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
inapropriate 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.
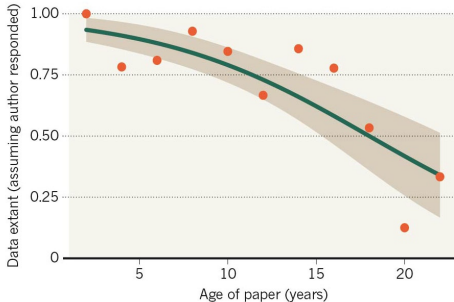
Disadvantage of classic storage file (ex. Excel)

❶ Flat file

❷ No dynamic query, only filters

❸ Large dataset could be messy

❹ Exportability

❺ Absence of fonctionnality on manage multiple users

❻ Missing options to create metadata

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 (Vines et al., 2013)

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.
- 80% of datasets are built on public funding (Graham, 2013) and should be accessible publicly
- All datasets could be re-used, recycle or valorize (as the 3-R in waste management: Reduce, Reuse, Recycle)

D. Graham. Academic Publishing - Survey of funders supports the benign Open Access outcome priced into shares. Technical report, HSBC - Global Research, 2013.

T. Poisot, R. Mounce, and D. Gravel. Moving toward a sustainable ecological science: don't let data go to waste! *Ideas in Ecology and Evolution*, 6(2):11–19, 2013. ISSN 19183178. doi: 10.4033/iee.2013.6b.14.f. URL `http://library.queensu.ca/ojs/index.php/IEE/article/view/4632`.

T. Vines, A. Albert, R. Andrew, and F. Débarre. The availability of research data declines rapidly with article age. *Current Biology*, 2013. URL `http://www.sciencedirect.com/science/article/pii/S096098221301400C`