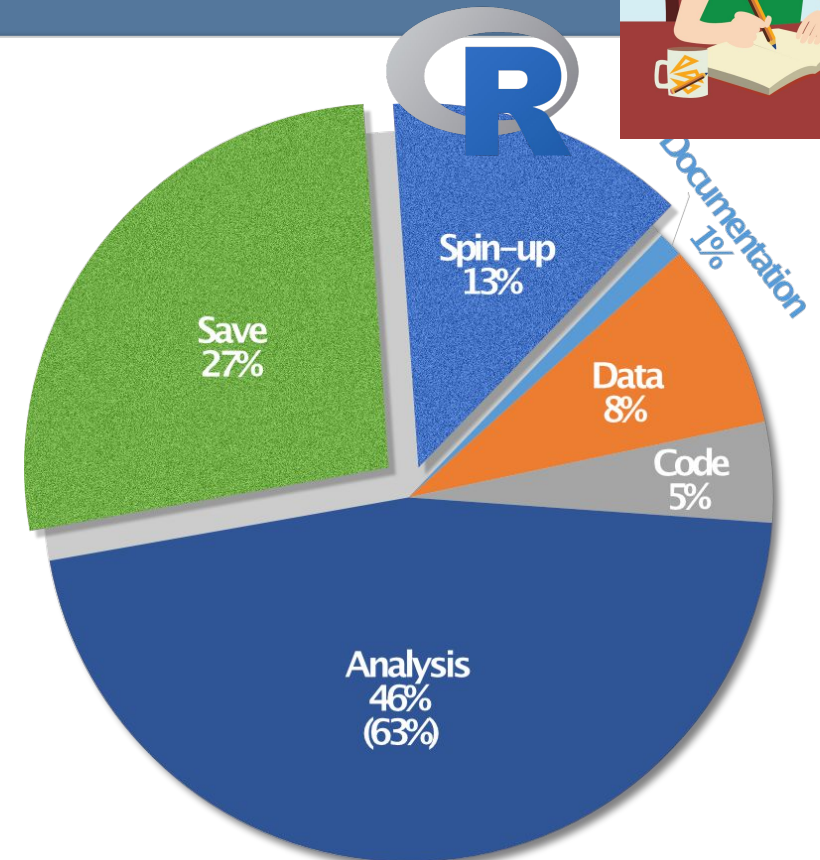
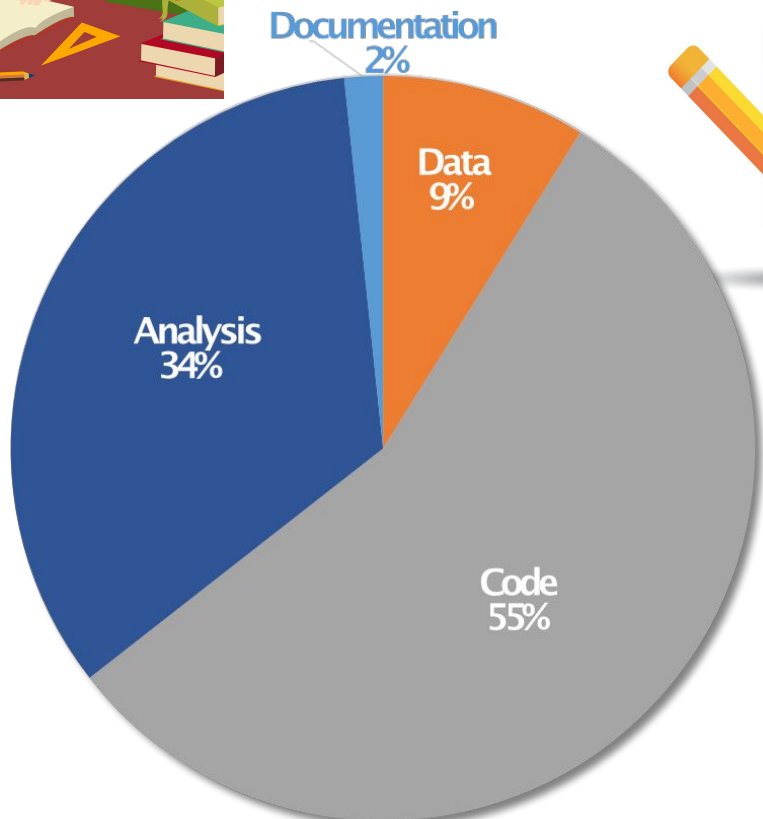


## Typical example from the literature

- Typical example from the literature



# Question: Environments Associated with Rapid Intensifying Hurricanes?



A stack of several sheets of white paper, slightly offset, creating a sense of depth and texture. The papers are set against a dark, textured background.

# Requirements / Relevance

- To complete a project / or someone else is using the data
  - Can the simulation be re-run?
- Required by a project or publication
  - Is documentation and code a good replacement for data
- Are new data/models available that are more relevant / up-to-date
- Are the data needed for comparisons
  - Are old data needed to help with comparisons/evaluations purposes
  - Model improvement
  - CMIP3/4/5/6?



# **Recommendation:** *A Scientist Perspective*

- I want to keep all my data, forever

