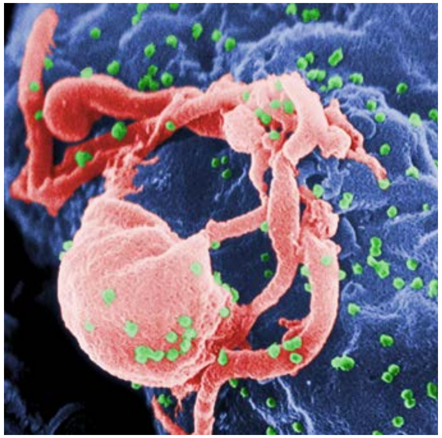
UNDERSTANDING SCIENTIFIC COLLABORATION

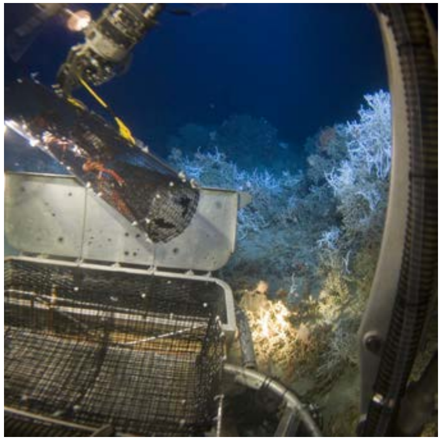
AAAI/CCC Symposium on Accelerating Science: A Grand Challenge for AI

Charlotte P. Lee
Associate Professor
Dept. of Human Centered Design & Engineering (HCDE)
University of Washington
Seattle, WA

Research Program: Empirically grounded collaboration theory

How can we better understand and support how scientists collaborate?





Approach: how scientists do research *and* how they organize in order to do their work

What is Computer Supported Cooperative Work (CSCW)?

- Subfield of Computer Science field of Human-Computer Interaction (HCI)

- Notion of “Human” in HCI keeps expanding
  - Ergonomics/Human Factors
  - Psychology
  - *Social Science – (Organization Science), Ethnomethodology, Anthropology, Sociology*
  - Humanities/Arts

- CSCW concerned with informing user-centered design of collaborative systems how people cooperate in social situations
  - how people use existing systems,
  - How prototype affects human activity

Social Science in HCI

Strongest influences

- Cultural Anthropology
- Ethnomethodology

Ethnography the first Social Science method widely adopted in HCI and CSCW (not the only one!)

- Rich descriptions
- Emphasis on the particular cases
- Implications for design
- Skepticism about theory, models

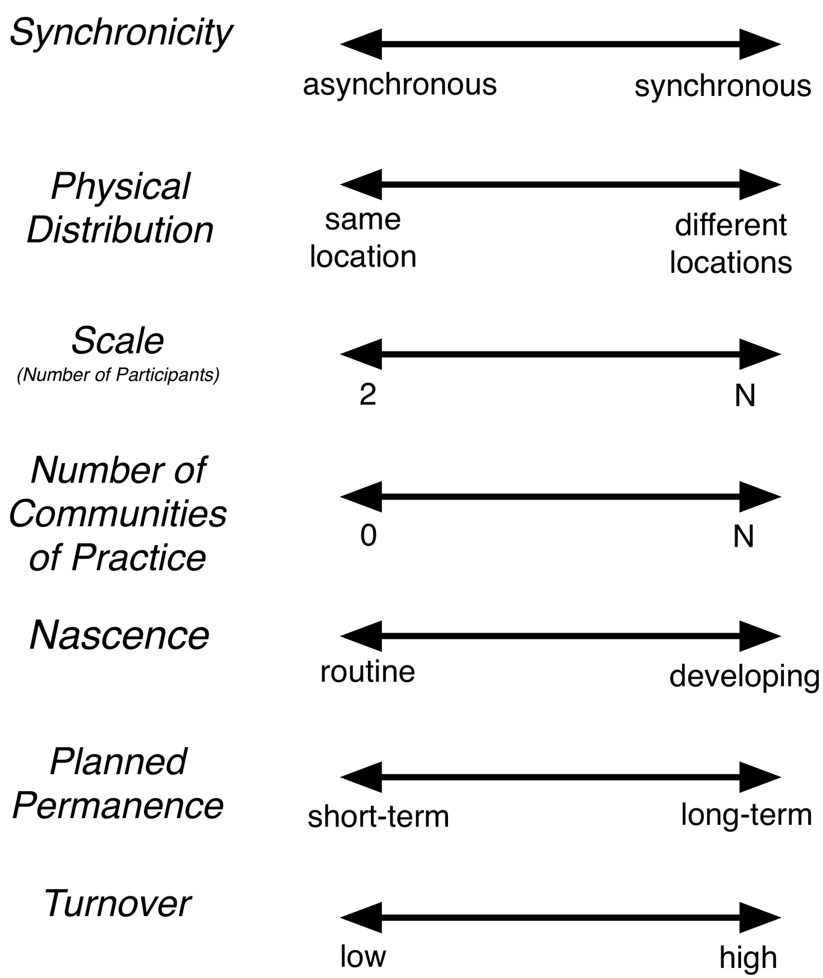
Releasing the stranglehold and breaking down barriers in CSCW

- Grounded, data-driven theories of collaboration
- Multi-sited ethnography of scientific research groups
- Studying and theorizing the “middle range” – meso level theory.

