



Open Data: A Modest Proposal

Steven M. Bachrach
Trinity University
sbachrach@trinity.edu



Sharing Data

- ✚ Allows for better peer review
- ✚ Allows for reproduction of experiments and verification of results
- ✚ Allows for comparisons of similar experiments
- ✚ Aids in collaborations



Internet Enablers

- ✦ All data fundamentally digital
 - Images, plots, spreadsheets, spectra, etc.
- ✦ Files readily stored and retransmitted
- ✦ Much easier to combine and maintain archive than with paper
- ✦ Distribution relatively inexpensive

Impediments

- ✘ Non-standard file formats
- ✘ Multiple file formats
- ✘ No standard deposition location(s)
 - *Cambridge Crystallographic Data Centre (CCDC)*
 - *Protein Databank*
- ✘ No incentives for authors
 - *NSF data requirement*
 - *Publishing x-ray structures*
- ✘ Lack of authoring tools



Proposal

Publishers can

- ✚ Partner with societies and librarians to create data archives
- ✚ Encourage editors to demand data submission
- ✚ Partner with software and instrument companies to create authoring/collaboration tools
- ✚ Partner with librarians and scientist to create metadata to enable search and retrieval

Plea

- ✦ Data is outside copyright
 - ✦ Establish all submitted data as “CC-BY”
 - **Free to Share** — to copy, distribute and transmit
 - **Free to Remix** — to adapt the work to make or even make commercial use of the work
 - **Attribution** — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work)
- (from Creative Commons: <http://creativecommons.org/licenses/by/3.0/us/>)