The Multidimensionality of Cyberinfrastructure: Objects, Practices, and Ideologies in the e-Science Community

Samantha Stein, Mona Sleiman, and Kerk Kee
COM499, Spring 2016 – Chapman University; Orange, CA

Introduction

As big data, e-science, and cyberinfrastructure (CI) gain increasing attention in the research community, there is a more pressing need to fully understand and define CI. There is great potential to advance the sciences through CI; however, we must first start by defining the components that are necessary for the full manifestation of CI.

Literature Review

In recent years, scientific research collaboration has become critically dependent upon effective access to and sharing of digital research data, especially as the e-science community is spread across a diverse range of disciplines (David, 2004). Previously, CI has been called "grid computing" and "collaboratories" (Lee, Dourish, & Mark, 2006). CI systems provide shared access to centralized or distributed resources and services, often in real-time (NSF, 2007). CI has the potential to enhance scientific collaboration because it will enable direct and shared access to more widely distributed computing resources than was previously possible (David, 2004). While the potential for scientific advancement through CI is undeniable, we must first fully understand cyberinfrastructure.

CI has been defined in previous literature by Atkins and colleagues (2003) as follows: "The newer term *cyberinfrastructure* refers to infrastructure based upon distributed computer, information and communication technology" (p. 5). Stewart (2007) defines CI as consisting of "computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible." In other words, CI is a complex system, involving a diverse network of interdependent technologies, remote instruments, big datasets, dispersed experts, diverse institutions, etc. (Kee et al., 2011).

Although the variety of definitions lists a range of components that make up CI, they neglect to define CI in terms of the over-arching dimensions involved. This investigation sets forth three over-arching dimensions: material objects, behavioral practices, and philosophical ideologies. This research serves to examine and define the dimensions with the following question: *"What material objects, behavioral practices, and philosophical ideologies collectively constitute CI?"*

Methodology

This poster employed the grounded theory approach (Corbin & Strauss, 1990) and analyzed 15 interviews conducted with members in the e-science community, such as directors, domain scientists, and PIs. Participants were from a diverse range of institutions and states across the U.S. Following a semi-structured protocol, interviews were conducted by telephone. Guided by the stated research question, the co-authors performed multiple iterations of data analysis and literature integration, yielding preliminary findings presented in this poster.

Findings

Material Objects

CI is constituted of hardware infrastructure and software infrastructure. Hardware infrastructure refers to the desktops/laptops, networks, HPC clusters, etc., while software infrastructure encompasses the software needed to run programs necessary for CI.

Hardware Infrastructure – Hardware includes the physical equipment necessary for the manifestation of CI (e.g., networks, workstations, etc.).

- "CI encompasses, beyond the desktop or laptop, some sort of computing facility or computing desktop that's beyond the typical desktop or laptop. So a workstation, a server that people would access, physical hardware." (Administrator, MA, 3/31/16)
- "I put into that category everything from the networks, the routers, the fibers, all that sort of stuff. The layer three type networking. There's the storage systems, there's the computers themselves, the high performance computer clusters, the desktops." (Technologist, IL, 3/22/16)

Software Infrastructure – Software includes the programs and other operation information created by a researcher used by a computer (e.g., HPC).

- "I think working equipment in general is something. I'm not coming from the hardware side I'm really coming from the software side, but if there's new hardware which will make my software faster." (Technologist, IN, 3/23/16)
- "You also have the layer of software and software tools that people will engage with… You have instrumentation. You got instrumentation, sensors, things like that." (Administrator, Washington D.C., 4/14/16)
- "That's what often many people think of as cyberinfrastructure, but what I think gets very short stripped in the community is all the software parts of cyber infrastructure." (Technologist, IL, 3/22/16)

Behavioral Practices

In the context of CI, behavioral practices bring the material objects to life. Such practices refer to: collaboration and financial management. As one informant expressed, "people drive the use of material objects. They help enforce the practices." (Administrator, CO, 3/21/16)

Collaboration – A collaborative 'ecosystem' is one where users share their research with peers in their discipline or across diverse disciplines, which ultimately drives collaborative success.

- "The different departments are connected with each other. It is very encouraged from the center for research computing to reach out to other departments for the research and work with biologists." (Technologist, IN, 3/23/16)
- "When everyone has to work for a goal in a healthy environment there's a lot of consensus about that. Everybody is trying to help everybody succeed. And if the collaboration can pull off that kind of collaborative success." (Administrator, IL, 4/5/16)

Financial Management – Financial management is crucial in CI, in that stakeholders must responsibly manage their finances in accordance with the funding source and established policies.