- **Early stages of any science focus on descriptive laws that summarize empirical regularities**

- Model of Coordinated Action (MoCA)
- Theoretical framing to generate new research questions for both qualitative and quantitative researchers

Model of Coordinated Action (MoCA)

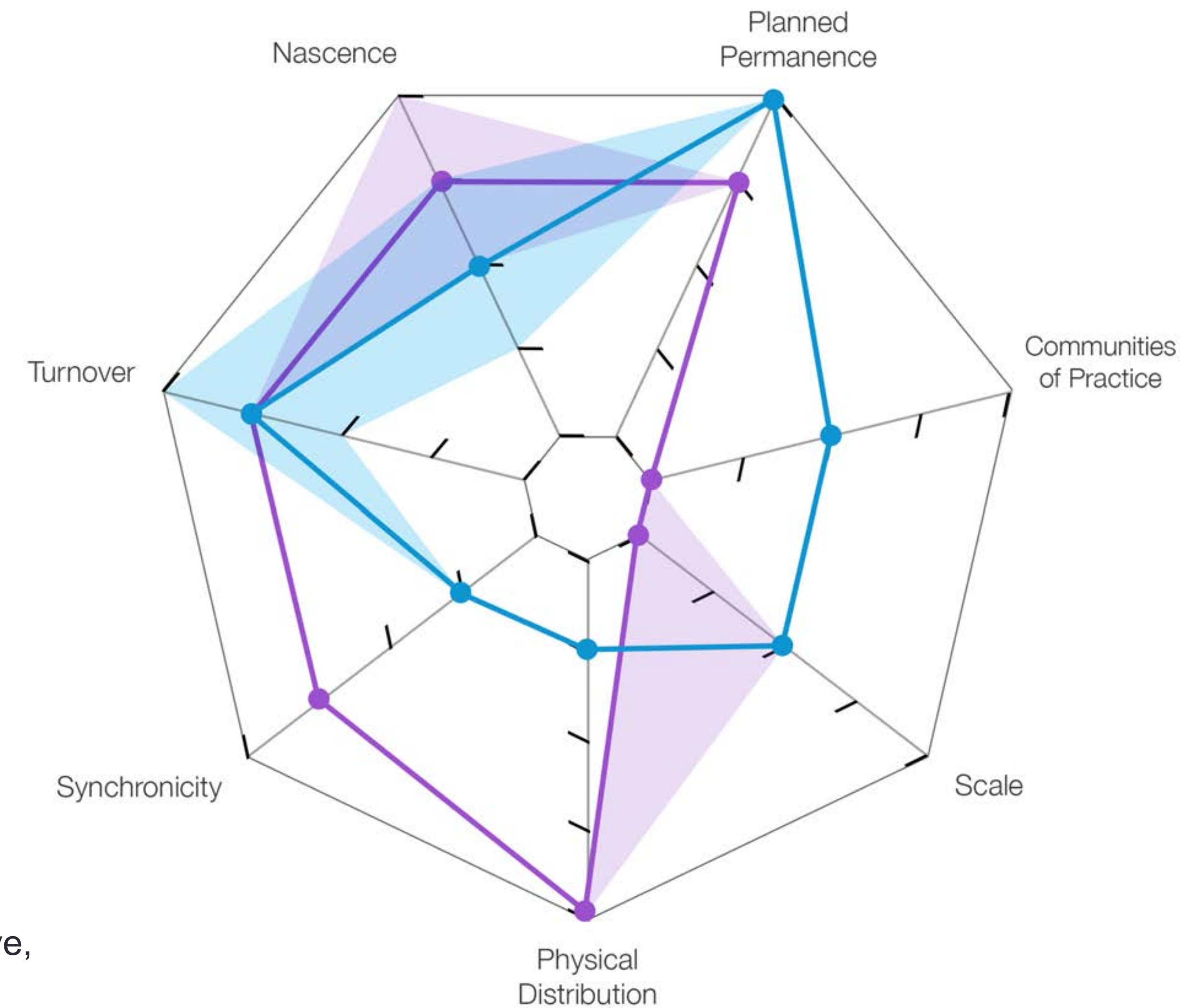


The Model of Coordinated Action (MoCA) and its seven dimensions with the end points of each continuum.

