

Les midis numériques

**Data conservation - Perspectives,
issues and solutions.**

Miranda Bryant and Steve Vissault
s.vissault@yahoo.fr

January 14, 2014

All biologists collect data during their career but most of them are using **inappropriate files** to long term storage:

- Open Office or Microsoft spreadsheet
- text and CSV files

Some risks attributes at those practices: Overwriting file, lost the full dataset or some records.

Disadvantage of classic storage file (ex. Excel)

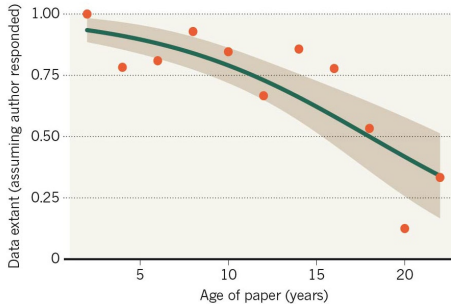
- ❶ Doesn't contain metadata informations (Authors, Year etc..)
- ❷ Data aren't normalize (Field Type)
- ❸

Why do something different ?

Data lost

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



Data for almost all studies published just two years ago were still accessible, the chance of them being so **fell by 17% per year**

We need to keep focus on those points as a part of our biologist culture:

- All datasets containing specific information given a time and a location are usefull.
- Most datasets are built on public funding and should be accessible publicly
- All datasets could be re-used, recycle or valorize (as the 3-R in waste management)