

THE 10 ATTRIBUTES THAT DRIVE ADOPTION AND DIFFUSION OF COMPUTATIONAL TOOLS IN E-SCIENCE

Kerk F. Kee
Chapman University
School of
Communication
Organizing,
Communication, &
Technology (OCT)
Group
One University Drive
Orange, CA 92866
+1-949-391-9988
kerk.kee@gmail.com

Mona Sleiman
Chapman University
School of
Communication
Organizing,
Communication, &
Technology (OCT)
Group
One University Drive
Orange, CA 92866
+1-714-532-6036
sleim100@mail.chapman.edu

Michelle Williams
Chapman University
School of
Communication
Organizing,
Communication, &
Technology (OCT)
Group
One University Drive
Orange, CA 92866
+1-714-532-6036
willi310@mail.chapman.edu

Dominique Stewart
Chapman University
School of
Communication
Organizing,
Communication, &
Technology (OCT)
Group
One University Drive
Orange, CA 92866
+1-714-532-6036
stewa167@mail.chapman.edu

ABSTRACT

As the computational movement gains more traction in the scientific community, there is an increasing need to understand what drives adoption and diffusion of tools. This investigation reveals what makes a computational tool more easily adopted by users within the e-science community. Guided by Rogers's [1] Diffusion of Innovations theory, we set out to identify the innovation attributes of a range of computational tools across domains. Based on 135 interviews with domain scientists, computational technologists, and supercomputer center administrators across the U.S. and a small portion from Europe, systematic analysis revealed 10 key attributes of tools. They are: driven by needs, organized access, trialability, observability, relative advantage, simplicity, compatibility, community-driven, well-documented, and adaptability. We discuss the attributes in the form of questions stakeholders should keep in mind while designing and promoting the tools. We also present diffusion strategies associated with each attribute. The 10 attributes and associated questions can serve as a checklist for e-science projects that aim to promote their computation tools beyond the incubators. This paper is submitted to the "Software and Software Environments" track because it has implications for engagement of user communities.

CCS Concepts

• **Software and its engineering~Designing software** • Software and its engineering~Open source model • Social and professional topics~Computer supported cooperative work • Social and professional topics~Socio-technical systems

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

XSEDE16, July 17-21, 2016, ,

© 2016 ACM. ISBN 978-1-4503-4755-6/16/07...\$15.00

DOI: <http://dx.doi.org/10.1145/2949550.2949649>

Keywords

Technology adoption; innovation attributes, strategic diffusion; Diffusion of Innovations theory; qualitative organizational studies; science & technology studies.

1. INTRODUCTION

High performance computing, high throughput computing, distributed/parallel computing, computational techniques, and cyberinfrastructure (CI) for large-scale science has a long history in the United States. Freeman [2] maintains that the concept of infrastructure in reference to labs, equipment, support personnel, etc., for conducting science did not become commonly used until the 1950s with the establishment of a National Science Foundation (NSF)-sponsored polar studies facility in Antarctica. However, he suggests that the accurate starting point of CI development was in the 1960s, when the NSF financially supported the establishment of academic computing centers on several U.S. campuses. The unique nature of these academic computing centers was that they were open to the general scientific community and not limited to specific projects. This allowed numerous faculty and students access to the facilities created with NSF funding.

In the 1980s, NSF made two strides that further strengthened the foundation of today's CI development. First, NSF invested in the creation of the first few supercomputer centers in the country [2, 3]; the establishment of the first supercomputer center sparked the vision of the open scientific community in the early 1980s. Through its Supercomputer Centers Program, NSF established five supercomputer centers across the U.S. between 1985 and 1986. These five centers include San Diego Supercomputing Center (SDSC) at the University of California at San Diego; National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign; Pittsburgh Supercomputing Center (PSC), a joint effort of Carnegie Mellon University and the University of Pittsburgh together with Westinghouse Electric Company; the Cornell Theory Center at Cornell University; and the

John von Neumann Center at Princeton University [4]. After about a decade, the NSF decided to focus its resources and funding on only three centers, and the Cornell and Princeton centers ceased to receive NSF funding in 1997 [5]. The remaining three centers, SDSC, NCSA, and PSC, continue to be the leading supercomputer centers in the present CI development in the U.S.

The second initiative that strengthened the foundation of today's CI development, was NSF's investment in distributed computing experiments and programs in late 1980s. Freeman [2] documented the Coordinated Experimental Research (CER) Program started in the late 1980s as one of the most important programs during that time. In fact, many basic concepts in distributed computing that are important to current CI development came out of CER.