Lee & Paine 2015

Implications for CSCW Research

- Change over time
- Differences ***within***
    - particular domains
    - particular types of activity
- Differences ***across***
    - particular domains
    - particular types of activity

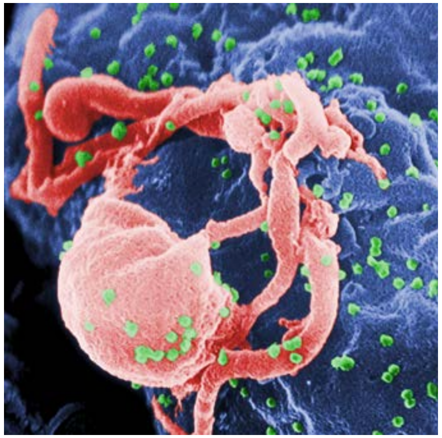


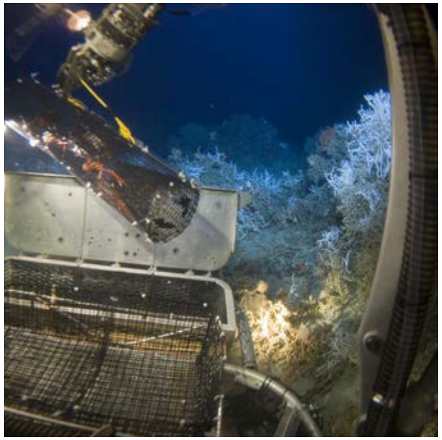
Early prototype of a comparative, overlaid visualization of the coordinated action examples discussed in article. Not actual data.

Research Program

How can we better understand and support how scientists collaborate?

- Scientific Cyberinfrastructures
- All types of collaboration for innovation
- Sociotechnical design for emergent organizations





CSCW Research on Sociotechnical Aspects of Scientific Collaboration

- Collaboratories
- *Cyberinfrastructure Development and Use*
- Data Sharing / Data Science
- Software Development
- Infrastructure Studies

Accelerating Science: A Computing Research Agenda

White paper by Honavar, Hill and Yelick (2016):

- Accelerating science requires rich model of entire scientific process
- Science increasingly a collaborative — need sharable structures and processes that facilitate collaborative science
- Need to do more supporting sharing: scientific workflows, mechanisms for decomposing tasks, assigning tasks, integrating results

Looking at All of Scientific "Process"

