Social Science Research Towards Widespread and Sustainable Smart Cyberinfrastructures

Kerk F. Kee
College of Media & Communication
Texas Tech University
Lubbock, TX, USA
kerk.kee@ttu.edu

***Abstract*—This paper suggests that in order to build smart cyberinfrastructures towards being widespread and sustainable, the efforts would benefit from complementary insights from social science research. It proposes five areas of social science research – user adoption and systemic diffusion, diverse workforce development, co-production between developers and users, thriving online communities, as well as inclusive and ethical infrastructures. The paper concludes with two recommendations for the NSF to consider – creating a 'communication management plan' requirement under the broader impacts criterion, and the (re)development of a social science research funding program under the Office of Advanced Cyberinfrastructure.**

Keywords—social science research, sustainable cyberinfrastructure, technology adoption, widespread innovation

I. Introduction

The 2020 vision of making cyberinfrastructure (CI) 'smart' with artificial intelligence (AI) and machine learning (ML) is exciting. Smart CI will take e-science, computational social science, and digital humanities to new heights. Over the last two decades or so, CI has been steadily developing and maturing. This can be seen in the strategic investment of the TeraGrid in 2001, the publication of the Atkins Report in 2003, the initial establishment of the NSF Office of Cyberinfrastructure (OCI) in 2005-2006, the iteration of the TeraGrid into XSEDE in 2011, the reinventions of OCI into the Division of Advanced Cyberinfrastructure in 2012-2013 and the Office of Advanced Cyberinfrastructure in 2015-2016.

CI development started as a technical phenomenon. However, CI has also been argued as a socio-technical system [1]. This recognition is evident in the establishment of the Virtual Organizations as Sociotechnical Systems (VOSS) program under NSF OCI, which funded social science research on CI and technologies roughly between 2006 to 2013. With AI and ML being key components of smart CI, understanding the human and social dimensions is more important than ever.

This paper proposes five social science research areas that can help generate insights towards making smart CI widespread and sustainable over the long-term. The premise here is that the NSF can maximize its investments in smart CI when CI also achieves widespread adoption and has a thriving ecosystem around it. This outcome requires research insights and practical strategies derived from the social sciences, and social science research can complement CI's technical foundation.

This material is partially based upon work supported by the US National Science Foundation (NSF) under awards #1322305 and #1453864.

II. Social Science Towards Building Widespread and Sustainable Smart Cyberinfrastructures

A. User Adoption and Systemic Diffusion

A smart CI without users is void of data, activities, and outcomes. It is like a mall without customers, impressive in structures but without impacts on the economy. Moving from traditional experimental, theoretical, local, and small-scale research towards large-scale computational research with big data and empowered by smart CI requires a fundamental shift in the way researchers do their work. Meaningful use of smart CI begins with adoption by individual users.

User adoption occurs at the individual level, and for smart CI to have the greatest impact on research and its applications for scientific breakthroughs, the US economy, citizens' well-being, national security, etc., user adoption needs to expand into systemic diffusion. 'Diffusion' refers to the widespread of smart CI as a platform innovation in the overall research community [2]. Such a diffusion of innovations that transformed the research community have been seen in the adoption of personal computers in the 1980s and the Internet in the 1990s. Smart CI is initiating a new wave of transformations.

The need to conduct social science research to understand user adoption and systemic diffusion is the first strategic move towards becoming widespread. Le et al. [3] advanced the argument to debunk the romantic ideal of – "*If you build it, they will come*," which they termed a 'fallacy'. Without strategic promotions, smart CI may struggle with attracting users and attracting them quickly. Social science research on technology adoption and innovation diffusion [2] can support such an effort.

B. Diverse Workforce Development

To continue building on the metaphor of the mall for smart CI – a buzzing mall also needs shopkeepers, cleaning crew, and security guards to ensure the smooth operation of businesses for customers. For smart CI to be successfully implemented, understanding how to develop and maintain a diverse workforce is critical. Social science research has shown that teams with diverse members are generally 'smarter' – better at generating new ideas and increasing innovations in projects [4].

Smart CI needs a diverse and 'smart' workforce to facilitate user training and user support, following widespread adoption and diffusion. What are the effective strategies in training and

support? This remains an open research question. CI users do not usually come into a training workshop like a traditional student in the classroom – a blank slate to be given generic information about a topic. Most CI users come with specific questions and needs. Social science research on problem-based learning can optimize training. CI trainers also require having effective communication skills, beyond technical knowledge.

C. Co-Productions between Developers and Users

Generic user training is sufficient when users only need CI technologies 'straight out of the box'. However, for many projects, users need custom-made technologies for specific research questions, methodologies, and datasets. In these cases, developers will work with users to identify their needs and design custom-made technologies. The co-production process is complex because many users do not fully know what is possible and/or what they really need. It is through multiple meetings between developers and users that they identify the solutions together. Effective co-productions will require research insights from user-centered design and human-computer interactions.