Perhaps the most recent and familiar development of CI is the creation of the Internet. Kahin and Jackson [6] claim the term 'cyberinfrastructure' was initially used as a shorthand for "Internet-based information infrastructure" (§ 1). Leiner and colleagues [7] credit the beginning of the Internet to ARPANET (Advanced Research Projects Agency Network), a network established by the Defense Advanced Research Projects Agency (DARPA) in the 1960s. Freeman [2] documented another critical beginning of the Internet as David Farber's distributed computing project at the University of California, Irvine, in the early 1970s, a project funded by NSF. Later, the project developed into Theory Net and then CSNet in the 1970s, and then NSFNET in the 1980s [8, 2, 9]. In the 1990s, along with the deployment of NSF supercomputer centers in the country, NSFNET eventually became the Internet we know today.

The revolutionary turning point of CI emergence traces to 1999 with the creation of the Open Science Grid (OSG), which is considered one of the earliest and most successful CI projects [10]. During that year, domain scientists working on large-scale projects in physics and astronomy in conjunction with computer scientists (with prior experience in distributed computing) came together and discussed the development of a grid-based computing infrastructure to facilitate data-intensive experiments in physics and astronomy.

Shortly after the initial conception, three pioneering grid projects were funded: Particle Physics Data Grid (funded by the Department of Energy in 1999), GriPhyN (funded by the National Science Foundation in 2000), and the International Virtual Data Grid Laboratory (funded by the National Science Foundation in 2001). Due to the significant overlap of human and institutional representations in these projects, they began to consolidate resources and merge efforts by creating a national-scale grid cyberinfrastructure called the Trillium Consortium in 2002. In the following year, the consortium grew steadily and created the Grid3 prototype CI project to support disciplines beyond physics and astronomy. It ran 1,000 concurrent applications successfully in October/ November 2003.

Trillium continued to expand for two more years with recruitment of more domain and computer scientists interested in this new development. This expansion led to additional funding from the National Science Foundation and the Department of Energy for the official establishment of the Open Science Grid (OSG) on July 20, 2005, while funding for the Trillium consortium (and its three inception projects) expired in 2006. OSG is still in stable operation today.

The CI movement has gained tremendous visibility in the US since the early 2000s. While there are many factors behind this

movement, one can argue that the emergence could be attributed to Atkins and colleagues' [3] NSF blue ribbon report on cyberinfrastructure in 2003, the establishment of the Office of Cyberinfrastructure in 2006 at the NSF [11], and the reinvention of this office into the current Division of Advanced Cyberinfrastructure under NSF's Computer & Information Science & Engineering Directorate in 2012 [12]. Over the years, a significant investment has gone into CI development by funding a range of CI and e-science projects [13, 14].

Stewart [15] 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. [16]. Among these interdependent technologies are computational tools designed to analyze big datasets. Computational tools enable scientists to harnesses big data and explore complex, large-scale problems. Although the concept of CI encompasses many socio-technical components, the focus of this paper highlights the particular role of computational tools.

A unique aspect of CI and e-science projects is the development of specialized computational tools for investigating grand challenges in science. In other words, these projects bring together multidisciplinary experts with knowledge across different domains and computer/computational sciences to develop tools to process large-scale data and solve problems at the frontier of science. This work will actively benefit research, society, and humanity. Given the demand for computational tools that will advance research, it is critical to understand what makes a tool diffuse, ultimately enabling researchers to carry out their work.

In an effort to advance research computing, the current paper reports the findings from an NSF-funded organizational study looking at the adoption and diffusion of computational tools in the scientific community. More specifically, we seek to answer the research question, "*What attributes promote the adoption and diffusion of computational tools for e-science?*"

2. DIFFUSION OF INNOVATIONS

In order to answer the research question, we turned to the Diffusion of Innovations theory by Everett Rogers ([1]; hereafter referred to simply as diffusion theory). Diffusion theory is the most cited communication theory, and the second most cited social science theory to date [17, 18]. The theory describes a complex model based on innovation attributes, adopter categories, diffusion networks, and opinion leadership which enables the prediction of adoption and diffusion of a wide range of innovations.

Diffusion theory describes diffusion as the communication process underlying the natural spreading of an innovation (i.e., a new technology, behavior, or idea) over time within a social system. On the other hand, adoption is an individual decision to accept or reject an innovation, along with its subsequent implementation, discontinuance, and/or modification. Therefore, adoption is an individual process that leads to diffusion as a systemic process [17].