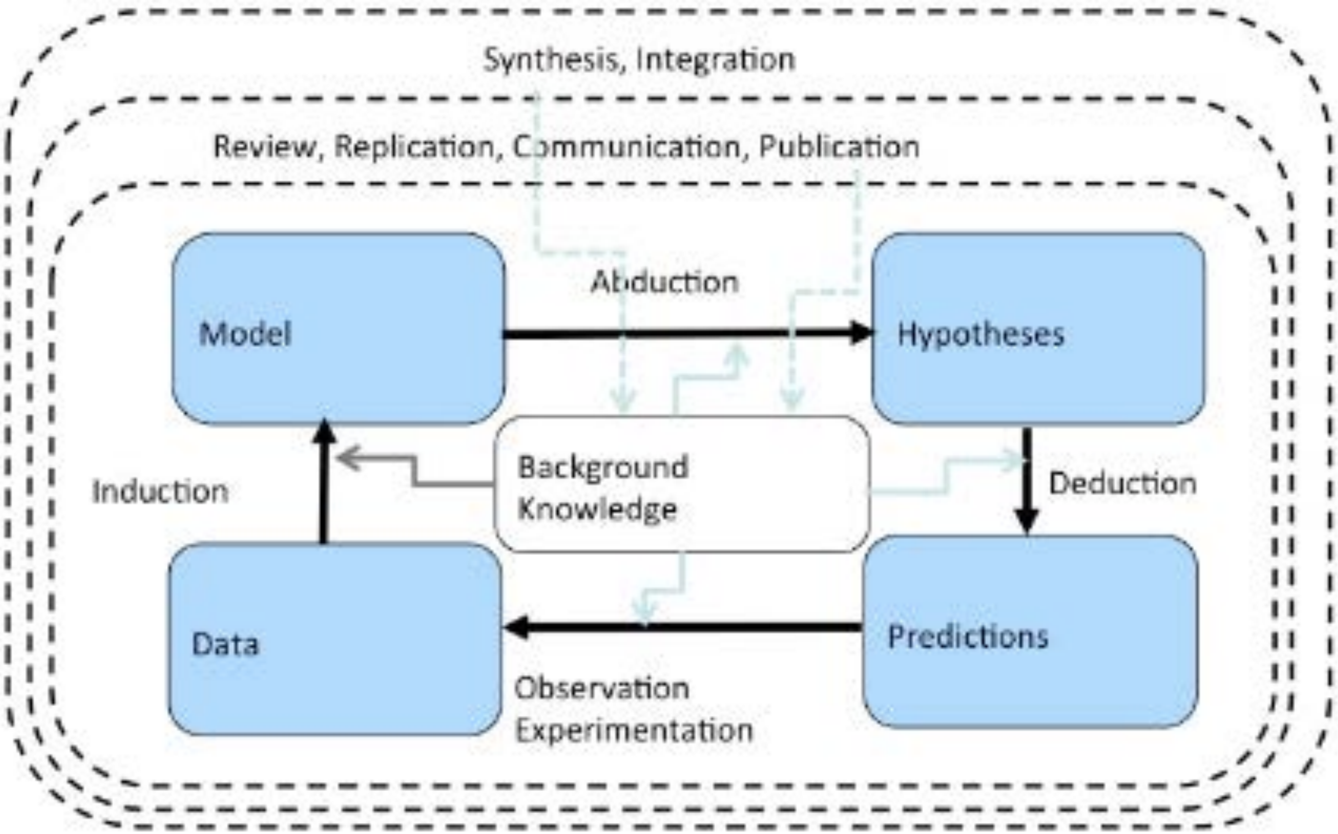


Figure 1: A Cartoon of the Scientific Process

CROSS-DISCIPLINARY COMPARATIVE STUDY OF SCIENTIFIC COLLABORATION

Research Sites

- University of Washington, Seattle WA
- Four research groups in different sciences doing data-intensive research

Image from Wikimedia Commons

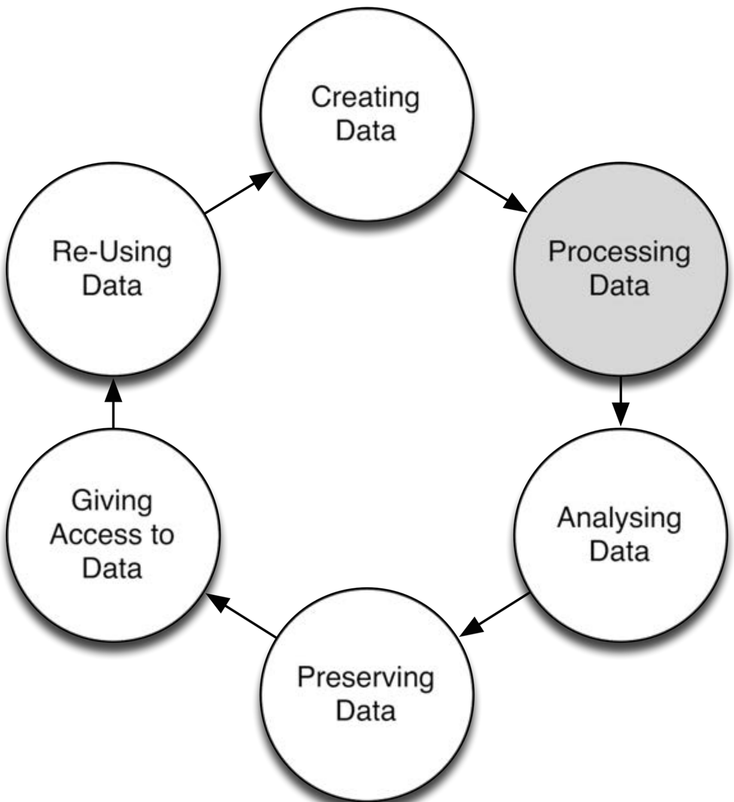
Research Sites

PI	Phenomena of Study	Research Group
Hank *Atmospheric Science professor*	Interaction of "dynamics, radiation, and cloud processes"	4 Doctoral Students
Waldo *Marine Geophysicist professor*	Submarine volcanoes and mid-ocean ridge hydrothermal systems	3 Doctoral Students
Martin *Microbiologist professor*	Human Immunodeficiency Virus (HIV)	1 Doctoral Student 3 Research Scientists
Magnus *Empirical Cosmologist professor*	Epoch of Reionization through the development and application of novel radio telescopes	3 Post-Doctoral Researchers 3 PhD Students 2 Undergraduate Students