- "Taxpayers would be proud of funding. They would see it as a good use of their money and they would see it as a diligent and conserving of the resources and a wise use of tax money." (Administrator, IL, 4/5/16)
- "One thing I've never appreciated is more of the financial, administrative side of this best practices trying to govern in in terms of the inputs, in terms of growing a center to support administrative as well as research capabilities." (Administrator, MA, 3/31/16)

Philosophical Ideologies

In addition to material objects and behavioral practices, there are certain philosophical ideologies that drive CI. In other words, there are certain beliefs, mindsets, cultural values, and norms surrounding these objects and practices. Among the multitude of ideologies present in CI, two concepts emerge as prominent themes in the data: data sharing vs. data hoarding; big data-driven vs. theory-driven.

Data Sharing vs. Data Hoarding – In the world of CI, stakeholders face the dilemma of whether or not to share data and must take into consideration the norms of the field and the implication of their actions.

- "One of the themes was that anyone should be able to access any piece of data anytime to basically enable bright student from coming in and to have a chance to actually be a part of a discovery process." (Administrator, IL, 4/13/16)
- "Do I want to share my data or not? Everybody would like to say they want sharing of data, but in practice, there is often great competitive advantage to be got by not sharing your data…Everyone wants to receive everybody else's data, but I think a lot of people want to keep their own data as long as they can to themselves as their own competitive advantage." (Technologist, IL, 3/22/16)

Big Data-Driven vs. Theory-Driven – Science has traditionally been guided by theoretical research based on equations, logics, and thought experiments. CI presents a new way of research that is driven by big data (e.g., using population data, intensive computation, simulation and visualization).

- "It always used to be computational cycle-driven. But now, there is a lot more data-driven work and being able to meet the needs of our users and adapt to the current computational fields as they adapt." (Administrator, UT, 4/6/16)

Conclusion

In answering the research question *"What material objects, behavioral practices, and philosophical ideologies collectively constitute CI?,"* analysis revealed various subcomponents of each category. Specifically, the material objects include: hardware infrastructure and software infrastructure. Collaboration and financial management are the behavioral practices that bring these objects to life. Moreover, one must consider the ideologies, such as 'data sharing vs. data hoarding' and 'big data-driven vs. theory-driven,' which drive CI. This investigation lays the foundation for future research on CI adoption, diffusion, integration, and use by clearly defining the components of CI.

Future analysis of the material objects, behavioral practices, and philosophical ideologies may reveal certain relationships between each category. One approach may be to explore these components as a constellation (Lievrouw, 2006), considering them as intertwined. It is likely that further analysis will reveal how (and if) these components work together to collectively constitute CI. Furthermore, future investigations may reveal other categories beyond this preliminary list of three. This investigation uncovers that CI can be conceptualized as a sociotechnical and multidimensional system, in that it is constructed of not only objects, but also, practices and ideologies. Although the relationship between each component is not known yet, it is evident that each component plays a role in constituting CI, to some degree. Considering each of these dimensions, and eventually the relationships between, allows for a full understanding of CI, which will ultimately enable the full manifestation of CI.

References

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., …Wright, M. (2003). Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-ribbon Advisory Panel on Cyberinfrastructure. Washington, DC: National Science Foundation. Retrieved December 19, 2006 from http://www.communitytechnology.org/nsf_ci_report/.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, *13*(1), 3-21.

David, P., & Spence, M. (2004). *Towards Institutional Infrastructures for E-Science: The Scope of the Challenge.* Oxford Internet Institute.

Edwards, P. N., Jackson, S. J., Bowker, G. C. & Knobel, C. P. (2007). *Understanding Infrastructure: Dynamics, Tensions, and Design*. National Science Foundation. Retrieved from http://www.si.umich.edu/InfrastructureWxorkshop/ on 21 April 2016.

Kee, K. F., Cradduck, L., Blodgett, B., & Olwan, R. (2011). Cyberinfrastructure inside out: definition and influences shaping its emergence, development, and implementation in the early 21st century.

Lee, C. P., Dourish, P., & Mark, G. (2006). The human infrastructure of cyberinfrastructure. In *Proceedings of the 2006 20th anniversary conference on computer supported cooperative work* (pp. 483-492). ACM.

Lievrouw, L. A. (2006). New media design and development: Diffusion of innovations v social shaping of technology. In L. A. Lievrouw & S. M. Livingstone (Eds.), Handbook of new media: Social shaping and social consequences of ICTs (pp. 246-265). Thousand Oaks, CA: Sage Publications.

NSF. (2007). Cyberinfrastructure vision for 21st century discovery Retrieved October 15, 2009, from http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

Stewart, C. (2007). Indiana University cyberinfrastructure newsletter. Retrieved November 1, 2008, from http://racinfo.indiana.edu/newsletter/archives/2007-03.shtml

This material is based upon work supported by the National Science Foundation under Grant No. #1453864. IRB Approved: #1516H017.

This poster was presented on May 11, 2016 at the Chapman University Student Research Day, Orange, CA.