Rogers classified five types of adopters based on their timing of adoption. Their distributions on a normal curve (see Figure 1) are indicated by the percentages after the categories, based on the

percentages associated with standard deviations of the normal curve of innovation adoption: innovators (2.5%), early adopters (13.5%), early majority (34.0%), late majority (34.0%), and laggards (16.0%). He argued that these five types of adopters have different characteristics and tendencies that create segmentations in the marketplace.

Innovators (2.5%), also known as technology enthusiasts, are usually venturesome. They are knowledgeable of complex and technical matters. *Early adopters* (13.5%), also known as visionaries, adopt an innovation after a careful evaluation. Therefore, others respect them and like to seek out the opinions and evaluations of early adopters. Individuals in the *early majority* (34.0%) are also known as pragmatists; they usually get on board right before the average members of a system do. Those in the early majority usually considered the innovation for some time before adopting it. Members in the *late majority* (34.0%), also known as conservatives, are always cautious of anything new. For those in the late majority, adoption is usually the result of an economic necessity or peer pressure. Lastly, *laggards* (16.0%), also known as skeptics, tend to be suspicious of innovations and the people who try to persuade them to adopt. They are also often the last people to learn about an innovation.

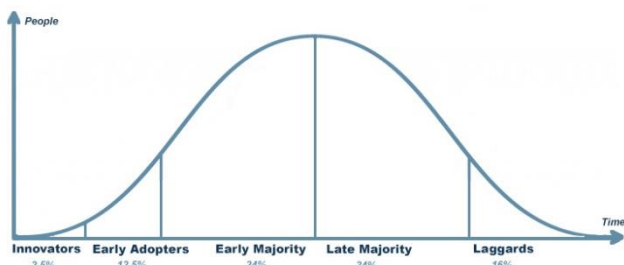


Figure 1. Five Types of Adopters

Another important concept in diffusion theory is opinion leadership. Opinion leadership refers to the informal influence exerted by a particular subset of members (usually about 5%-8% of the total population) within a social system. These members, called opinion leaders, are usually respected by their peers. Therefore, their opinions about new issues are highly influential. In fact, members actively seek out their opinions. In the case of a new technology, an opinion leader's comments can sway others' opinions about the new technology, either for or against it.

Opinion leaders are often found in the early adopters category (as opposed to the innovators category) because early adopters are characteristically closer to the rest of the population and have more in common with their peers. Innovators are too different from the rest of the members, and are thus often regarded as extreme individuals. When innovators endorse a new technology, the majority of other members may further think that the technology is not for them, because the innovators' needs and circumstances are very different from the regular members. This is also the reason why sometimes using the first few adopters to help promote a technology further alienate the technology for most of the population.

Once a new product attracts a substantial number of adopters, a condition to spark a takeoff of the product is referred to as achieving critical mass. Achieving *critical mass* refers to attracting the number of users necessary to influence and trigger a rapid and wide adoption by other members within a social system. The concept of critical mass implies that adopters have varying network thresholds that prevent individual adopters from acting until the

thresholds are met, at which point adoption occurs. When more people in a social system adopt, more non-adopters' thresholds are met, triggering a domino effect of adoption. These concepts of diffusion theory have been observed in a wide variety of technologies over the decades. In this paper, we use diffusion theory to explore the adoption and diffusion of computational tools for e-science.

3. METHODOLOGY

The 10 key themes reported in this study came from a pool of 135 in-depth interviews with a mix of domain scientists as lead users, computational technologists as developers, and supercomputer center administrators as key facilitators. The majority of participants came from across the U.S. (n = 130), and a small portion from Europe (n = 5).

The first interview was conducted on November 16, 2013 and the last one on July 29, 2015. Ninety-five interviews were conducted in person at 4 conferences: Supercomputing 2013 in Denver, XSEDE 2014 in Atlanta, Supercomputing 2014 in New Orleans, and XSEDE 2015 in St. Louis. Forty telephone interviews were conducted year-round based on simple random sampling technique (using a random number generator) and the conference speaker lists of XSEDE conferences in 2013, 2014, and 2015. The semi-structured interviews were guided by a pre-designed protocol; however, the semi-structured nature allowed for flexibility to adapt to each informant's unique background.