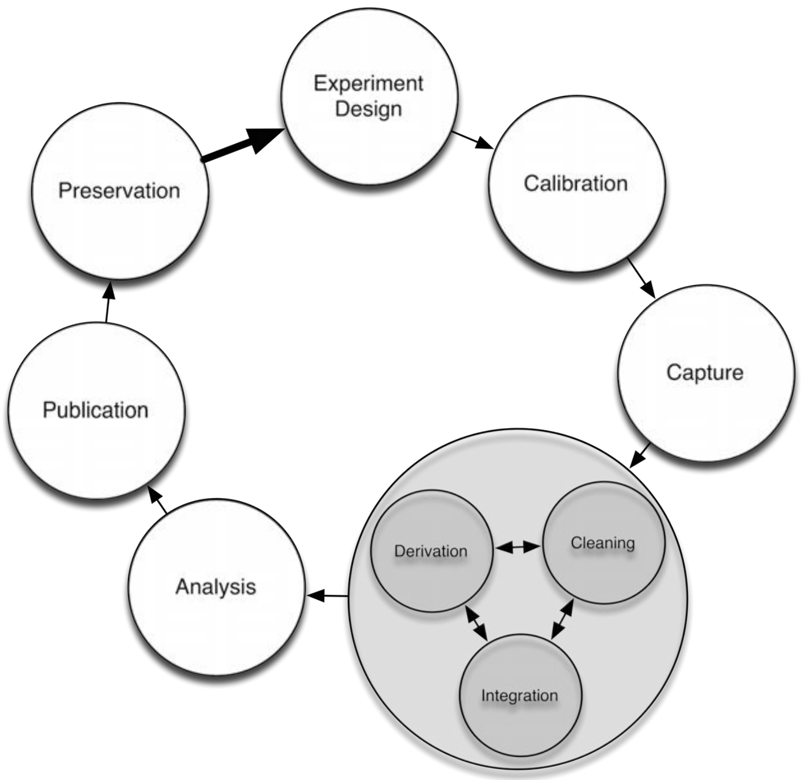
Research Methods

- Qualitative study
  - Observations of group meetings
  - Three rounds of semi-structured interviews over 3 years
  - Artifact analysis
    - Examination of Wikis, software code, publications, websites
- Iterative qualitative data collection and analysis to elicit and develop themes

Two Research Data “Lifecycle” Models



UK Data Archive (2013)



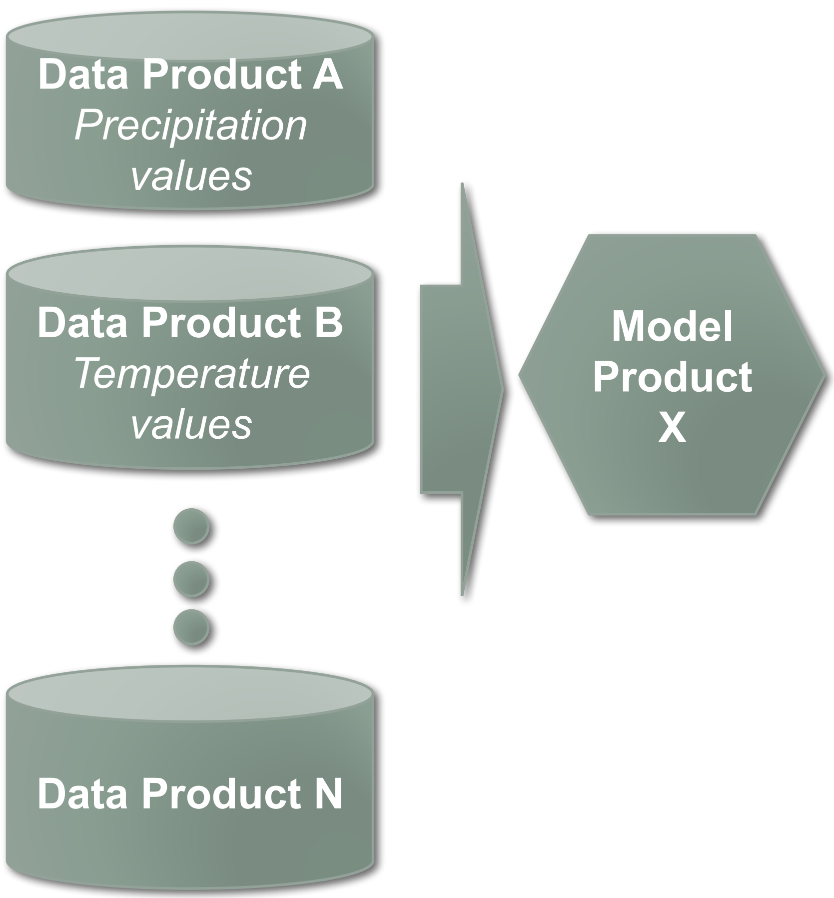
Wallis et al. (2008)

Models necessarily abstract and concretize stages, however work is much more fluid and interconnected!