To better identify users' needs, and the requirements to prototype and test new tools through multiple iterations, another research topic is identifying a software development methodology, such as agile software development (ASD), for effective co-productions. The argument here is not to promote ASD, as ASD and its variations have practical limitations. The point here is to highlight the need to study the social dimensions in co-production in the development of smart CI.

D. Thriving Online Communities

Smart CI can develop thriving online communities of researchers across time and space. With all the pieces interwoven into a robust CI, researchers within and across domains can share and integrate data now and over a long period of time, to do large-scale and longitudinal research otherwise not possible. However, understanding virtual organizations, interdisciplinary collaborations, international partnerships, and online community building are all complex human endeavors that would benefit from the social sciences, such as organizational communication, organizational sociology, and computer supported cooperative work. The goal is to create and sustain a thriving ecosystem to carry smart CI forward.

E. Inclusive and Ethical Infrastructures

Finally, a smart CI is also an ethical CI. It pays attention to who are included and excluded in the process of advancing CI research. Leigh Star and Geoff Bowker in science and technology studies wrote about the concept of the 'installed base' [5]. They caution against the activities in laying the foundation of any infrastructure, such that an 'installed base' with specific values and priorities are built into an infrastructure, creating a situation where future technologies and communities not fitting this base will be excluded. In the early phases of infrastructural developments, it is likely that decisions are made based on what makes sense for the immediate problems. It is likely that no groups are intentionally trying to exclude others in the future of the infrastructure being built at the moment. However, the argument here is to raise the awareness of how the decisions made together today will have future implications, and CI developers would be wise to keep the 'long now' [6] in mind.

Furthermore, for smart CI to be optimal, it requires continuous gathering and processing of data, including data with sensitive information. This is especially true in the case of biomedical research, for example. Also, university committees, such as the Institutional Review Boards (IRBs), may have strict policies on governing the collection and longitudinal use of sensitive data for research. In the case of medical research with smart CI, AI algorithms and ML techniques may rise to becoming 'deterministic'. How can patients' wishes and their human values not be override by AI and ML in the name of being 'smart' is an important ethical question. It would be wise for smart CI developers to be mindful of the human ethical dimensions of CI, and/or work with social scientists and critical scholars who can help them attend to this particular challenge.

III. Conclusion

This paper proposes five specific areas in social science research that can help build smart, widespread, and sustainable CI towards transformation of research across domains, locations, and time. Recounting the argument advanced by Le et al. [2] – "*If you build it, users may not come*". In order to address this concern, the NSF can consider two suggestions. First, similar to the requirement introduced in 2012 for proposals to include a data management plan, the NSF could consider adding a similar requirement (or an optional supplemental) for a 'communication management plan' under the 'Broader Impacts' criterion. Such a plan would encourage PIs to be more thoughtful about promoting user adoption and systemic diffusion, cultivating a diverse workforce, facilitating co-production between developers and users, building thriving online communities, and/or designing inclusive and ethical infrastructures in their smart CI projects. These social science topics are inherently 'communication' in nature. Second, the NSF may consider (re)developing a funding program similar to VOSS, and encourage more involvement of social scientists who can carry out research on the five areas (and other important topics) under OAC, towards a strategic and collective effort of building smart CI in a sustainable way with widespread adoption and successful implementation. This program should be under OAC as the context of CI is unique. Social science conducted within the context of CI is critical for accurate applications.

References

[1] C. P. Lee, P. Dourish, and G. Mark, "The human infrastructure of cyberinfrastructure," CSCW '06: Proceedings of the Conference on Computer Supported Cooperative Work, pp. 483-492, New York: ACM Press, 2006.

[2] K. F. Kee, "Adoption and diffusion," In International encyclopedia of organizational communication, C. Scott and L. Lewis Eds. Hoboken, NJ: Wiley-Blackwell, 2017, pp. 41-54.

[3] B. Le, F. Escalera, K. Jitkajornwanich, and K. F. Kee, "External communication to diffuse science gateways and cyberinfrastructure as innovations for research with big data," Gateways Conference, San Diego, CA, 2019, available at: https://osf.io/mz8bv/

[4] D. Rock, and H. Grant, "Why diverse teams are smarter," Harvard Business Review, 2016, available at https://hbr.org/2016/11/why-diverse-teams-are-smarter

[5] S. L. Star, and G. C. Bowker, "How to infastructure," Handbook of new media: Social shaping and social consequences of ICTs, L. A. Lievrouw & S. M. Livingstone, Eds. Thansand Oaks, CA: Sage, 2006, pp. 230-245.

[6] D. Ribes, and T. A. Finholt, "The long now of technology infrastructure: Articulating tensions in development", J. of the Assoc for Information Systems, vol. 10, pp. 375-398, 2009.