After the interviews were transcribed verbatim, we systematically coded the transcripts on QSR International's NVivo qualitative analysis software [19] using the grounded theory approach [20]. According to grounded theory, coding was bottom-up (based on what informants said in response to open-ended questions), instead of top-down (based on pre-determined categories or hypotheses). For example, analysis begun with selective coding [20], during which all the transcripts were carefully read, and large sections of text were coded based on the relevant research questions that were discussed throughout each of the interviews. Selective coding involved flagging text based on major themes of the project, which in turned allowed some sections of text to belong to multiple themes and/or research questions.

When we wrote this paper, based on the code of "Innovation Attributes," we pulled all of the text related to this theme, and systematically coded/flagged quotes to identify the key themes presented. Selective coding on NVivo is akin to sorting photographs into different albums on Facebook. If you tag each photograph based on what is presented in the image, sometimes the same photographs will simultaneously appear in different albums because they do not strictly belong to one album. That way, while looking at each album, viewers get photos that are relevant to each event or tag they were placed under.

After selective coding, open coding and axial coding [20] allowed us (the authors) to compare and contrast emerging themes, and group them into sub-themes. For example, the theme of "organized access" was a sub-theme for innovation attributes and reveals the need for tools to be featured on an organized site. For this particular code, the tool and/or site needs to be accessible to potential users as well.

To ensure reliability, selective coding was first performed by one team member, then reviewed by two other team members. Subsequently, key themes were identified and representative quotes were pulled to illustrate each theme. Finally, the PI and the key

interviewing team member discussed and finalized the themes. As a team, we met multiple times to ensure collective consensus on the final themes before writing the paper. This bottom-up grounded theory technique is especially suitable when the purpose of the research is to explore and discover a new topic from the presented data.

The final prominent themes came from multiple iterations of the explained analysis. They were determined based on Owen's [21] criteria of frequency, repetition, and forcefulness. Emerging themes were constantly compared and contrasted throughout the interlocked interview and analysis processes. Staying true to the grounded theory philosophy, the protocol went through two major revisions and multiple minor revisions to better adapt to emerging themes. In other words, instead of being constrained by pre-formulated hypotheses to be tested without alteration during data collection, our study evolved as more information was collected from the interviewees. For example, the original research proposal only considers innovation attributes (what is reported in this paper) and governing macro conditions (data reserved for another paper) are sources of influence on the adoption process. However, one of the first informants responded to our questions and suggested that the organizational attributes of the e-science team also play a key role the adoption process. We quickly adapted the interview protocol to include questions about organizational attributes, which yielded findings that will be reported in another paper. The rationale for using this flexible (grounded theory) approach is to remain open minded during data collection. Interview questions were also broken down to allow for more participant understanding and specific responses. Finally, one earlier informants commented on how the interview questions were written in the style of a formal proposal rather than a natural flowing conversation, which was better suited for this research. Therefore, academic language was scaled back and more plain language was used to increase comprehensibility and reduce priming effects.

4. FINDINGS

Based on systematic analysis of the investigative interviews, 10 themes emerged to answer the research question, "*What attributes promote the adoption and diffusion of computational tools for e-science?*" We present the themes in a sequence that we believe likely to unfold for potential users. While presented in a sequence, it is important to keep in mind that the 10 attributes influence each other in a dynamic and complex fashion. The attributes discussed came from informants' descriptions, reflections, and suggestions that helped us formulate what it may take for a computational tool to become successful in attracting new users and diffuse from one project to another to establish widespread adoption. We present the attributes in the form of critical questions that lead users, developers, and key facilitators as stakeholders should keep in mind while promoting their tools. Furthermore, we also present associated diffusion strategies with the attributes.

4.1 Driven by Needs

A computational tool is typically designed to solve a particular complex problem or a grand challenge in science. In other words, a computational tool meets certain needs. Given the demands of large-scale scientific data, domain scientists turn to computational resources to better carry out their research. Each group of scientists has particular requirements for computational tools. Thus, developers must cater to the needs of the users. A technologist from Colorado highlights the importance of addressing scientists' needs:

"Scientists don't need cool things. They need things that provide tangible benefits to solving their immediate problems. And so the first thing I look for is: what does a scientist actually need? What are they trying to do? What are the problems that they're facing? And [then] solve a problem that actually exists." (November 19, 2014)

After a tool has been developed to address a specific in-house problem, technologists are often interested in promoting the tool beyond the incubator. That is when the adoption issue becomes a challenge for the inception team. During this phase, stakeholders should put themselves in the mindset of potential users and ask themselves: "*Does this tool meet my needs?*" Similar to how the inception group developed a tool to meet a specific need, potential users will also ask if the tool will meet their research needs. A diffusion strategy in this case is to help potential users identify their pain-points, and explore how a tool can help them address the pain-points.