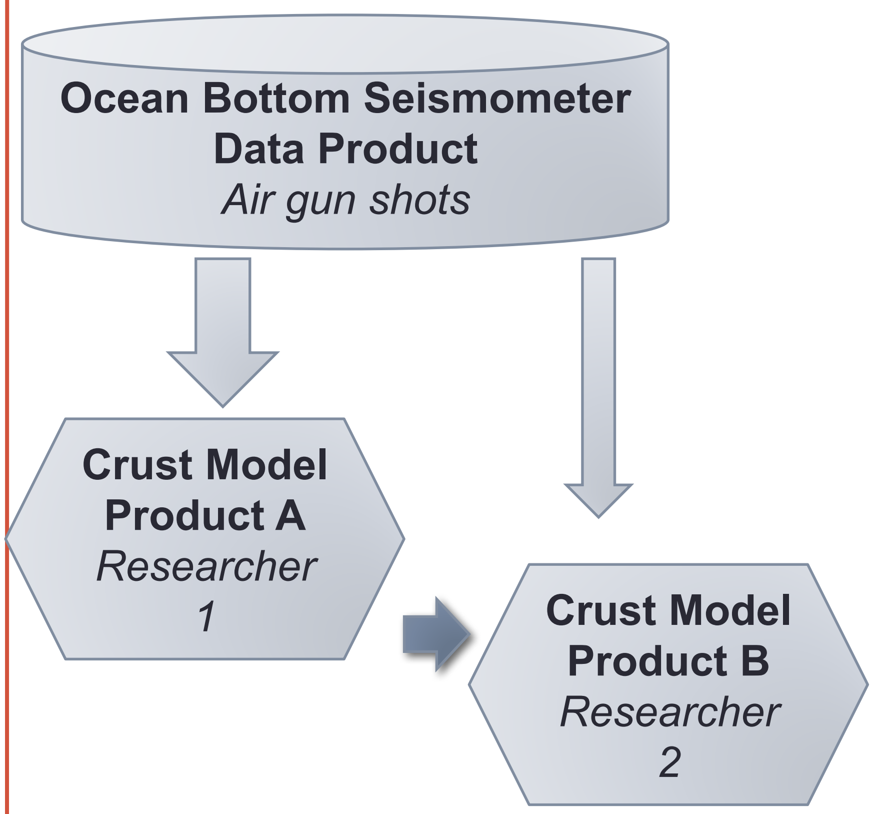
Paine, Sy, Piell, & Lee 2015

Assembling or Selecting a Subset of Data - Examples

Hank - Atmospheric Science



Waldo – Marine Geophysics

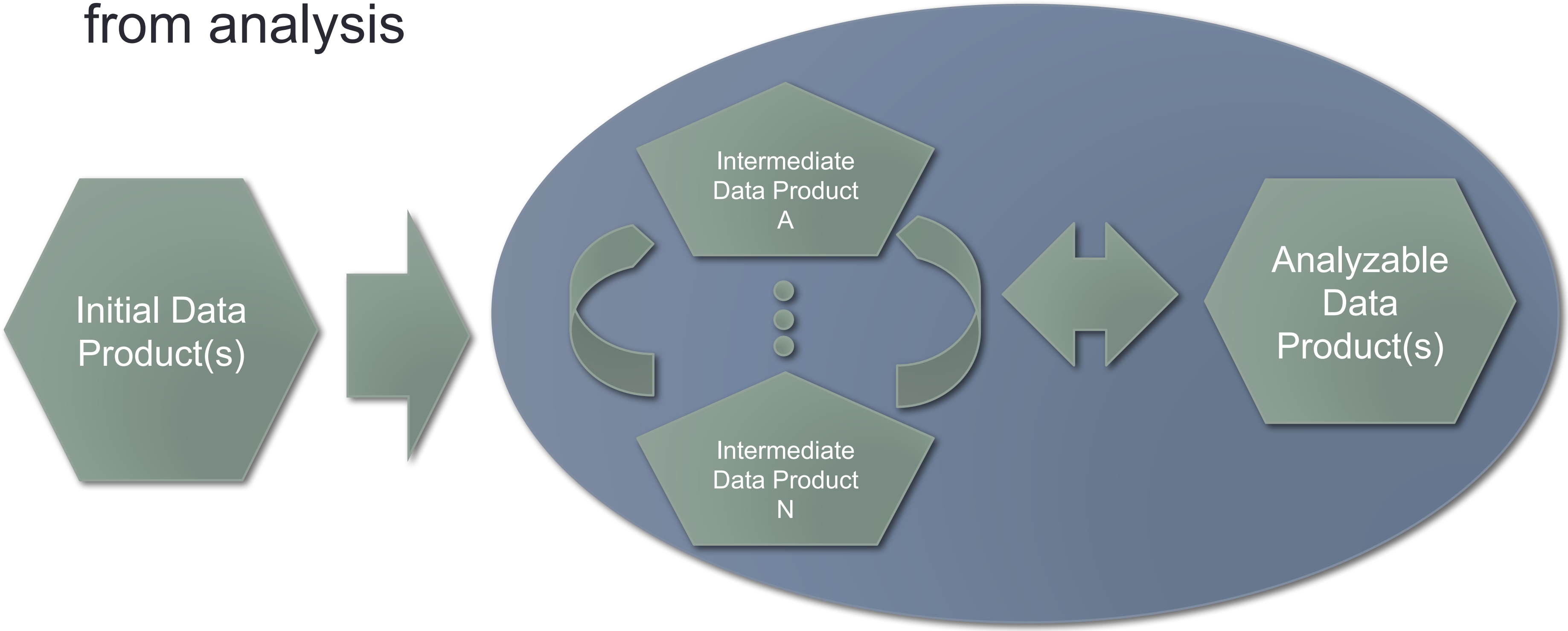


Assembling or Selecting a Subset of Data

- In Waldo’s group one data product supports multiple sub-projects, each researcher selects elements from it:
  - “… *we can go through and look at the arrival of each packet of energy from those air gun shots, and you just make little clicks along in the GUI in order to identify where that arrival is.” (Rollin, PhD student)*

Processing Work

- refines data products for particular scientific goals
- requires a high level of scientific and technical competence
- Often shapes and is shaped by iterations of research questions themselves—not necessarily separate from analysis



Paine, Sy, Piell, & Lee 2015

Coordinative Entities Framework (CEF)

How can we better understand and support how scientists collaborate?

