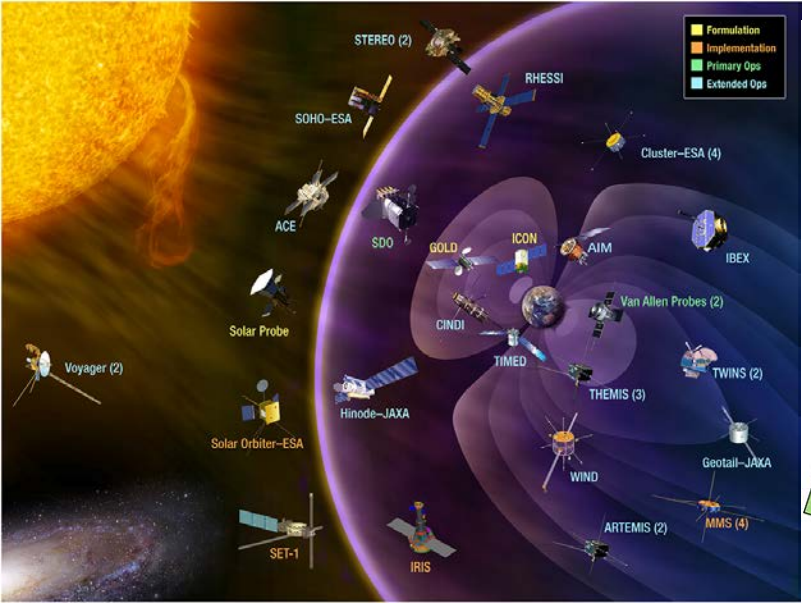
Reproducibility and FAIR Data in the Earth and Space Sciences

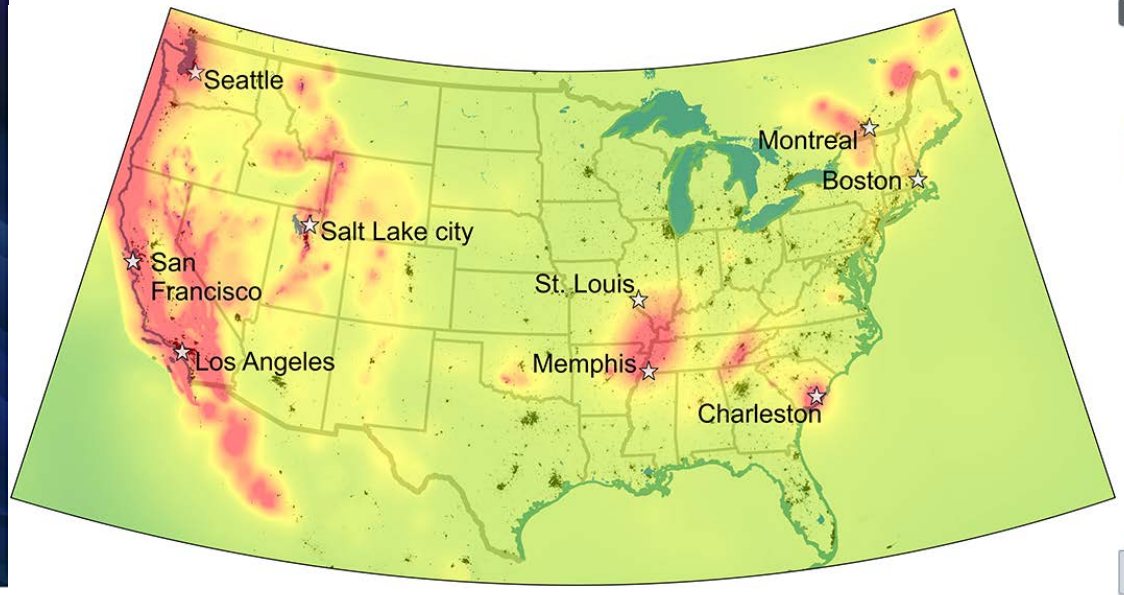
December 2017

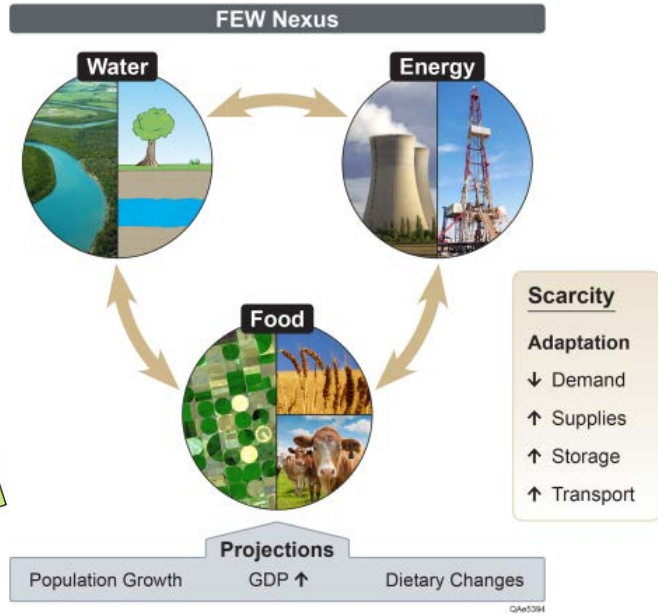
Brooks Hanson

Sr. VP, Publications, American Geophysical Union

bhanson@agu.org







Earth and Space Science is Essential for Society

Special AGU Journal Issue, April 2017

27 Commentaries across AGU journals illuminate the deep and growing benefits of research in the Earth and space sciences for humanity.
Driven by international shared data

Open Access; see link at http://publications.agu.org

AGU data position statement (began in 1997; updated in 2016)

Earth and space sciences data are a world heritage.

