

What are the most urgent research questions to address about KI? Why?

For several years now, I've worked with colleagues to ask questions related to what is stalling the progress of the knowledge infrastructure (KI) expected to offer unfettered discovery, access, and reuse of research data. We started by studying data reuse and then investigated data production, management, sharing, and curation. More recently, we've used the results of these studies as a means to resolve issues at the point of data creation and management with mixed results. Based on these experiences, I believe that urgent questions to address about KI include:

- How can parts of KI that are opposing, independent, and lagging be bridged? Which bridges facilitate success under these different circumstances? When in the life of KI is bridging more or less successful?
- How can KI revive itself? What circumstances enable successful KI revivals?
- How does one part of KI take in and adapt what makes other parts of that same KI successful? What about other parts of different KI?
- What makes for successful transfer of KI scholarship into KI practice?

These questions can provide insight into not only how KI operates, but also how to proactively improve its operation.

Identify a KI whose survival is under threat. What led to these threats? Over what time frame? What actions or changes in circumstances might lead to its survival? What will be gained/lost, by whom, if this KI fails to survive?

The survival of the KI needed to offer unfettered discovery, access, and reuse of research data is under threat. The threat loomed at the start, because the KI was being built without taking reuse needs into consideration. Taking data reuse needs into account would have helped others (e.g., data producers, data curators, technology developers) understand what it means to build a data supply that meets demand, but accounting for reuse in isolation would not have been enough.

Survival of the KI also is threatened because due consideration has not been given to the interconnectedness among activities that occur over the life of the data. What happens at the point of data creation, documentation, and management can positively or negatively impact what happens at the point of data deposit, curation, and reuse. Furthermore, experiencing data deposit, working with curators, and reusing others' data can impact later data creation, documentation, and management. Less is known about these interplays, which means less can be done to effectively align these activities and the needs and resources associated with them.

Even with these threats, I don't think it is a matter of whether the KI will survive. It will. Current parts of the KI that support research labs, communities of practice, and disciplines are surviving, some are even thriving. A better question is whether the KI as imagined can be revived, so that the global community can fully participate.

Building relationships among the people working at key touchpoints over the life of the data might lead to the survival of this KI as imagined. The goal would be to use these relationships to develop mutual understanding and negotiate what can be done to support their work with the data in ways that positively impact others who also are working with the data. For instance, how can producers manage and document data in ways that facilitate personal use while also reducing the costs of curation and

increasing reuse satisfaction? What education and consultation services can curators design and scale to help producers manage data for effective personal use, deposit, and reuse? What experiences facilitate consumers' successful reuse, while also improving their data production and management practices? What can developers do to ensure their technology solutions are mutually reinforcing across key touchpoints over the life of the data? These questions can only begin to be successfully addressed if the people who participate in the life of the data work together to understand and support one another's needs.

If the KI as imagined does not survive, gains will be smaller in scale, perhaps held primarily among those already surviving. Failure of the more ambitious mission may lead surviving parts of the KI to turn inward and focus on their fitness. A positive outcome of this focus may be that they work to expand support for data reuse beyond their research communities, to include policy makers, K12 and higher education instructors, and the public. To date, few have pursued this goal. A negative outcome may be that they restrict participation in some way. For instance, they may require data contributions in exchange for data reuse and/or decide not to pursue data interoperability needs beyond their group.

How do KI spread information? Misinformation? Alone and in conjunction with other infrastructures?

KI spreads information and misinformation through the people, content, and technology that comprise it. Moreover both kinds of information can be disseminated in conjunction with other infrastructures. However, I expect that misinformation is more likely to spread via infrastructure that is more homogeneous, particularly when it comes to the representation of the people that comprise it. It is easier to bake similarity into homogeneous infrastructure, which means it would be more difficult to counter the messages being developed and spread. At worst the result is misinformation, at best it is partial information, because the points of view used to generate the information are limited in number and kind. I expect the opposite to be true for infrastructure that is more heterogeneous, because it is more fragmented. The fragmentation provides the potential to create and spread information that is more representative of the people who are impacted by it.