- Multi-sited ethnography
- 4 coordinative "entities"
- prototypical organizational or cooperative work arrangements (Schmidt 1990)
- social system of work with different forms of interaction
- Identifying prototypical types and learning about differences between them

Entity	Key facets
Principal Group (PG)	1 Focal group of a CSCW study organized by a principal investigator who plans and runs projects to advance their research agenda 2 Composed of PI, students, postdoctoral researchers, research scientists, or other staff 3 Endures over time so long as the PI is engaged in research
Project Aggregation (PA)	4 Group organized or co-organized by a PI or member of their PG around a shared research question or goal 5 Composed of individuals from the focal PG working with individuals, groups, or organizations (including entities) from outside the PG 6 Endures for either short or long periods of time until shared questions or goals are resolved or put aside 7 Lacks formal organizational structure and depends upon continuing engagement between PGs
Project Federation (PF)	8 Formal partnership of individuals, groups, or organizations developing and using resources for multiple scientific questions or goals 9 Composed of members from a focal PG as well as other PGs, PAs, or organizations 10 Endures over time to sustain and use resources produced 11 Membership is formal with defined rules and requirements and a CSCW study's focal PG must be members
Facility Organization (FO)	12 Organization producing and sustaining resources in a general research area that a PG or PA can draw upon in their research projects 13 Distinct from other entities in CEF because it is a resource provider to PGs or PAs, it is not composed with members from a focal PG 14 Endures over time to ensure availability and durability of its resources

Entities by Formalization

Less Formal

Project Aggregations (PAs)

Formalization
&
Entrenchment

Principal Groups (PGs)
Project Federations (PFs)

Facility Organizations (FOs)

More Formal

Principal Group (PG)

- A **focal** group being studied
- Organized by a Principal Investigator (PI)
  - Has projects planned & run to advance PI's agenda
- Composed of
  - Undergraduate & graduate students
  - Postdocs
  - Research scientists
  - Other research staff
- Endures over time with PI's career

Project Aggregation (PA)

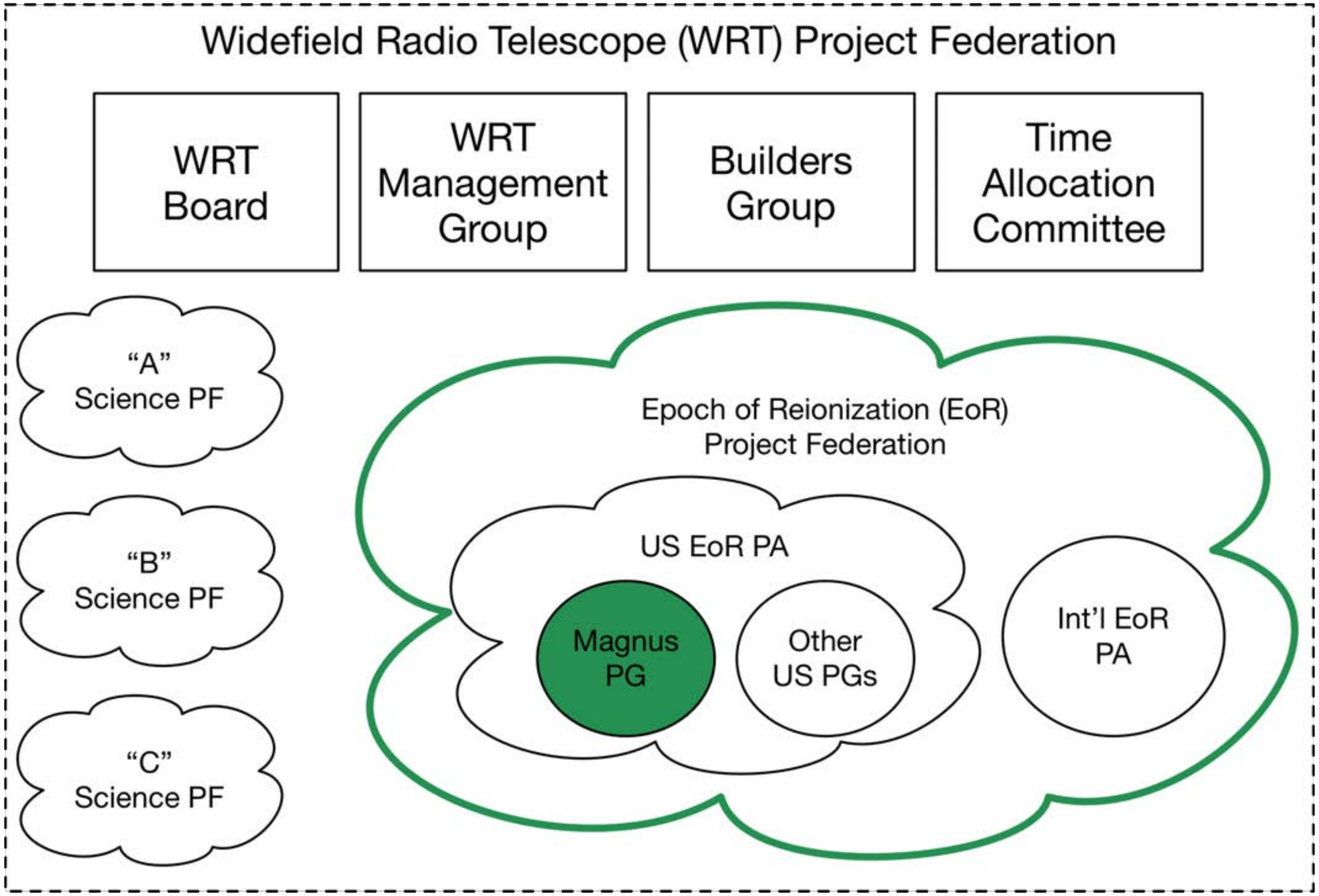
- Group engaging member(s) of one PG with member(s) of other PG(s)
- Organized around **shared** research goal or task
- Short or long-term endurance
  - Dissolves when questions/goals are resolved or shelved
- **Lacks** formal organizational structure
  - Depends on **ongoing** engagement between PGs