4.2 Organized Access

Organized access refers not simply to a tool being available for potential users, but it also means that a computational tool can be found and accessed on an organized marketplace. Given that open source tools are technically free and available for all, everyone has access to them. However, even if a tool is open source and available online, it does not necessarily mean that the tool can be easily found by users. Therefore, the tool should be posted on an organized marketplace, increasing the visibility of the tool itself. A research scientist from Indiana explains the function and value of an organized marketplace:

"We kind of look at the hubs, like with HUBZero, as sort of a marketplace. On the one hand, they're suppliers and the suppliers are building apps or applications/tools and they're putting them up in the supply chain. And then on the other side, you have consumers using those tools or resources that people are posting... You need ratings and reviews. You need to be able to submit bug reports, you might have questions because you're not sure how it works, you need a little bit more help, or you have ideas for improvement. So it's very much organized in those terms." (December 4, 2014)

With the convenience, design features, and reputability of an established marketplace, users can more easily find and access tools. Users can also take advantage of the resources available on the marketplace websites in order to use the tools with more ease. Furthermore, it is important for users to have easy access to the tools so that the benefits of open source collaboration can materialize. In order for stakeholders to make the tool very accessible to potential users, they need to ask themselves: "*Is this tool easily available?*" and "*Can I find it at a known location?*" A diffusion strategy in this case is to post the tool on an organized marketplace, such as HUBZero or the Galaxy Toolshed, where potential users may look for tools that meet their needs. Additionally, the organized online marketplace should feature ratings and reviews to provide context for potential users to assess if a particular tool would fit a given problem and/or context.

4.3 Trialability

Trialability refers to whether a potential user can experiment with the tool, prior to full adoption. For most traditional technologies, financial cost and time investment pose barriers to experimenting with a new tool. Fortunately, due to the open source nature of most computational tools in e-science, financial cost of the tool is not a factor. However, the barrier of time remains to be one of the biggest challenges to get potential users to experiment with computational tools. An administrator from Tennessee expresses the value of trialability in influencing adoption decisions:

“It’s a little more difficult to get...people to try it. And once they try it, they will like it, and it goes on.” (July 16, 2014)

In order to increase trialability, stakeholders should put themselves in the shoes of users and think: “*Can I try this tool without much investment of time?*” Using this type of strategy can provide timely support for users and some ‘hand-holding’ in the early diffusion cycles to further the diffusion of the tool.

4.4 Observability

Observability refers to the visibility of a computational tool in one’s environment. In other words, whether potential users see others using the same tool, or if people they respect talk about the tool in social conversations or professional presentations.

French sociologist Tarde wrote about the law of imitation in 1903 [21]. He identified the human nature of imitating the people around us: we are always under the influence of those we associate with. In the context of e-science, scientists adopt tools that they observe their peers using. This may be due to the fact that the tool has been tested and vetted by their trusted colleagues. An administrator from Scotland describes the process in which tools become visible through ‘word-of-mouth’:

“Say they are projects, if your tool is successful and people beyond your own projects aren’t using it, then the hope is that it’ll be word-of-mouth that will spread and it’ll be used in many different projects... spread through recommendations from other people. That’s generally how it works in this community, at least. You use what other people recommend, because it is quite a small community. It is very niche, so word-of-mouth still works very well.” (November 18, 2013)

When considering a tool, potential users may ask the question, “*What are my peers using, and how are they using it?*” A diffusion strategy in this case is to maintain an active website/forum, and heavily engage in demonstrations at professional conferences with strong papers and publications by their peers. The strategy of maintaining an active website/forum goes hand-in-hand with the aforementioned strategy of making a tool accessible in an organized marketplace.

4.5 Relative Advantage

Relative advantage refers to the comparison between an innovation and an existing technology, in terms of usefulness and benefits. For users who already belong to the CI and e-science movement, they will be comparing a new tool with their existing tool to assess if they should switch over to the new tool. This comparison will be based on whether the new tool is going to offer advantages over

their existing tool. For potential users who are not yet doing e-science, they may need to assess if the tool (along with the new way of doing science, the computational way) is more advantageous than the way they currently do research.

