Data management across academic disciplines
How have other disciplines and communities addressed data citation and attribution?

Linguists are not alone in understanding the necessity, challenges and benefits of the citation and attribution of linguistic data. This poster highlights some community based responses to the issue and specific tools designed to accomplish the goals at hand.
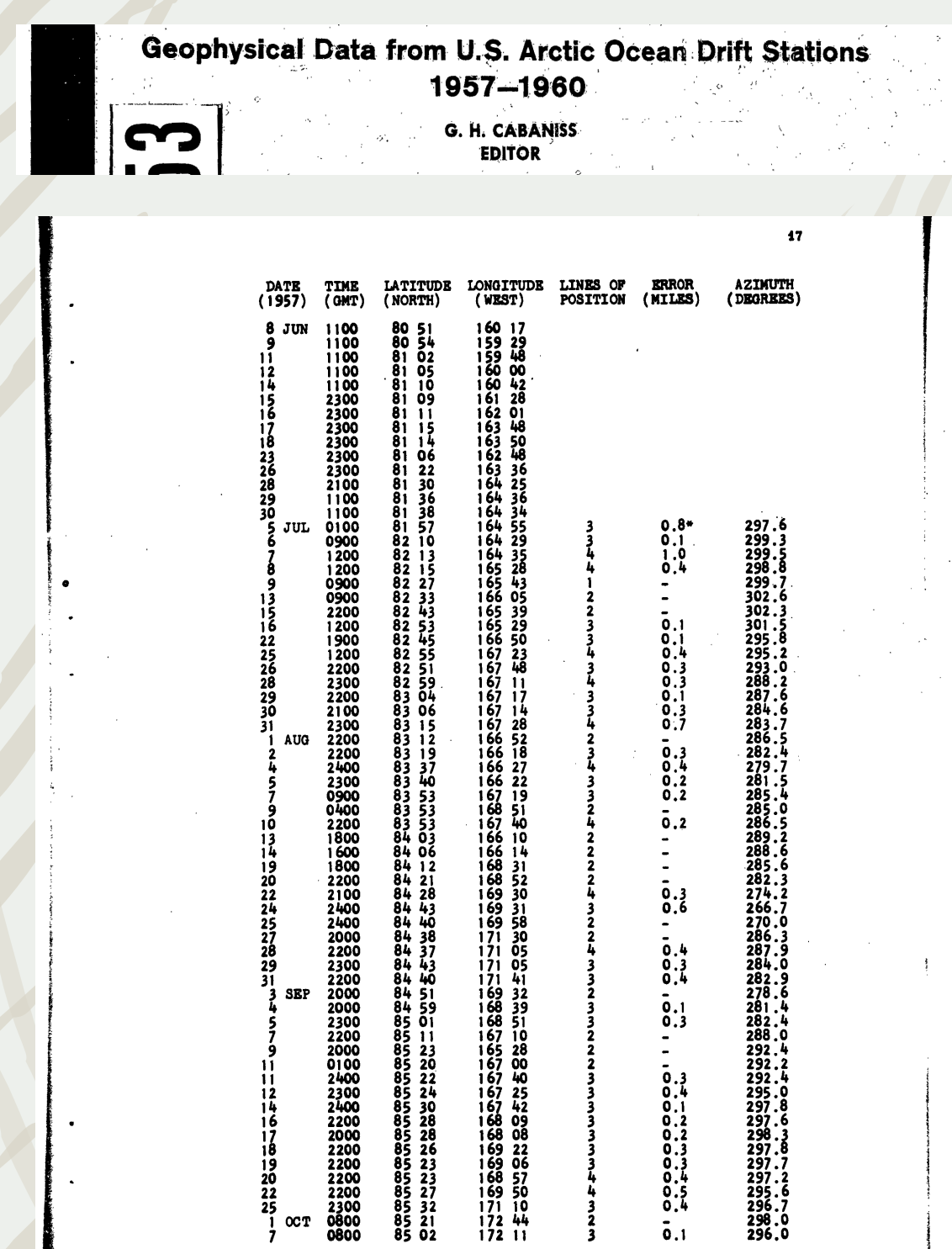
THE PROBLEM

In the past, data was published like research findings.

Due to technology, data became more structurally complex and quantities grew rapidly.

Traditional publishing of data waned.

Technology also facilitates grey publishing of data.

Excerpt from traditionally published geophysical data [1]

COMMUNITY BASED RESPONSES

E-IRG WHITE PAPERS [2]

Open access and free movement of knowledge are the key policy Issues.

Barriers towards goals are structural and organizational, rather than technical.

Appropriate incentives need to be cultivated to develop a sharing mentality among the users.