- They should be preserved long-term for future use.
- They should be made openly available to the scientific community and the public as soon as possible.
- They should be accessible in usable formats with sufficient machine-readable documentation to allow informed re-use.
- These responsibilities are an integral part of scientific research shared by individual scientists, data stewards, research institutions, and funding organizations.

Enabling FAIR data in the scholarly literature

- Embrace that published papers are only part of the research record.
- They must contain useful and reliable 2-way links and identifiers to other secure resources for integrity and discoverability:
  - Context (metadata) around these links is critical
  - Data, software, repositories, samples (IGSN)
  - Funding information
  - Author information (ORCID, CREDIT, institutions)
  - Reference information (semantic context is coming)
- We need efficient ways to help authors, publishers and repositories preserve these links: Standard, expected, sensible, easy.
- Data needs quality curation, metadata, provenance, linking.

Recent Alignment by Publishers, Repositories, and Funders Around Best Practices

- **TOP** (transparency and openness promotion guidelines)--2900 journals
- **http://COPDESS.org** (Coalition on Publishing Data in the Earth and Space Sciences)—Statement of Commitment--most publishers and repositories in the Earth and space sciences
- **Joint Declaration of Data Citation Principles**–114 organizations
- **Software Citation Principles**: https://doi.org/10.7717/peerj-cs.86
- **Reproducibility conferences** and outcomes (AAAS and other orgs)
  - Best practices around: Clinical trials, Lab studies, Field data, Software, Industry-academic research
- **Authorship** (https://doi.org/10.1101/140228 submitted to PNAS)
- Quality/certification standards for repositories

Challenge is practicing what you preach

Current status:

- Publishers increasingly requiring data (and code) availability
  - Supplements still being heavily used (no metadata, pdf often)
  - Growing use of repositories, domain and general (Figshare, Dryad, institutions).
  - Few standards on metadata or linking (limiting discoverability; interoperability).
  - “Available from authors (yeah, right)” still common
  - “unpublished” references still commonly allowed
- Publishers implementing other identifiers (ORCID, IGSN, etc.)
- Author standards and expectations and declarations standardized
- Best practices for FAIR data are available
- Great examples in some disciplines/repositories of successful implementations and solutions but not widely adopted.
- Leveraging and scaling these solutions

Software: Current Status

- Leading journals have software transparency standards
- Community best practices emerging
- But…little uniformity in those best practices and limited awareness among authors, editors

- Key issues:
  - Licenses—Use MIT or other software license, not CC-BY (which require attribution and documentation of any changes)
  - Github has limited metadata (can use zenodo as a landing page).
  - Different IP issues than for data (treated differently by universities)

Grant from Laura and John Arnold Foundation (LJAF) to AGU

- **Align/develop best practices and standards across the Earth and space sciences to enable FAIR data**

- Develop common solution(s) for researchers, publishers, editorial systems, and data repositories

Community-Driven Project – Partnership Includes:

- **Science Data Communities**
  - AGU
  - Earth Science Information Partners
  - Research Data Alliance (RDA)
  - COPDESS
  - Earthcube/CDF
  - DataCite
- **Publishers**
  - AGU
  - *PNAS*
  - *Nature*
  - *Science*

- **Repositories and COPDESS Signatories**
  - National Computational Infrastructure (NCI)
  - AuScope
  - Australian National Data Service
- **Infrastructure**
  - Center for Open Science

And Growing!!

The Goal:

Publishers will adopt common standards in their editorial systems around workflows, datasets, metadata, acceptable repositories, and data citation. Will connect w/ repositories via api's for efficient metadata exchange.

Repositories will adopt standards and best practices around, persistent identifiers, landing pages (for exposing metadata), access, embargoes (for peer-review), data citations, and licenses.

Researchers will know what to expect across all journals and be able to prepare data early in workflow.

Progress so far, and Directions:

- 6+ working groups (we're calling them temporary action groups) to address needs among publishers and repositories.
- Major ESS publishers have agreed to eliminate data supplements—all data in repositories with DOI's, curation.
- Repositories working on solutions to improve data curation and help researchers.
- Funders also engaged; strengthen and provide more directed DMP's (NRC can help here especially).
- Expanded declaration or principles around FAIR data supported and in progress
- Implementation begins in summer 2018

Larger Effort Needed

- Support and publicize these community efforts around best practices
- **Publishers** need to follow current best practices
  - Get references out of supplements in online versions (all references in main text); open up references at Crossref (I4OC)
  - Help authors (include data best practices and expectations into workshops, instructions…)
  - Ensure integrity (no unpublished references; data availability statements).
- **Societies** should recognize data stewardship in awards and recognition (fellowship, academy membership).
- Funders need to standardize and strengthen DMP's guidelines and follow through on these (be more directive)
  - Update guidelines and FAQs to follow best practices
  - Support leading publishers
  - Support leading repositories
- Implement identifiers fully (samples, affiliations, repositories….)

Efforts and Culture aligned for Reproducibility

- FAIR data effort across the Earth and space sciences—publishers and repositories are implementing
- Recent surveys indicate much increased acceptance of need for data sharing and curation in the Earth and space sciences
- Need to reward groups/people that are following best practices
- Help, direction, and common outreach needed from funders
  - NRC can greatly help here in providing directions to funders/societies/researchers
- Society alignment and support (AGU willing to lead and share what we have done with other societies)