In organizational studies, there is a concept called the “9X Effect” [23]. This concept means for people to give something up, the new idea needs to be at least 9 times better than the existing one. In science, the main goal is to publish more research, and more importantly for some, publishing better quality or groundbreaking research. Therefore, for a scientist to give up an existing tool or their current way of doing research, the innovation has to be multiple times better than the current ways. An administrator from Scotland explains the concept of relative advantage in the e-science community:

“Can they be assured that this is doing things more efficiently or solving a larger problem or more complex problem than they could do before? Because if it isn’t, it’s not worth the time investment.” (November 18, 2013)

The question stakeholders should keep in mind on behalf of potential users is: “*Is this option a lot better than what I have right now?*” A diffusion strategy in this case is to generate a solid list of advantages associated with a computational tool; the 10 attributes documented in this paper is a good start.

4.6 Simplicity

When potential users consider a new tool, they want to make sure that the tool is simple and easy to use. Most, if not all, scientists are busy; they work long hours and they are racing with their peers towards the next great discovery. Undoubtedly, time is of the essence. Therefore, the simpler the tool, the less time they have to spend on learning how to use the tool. An application specialist from Georgia describes the importance of simplicity for potential adopters:

“There are specific sets of tools that may be really good at solving a problem in a unique way... Typically, you will pick it up and say, ‘This looks cool, it does what I want,’ – like I was talking to somebody ...and he said, ‘It looks really cool but I can’t figure out how to use it,’ then you are not going to pick it up.” (July 16, 2014)

The question stakeholders should ask on behalf of potential users is: “*Is this tool easy to use?*” In other words, usability is critical. In this case, a diffusion strategy could be to keep testing a tool with willing users across domains and varying levels of technical knowledge, so the tool can be further fine-tuned to achieve elegant simplicity for a range of potential users, not simply those with high technical skills and deep computer knowledge.

4.7 Compatibility

Compatibility is the perception that an innovation will work smoothly for an adopter (i.e., not disruptive). Scientists are like most other professionals in that they are busy individuals. When an innovation is not compatible with their existing routines and/or practices, it becomes non-productive for scientists to adopt an innovation. For instance, scientists may think that learning how to use a new tool will actually take time away from their existing work and responsibilities. It is also important to note that most scientists do not work alone; therefore, they must take into account what will

work for them and also their collaborators. A technologist from Colorado explicitly states the importance of developing tools that are compatible with scientists' existing practices:

“You have to build something that doesn't create barriers. So you have to somehow make it fit into the way they're already doing things... Because if the scientist has to change a lot of what they're doing, they're not going to adopt the tool.” (November 19, 2014)

Evidently, a stakeholder must put themselves in the mindset of users and ask: “*Can I easily integrate this tool into my existing routine and collaborations?*” When a potential user is considering a new tool, he/she will weigh the pros and cons as benefits vs. costs. When a tool is compatible, there will be fewer costs to adoption. A diffusion strategy for ensuring compatibility is to recruit adopters from multiple domains in the early phase to help make suggestions for how a tool can adopt to a range of scenarios related to compatibility.

4.8 Community-Driven

Most of the computational tools available today for e-science are open source technologies. In the open source community, a developer and/or their team will share the code online, and the community will try it out, integrate the tool, fix the bugs, update the codes, improve its functionality, and extend its usage.

A powerful tool is one that thrives among an active community of users. However, Shirky [24] reported that most Internet-based groups fail due in part to a lack of planning. Ultimately, the success of a tool is driven by whether the user community is participatory. When someone has a question, there is a group of other users out there who can help answer some questions and/or share their adoption experience; this group must be actively engaging in online groups to help address the needs of others. If there isn't an active community behind a tool, potential users may be hesitant to adopt, because it means that they may not be able to get community support in a timely fashion, and the tool may not be updated constantly. A scientist-developer/administrator from Indiana describes the value of the community:

“The tools that are really taking off are the tools that enable the user-community to participate in the development, because any tool that is going to be successful in creating a virtual organization around it is going to be successful in part by empowering the community to grow and have their own needs met by having access to some of the code and being able to implement those codes.” (October 15, 2014)

In examining the sustainability of a tool, stakeholders need to keep in mind the question: “*Is there an active community that will carry this tool forward?*” In this case, a diffusion strategy is to partner with communication professionals (in public relations and/or marketing) to intentionally create an active user community [25].