Involve stakeholders of the data infrastructure including resource providers, existing infrastructures and initiatives, and user communities (especially communities created by national borders)

e-IRG e-Infrastructure Reflection Group

facilitate integration of European e-Infrastructures

support strategic policymaking and sustainable services

Author
Kavon Hooshiar
kavon@hawaii.edu
University of Hawaiʻi at Mānoa

COMMUNITY BASED RESPONSES (CONT.)

FORCE11 DATA CITATION PRINCIPLES [3]

1. Importance
Data are citable products of research.
Data is as important as publications.

2. Credit and Attribution
Citations facilitate giving scholarly credit, and attribution to all contributors.

3. Evidence
Scholarly claims relying on data should have that data cited.

4. Unique Identification
A data citation should have a persistent identifier that is:
machine actionable, globally unique, and widely used by a community.

5. Access
Data citations should facilitate access to the data themselves, including associated metadata, documentation, code, and other materials, so they are readable by both humans and machines.

6. Persistence
Unique identifiers and metadata should persist, even beyond the lifespan of the data they describe.

7. Specificity and Verifiability
Citations should facilitate identification, access, and verification of the data supporting a claim. Citations/metadata should include information about provenance and fixity to ensure data retrieved is the same as was originally cited.

8. Interoperability and Flexibility
Data citation methods should be both:
sufficiently flexible to accommodate different communities, and sufficiently interoperable to facilitate reference across communities

FORCE11 The Future of Research Communications and e-Scholarship

improve research practices through innovations in the ways knowledge is created and shared across research disciplines, communities, sectors and timeframes

community of scholars, librarians, archivists, publishers and funders

DIGITAL CURATION CENTRE LIFECYCLE MODEL [4]

DCC because good research needs good data

focus on building capability and skills for research data management

provide expert advice and practical help to research organizations

TOOLS IN USE TO ACHIEVE THE GOALS

doi Digital object identifier: identifier of an entity on digital networks; system for persistent and actionable identification and interoperable exchange.

Provides an identifier for individuals to use with their name as they engage in research, scholarship, and innovation activities.

ORCID Connecting Research and Researchers

OpenAIRE Through the OpenAIRE portal, makes as much European funded research output as possible available to all.

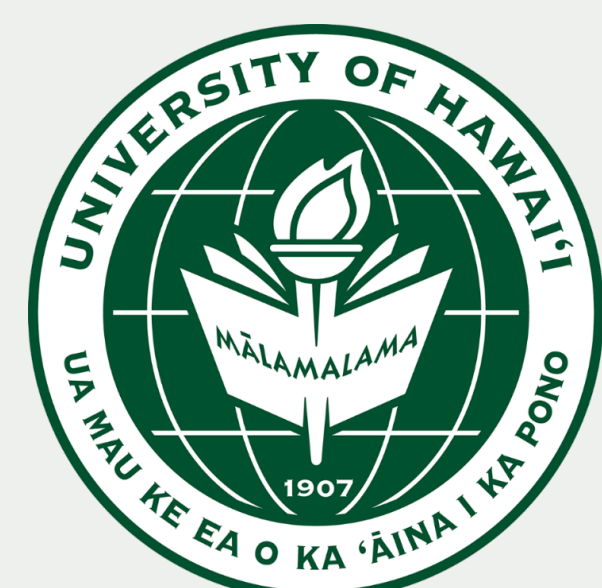Provides DOIs for research data, to help locate, identify, and cite research data with confidence.

DataCite FIND, ACCESS, AND REUSE DATA

THE DATA CITATION INDEX ON WEB OF SCIENCE Single point of access to research data from repositories across disciplines and around the world. Data is linked to the scholarly research.

CONCLUSIONS

-Fields across academia are working toward data citation and attribution.
-The guiding principles of organizations dedicated to data citation and attribution match our own.
-Tools are already under development to facilitate specific communities while maintaining interoperability.
-Our efforts should be informed by and mesh with the work already accomplished by the greater scientific community.

REFERENCES

[1] Cabaniss, G. H. (1962). *Geophysical Data from US Arctic Ocean Drift Stations 1957-1960*. AFCRL-62-683, pp. 1-43, Air Force Cambridge Res,. Lab., Bedford, Mass.
[2] Vandenbroucke, Rosette (ed.) (2011). *e-IRG White Paper 2011*. e-Infrastructure reflection group. http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf
[3] Martone M. (ed.) (2014). *Joint Declaration of Data Citation Principles*. Data Citation Synthesis Group. San Diego, CA: FORCE11.
[4] Higgins, S. (2008). *The DCC Curation Lifecycle Model*. The International Journal of Digital Curation, 3(1), 134–140.