

Data Curation in Scientific Teams: An Exploratory Study of Condensed Matter Physics at a National Science Lab

Introduction

Research processes are increasingly data driven, and there is a growing need to share, reuse, and aggregate data before, during, and after experiments are conducted. Access to original data and the record of its provenance is also necessary to replicate and validate the findings of scientific experiments.

This pilot study explores the data curation and sharing practices of scientific teams within the **condensed matter physics** (CMP) community at the National High Magnetic Field Laboratory (NHMFL) in Tallahassee, Florida.



National High Magnetic Field Laboratory in Tallahassee, Florida

Methodology

In-depth semi-structured interviews with 5 key scientists

- Positions: 3 experimentalists, 2 theorists
- Institutional roles: 2 full-time staff scientists, 2 academic faculty, 1 external scientist

Interview Questions:

- lifecycle of scientific collaborations
- scientific work practices
- perceptions of data ownership
- rules, community norms, policies regarding data practices
- criteria for assessing quality
- data sharing

Interviews recorded, transcribed, coded with NVivo 9

College of Communication and Information Florida State University, PO Box 3062100 Tallahassee, FL 32306-2100

Charles C. Hinnant, Besiki Stvilia, Shuheng Wu, Adam Worrall, Gary Burnett, Kathleen Burnett, Michelle M. Kazmer, Paul F. Marty



Condensed matter physicists examine a magnet probe

Findings

Data from CMP experiments is generated by sensors that measure the effects of various stimuli (i.e. magnet field) on the sample being studied.

- Value is placed on developing better sensors and new analytical techniques to reduce "noise" in data.
- Development of such innovations is tied to the reputations of scientists.

Contextual information is crucial to understand the data.

Scientists maintain their own "notebook" that explains the experimental protocol and is not shared with other scientific teams or with publication outlets.

Ownership of data rests firmly with the PI.

• This is the norm within the CMP community and is also reinforced by strict policies at the lab.

The CMP community is highly diverse; work orientations may impact the ability to develop more formalized curation policies and rules at the NHMFL.

"[CMP may be] dirtier than many other fields...you do not have a standard instrumentation that gives sort of standard output finals or something, you know. So I think that people do things very different ways..."

Conclusions

Data practices are highly specific to individual CMP teams and thus may vary widely across teams.

- repurposing.

• For some CMP scientists the perceived cost of data management and curation may exceed their perceived value of sharing, reusing, and repurposing the data. Future research will conduct a comprehensive survey of data curation practices among the NHMFL users and staff.



900 megahertz, 21.1-tesla ultra-wide bore NMR magnet

Acknowledgements

This research was supported in part by the National Science Foundation under Grant OCI-0942855. This poster reflects the findings and conclusions of the authors, and does not necessarily reflect the views of the National Science Foundation.



• The variety in data management and curation practices may weaken the potential for data sharing, reuse, and

> Contact: Charles C. Hinnant Email: chinnant@fsu.edu Project website: *http://voss.cci.fsu.edu*