4.9 Well-Documented

A well-documented tool has both the historical record keeping, and the helpful user guides. Historical record keeping allows for historical documentation of the tool, such as various versions and updates over time. User guides include instructional material with examples for using the tool. Moreover, good documentation

includes prime examples to help potential users imagine how they could use the tools to help with their own science. A technologist from Illinois describes the documentation process can lead to potential adoption:

“The Open Grid Forum is an example of the community coming together and documenting specifications and interaction patterns for their software systems that has been very valuable in connecting different software components together. And so not just documenting how to use the software, but documenting the interfaces, the network protocols and doing things in a standard way... That's really crucial to enabling the software components to be collected together and then that's really important for adoption.” (July 16, 2014)

Stakeholders should ask themselves: “*Is the tool well-documented with a track-record and user-guides?*” A diffusion strategy in this case is to specifically assign at least one person to create and constantly update the documentation. This can be achieved by allocating specific funding for this role. Another strategy is to create multi-media tutorials to help new users to learn about the tool. With such an increase in the use of Web 2.0, users and developers may find such documentation more useful and adaptable.

4.10 Adaptability

Adaptability refers to a tool's flexibility to cross domains and/or find novel usage in a new context unintended by the original stakeholders. Most computational tools are developed to address a specific grand challenge in science, which is how a project gets funding to launch the effort in the first place. As most funders have a specific domain interest (e.g., NSF for mathematics and scientific research, National Institutes of Health for medical and health research, Department of Energy for nuclear and energy research, Department of Defense for military and defense research, etc.), funded projects develop their tools for a specific domain to start. However, if a tool can only be used to address problems within a specific domain and for a particular niche problem, its diffusion will be limited. In order for a tool to achieve a high status of diffusion, it needs to spread beyond its original domain and/or problem. A technologist from Georgia offers an analogy to describe the concept of adaptability:

“A hammer's use is generally to push in nails, and it's not meant to screw things in. But when it comes to virtual [computational] tools, like tools used in software, it's different and you want your Swiss Army knife instead. It might have one main purpose, but it can be flexible enough or it can bend enough to serve another purpose without having a major software rewrite and things like that.” (November 18, 2014)

Scientists have colleagues across domains and institutions; when they talk about a tool with others, it is an opportunity for the tool to jump from one domain and/or institution to another. It would be wise for stakeholders to think on behalf of potential new users: “*Can I take this tool from that domain and bring it into my domain?*” Stakeholders should not only keep in mind the original users, but also consider groups that are un-thought of at first. In this case, a diffusion strategy is to strike a balance between being

specific (for the inception project) and adaptable (for future adopters' projects) from the start. Instead of thinking about this as an "either-or" situation, one needs to make it a "both-and" choice.

5. DISCUSSIONS, IMPLICATIONS, AND CONCLUSIONS

The purpose of this paper is to answer the research question, "What attributes promote the adoption and diffusion of computational tools for e-science?" Systematic analysis of in-depth interviews with 135 lead users, developers, and key facilitators reveal 10 themes. These themes describe the attributes that will increase the likelihood of wider adoption of computational tools for e-science, which will ultimately lead to diffusion beyond the incubators.

To reiterate, the 10 attributes are: driven by needs, organized access, trialability, observability, relative advantage, simplicity, compatibility, community-driven, well-documented, and adaptability. Along with defining and explaining these attributes, we present associated questions that stakeholders should ask on behalf of potential users, as well as diffusion strategies that will generate these attributes. The attributes and associated questions are summarized in Table 1.

Innovation Attributes	Critical Questions
Driven by Needs	"Does this tool meet my needs?"
Organized Access	"Is this tool easily available?" and "Can I find it at a known location?"
Trialability	"Can I try this tool without much investment of time?"
Observability	"What are my peers using, and how are they using it?"
Relative Advantage	"Is this option a lot better than what I have right now?"
Simplicity	"Is this tool easy to use?"
Compatibility	"Can I easily integrate this tool into my existing routine and collaborations?"
Community-Driven	"Is there an active community that will carry this tool forward?"
Well-Documented	"Is the tool well-documented with a track-record and user-guides?"
Adaptability	"Can I take this tool from that domain and bring it into my domain?"

Table 1. The 10 Attributes that Promote Adoption and Diffusion

Given the findings presented, it is evident that there are several implications for the e-science community. We argue that the 10 attributes can serve as a checklist for stakeholders as they design, develop, and promote the tools to establish a community of users. Stakeholders can ask themselves the critical questions and also utilize the proposed strategies throughout the development process. Now that a systematic analysis has been conducted to identify the traits that give strength to computational tools, it would be wise for future projects to strategically design their tools with the 10 attributes in mind. Additionally, these 10 attributes, critical questions, and diffusion strategies can help promote engagement of user communities.