Project Federation (PF)

- **Formal** partnership of individuals, groups, organizations
- Developing & using resources for multiple research questions or goals
- Membership is formal with **rules & requirements**
  - Members come from constituent PGs, PAs, or organizations
- A focal PG must be members
- Endures over time to sustain resources

Facility Organization (FO)

- Organization producing & sustaining resources
- PGs and PAs rely upon FOs as a resource provider
- PGs or PAs are NOT members of FO’s organization
- Endures over time to keep resources available



- **Figure 1. Overview of the WRT Project Federation with different sub-PFs and their constituent PGs and PAs. Magnus's PG is highlighted in green, nested within the US EoR PA, the EoR sub-PF, and the overarching WRT Project Federation. We did not study the A, B, or C science PFs.**

		Data Collection	Data Processing	Data Analysis
Research Group	**Hank** Climate Modeling	Principal Group (PG) Facility Organization (FO)	Principal Group (PG) Project Aggregation (PA)	Principal Group (PG) Project Aggregation (PA)
	Waldo Underseas Seismology	Principal Group (PG) Project Aggregation (PA) Facility Organization (FO)	Principal Group (PG) Project Aggregation (PA)	Principal Group (PG) Project Aggregation (PA)
	Martin HIV Microbiology	Principal Group (PG) Project Aggregation (PA) Project Federation (PF) Facility Organization (FO)	Principal Group (PG) Facility Organization (FO)	Principal Group (PG) Project Aggregation (PA) Project Federation (PF)
	Magnus Cosmology	Principal Group (PG) Project Aggregation (PA) Project Federation (PF)	Principal Group (PG) Project Aggregation (PA) Project Federation (PF)	Principal Group (PG) Project Aggregation (PA) Project Federation (PF)

Conclusion

- Still know very little about the dynamics of scientific collaboration
- More complicated than imagined, but not intractable
- Need right people with the right tools

  →Climbing Everest →

We're working on it!

Papers

- Paine, Drew, & Lee, Charlotte P. (2014). *Producing Data, Producing Software: Developing a Radio Astronomy Research Infrastructure.* Paper presented at the 2014 IEEE 10th International Conference on e-Science (e-Science).
- Paine, Drew, Sy, Erin, Piell, Ron, & Lee, Charlotte P. (2015, March 24-27). *Examining Data Processing Work as Part of the Scientific Data Lifecycle: Comparing Practices Across Four Scientific Research Groups.* Paper presented at the iConference 2015, Newport Beach, CA, USA.
- Paine, Drew, & Lee, Charlotte P. (under review). The Coordinative Entities Framework (CEF): A Lens for Unpacking Scientific Collaboration.
- Lee, Charlotte P., & Paine, Drew. (2015). *From The Matrix to a Model of Coordinated Action (MoCA): A Conceptual Framework of and for CSCW*. Paper presented at the Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, Vancouver, BC, Canada.

Acknowledgements

This work has been supported in part by National Science Foundation grants IIS-0954088, ACI-1302272, OCI-1220269, IIS-0712994, and OCI-0838601

Thanks also to:

Ying-Yu Chen
Ilana Diamant
Scott D. Mainwaring
Ron Piell
Erin Sy
Study Participants

Computer Supported Collaboration Laboratory

HCDE Human Centered Design & Engineering

dub design: use: build: university of washington

UNIVERSITY of WASHINGTON