6. ACKNOWLEDGMENTS

We thank Rion Dooley, Nancy Wilkins-Diehr, and John Towns for their support of this project. We also thank Brett Robertson, Ryo Nakagawara, and Tyler Bice for their early analysis of data for this paper. This study was funded by NSF ACI 1322305.

7. REFERENCES

- [1] Rogers, E. M. 2003. Diffusion of innovations (5th ed.). Free Press, NY.
- [2] Freeman, P. A. 2007. Is 'designing' cyberinfrastructure -- or, even, defining it -- possible? First Monday 12 (Jun. 2007), <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1900/1782>.
- [3] 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/.
- [4] NSF. 2009. From supercomputing to the TeraGrid Retrieved October 5, 2009, from http://www.nsf.gov/news/special_reports/cyber/fromscototg.jsp
- [5] Markoff, J. 1997. U.S. to support fewer supercomputer centers. The New York Times on the Web, from <http://partners.nytimes.com/library/cyber/week/033197super.html>
- [6] Kahin, B. and Jackson, S. J. 2007. Preface. First Monday, 12, (Jun. 2007), <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1865/1748>.
- [7] Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., . . . Wolff, S. (2009). A brief history of the Internet. ACM SIGCOMM Computer Communication Review, 39 (Nov. 2009), 22-31.
- [8] Comer, D. 1983. The computer science research network CSNET: A history and status report. Communications of the ACM, 26 (Oct. 1983), 747-753.
- [9] Jennings, D. M., Landweber, L. H., Fuchs, I. H., Farber, D. J., and Adrion, W. R. 1986. Computer networking for scientists. Science, 231 (Feb. 1986), 943-950.
- [10] Avery, P. 2007. Open science grid: Building and sustaining general cyberinfrastructure using a collaborative approach. First Monday, 12, (June 2007), <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1866/1749>.
- [11] Seidel, E., Muñoz, J., Meacham, S., and Whitson, C. A. 2009. A vision for cyberinfrastructure. Computer, 42 (Jan. 2009), 40.
- [12] NSF. 2012. National Science Foundation realignment plans, from http://nsf.gov/news/news_summ.jsp?cntn_id=125381
- [13] Edwards, P. N., Jackson, S. J., Bowker, G. C., and Williams, R. 2009. Introduction: An agenda for infrastructure studies. Journal of the Association for Information Systems, 10 (May 2009), 364-374.

- [14] Kee, K. F., and Browning, L. D. 2010. The dialectical tensions in the funding infrastructure of cyberinfrastructure. *Computer Supported Cooperative Work*, 19 (Aug. 2010), 283-308.
- [15] Stewart, C. 2007. Indiana University cyberinfrastructure newsletter. Retrieved November 1, 2008, from <http://racinfo.indiana.edu/newsletter/archives/2007-03.shtml>
- [16] Kee, K. F., Craddock, L., Blodgett, B., and Olwan, R. 2011. Cyberinfrastructure inside out: Definitions and influences shaping its emergence, development, and implementation In D. Araya, Y. Breindl & T. Houghton (Eds.), *Nexus: New intersections in Internet research*. (pp. 157-189). Peter Lang, NY.
- [17] Kee, K. F. in press. Adoption and diffusion. In C. Scott & L. Lewis (Eds.), *International encyclopedia of organizational communication*. Wiley-Blackwell, Hoboken, NJ.
- [18] Rice, R. E. 2009. Diffusion of innovations: Theoretical extensions. In R. Nabi & M. B. Oliver (Eds.), *Handbook of media effects*. Sage, Thousand Oaks, CA.
- [19] Bazeley, P., and Jackson, K. 2013. *Qualitative data analysis with NVivo*. Sage, Thousand Oaks, CA.
- [20] Corbin, J., and Strauss, A. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13 (Mar. 1990), 3-21.
- [21] Tarde, G. 1903. *The laws of imitation*. Henry Holt and Company, NY.
- [22] Owen, W. F. Interpretive themes in relational communication. *Quarterly Journal of Speech*, 70, 274-284.
- [23] McAfee, A. 2009. *Enterprise 2.0: New collaborative tools for your organization's toughest challenges*. Harvard Business Press, Boston, MA.
- [24] Shirky, C. 2009. *Here comes everybody: The power of organizing without organizations*. Penguin, NY.
- [25] Kuchner, M. J. 2012. *Marketing for scientists: how to shine in tough times*. Island Press, Washington D.C.