Panel Article **Open Access**

Angela P. Murillo*, Renata G. Curty, Wei Jeng, Daqing He

Confronting the Challenges of Computational and Social Perspectives of the Data Continuum

https://doi.org/10.2478/dim-2020-0008
received March 24, 2020; accepted April 28, 2020.

Abstract: As the availability of data is increasing everyday, the need to reflect on how to make these data meaningful and impactful becomes vital. Current data paradigms have provided data life cycles that often focus on data acumen and data stewardship approaches. In an effort to examine the convergence, tensions, and harmonies of these two approaches, a group of researchers participated in an interactive panel session at the Association of Information Science and Technology Annual meeting in 2019. The panel presenters described their various research activities in which they confront the challenges of the computational and social perspectives of the data continuum. This paper provides a summary of this interactive panel.

Keywords: data acumen, data stewardship, agricultural data, biomedical data, archaeological data

1 Background

The current data paradigm needs to be examined to consider potential points of integration for a more comprehensive approach when working with data to make sense of the data. In the current literature, a variety of data life cycles exist to provide a framework on how to confront the data challenges (Cox & Tam, 2018). These data life cycles have often been framed by two approaches, data acumen and data stewardship.

***Corresponding author: Angela P. Murillo,** School of Informatics and Computing, Indiana University-Purdue University Indianapolis, Indianapolis, Indiana, USA, Email, apmurill@iu.edu
Renata G. Curty, Data Curation Program — UCSB Library, University of California, Santa Barbara, USA
Wei Jeng, Department of Library and Information Science, National Taiwan University, Taipei, Taiwan, China
Daqing He, School of Computing and Information, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

The concept of acumen can be defined as the capacity to make strategic decisions. Data acumen, therefore, refers to one's ability to create insights from data and apply them strategically. As described in the National Academies Press report on Data Science for Undergraduates (2018), data acumen experts will use tools to understand data, make good judgments, and make correct decisions with data, and use data analysis tools responsibly and effectively. This analysis and interpretation of data often rely on specific computational skills and strategies such as visualization, modeling, and machine learning techniques. One of the first data life cycles to describe this perspective was the OSEMN framework. The OSEMN framework described each step of a data science project, and it included obtaining, scrubbing, exploring, modeling, and interpreting the data (Lau, 2019). In short, the primary emphasis of the data acumen approach is to work with data to reach informed decisions and insights based on the data.

The concept of stewardship can be defined as supervising or caring for an item or a person. For example, a steward oversees and protects. Data stewardship can, therefore, be referred to as the management and curation of data. In many ways, data stewardship aligns closely with the research data management activities that focuses on the management of data among and between different data lifecycle stages. Digital Curation Centre's Data Lifecycle Model was one of the first data curation life cycles that examined stewardship practices (DCC, 2020.; Higgins, 2008). The DCC model describes each step of curation and includes the following activities: create, appraise, ingest, preservation action, store, access, use and reuse, and transform. The primary emphasis of the data stewardship approach is the management of data that includes the skills of data provenance, organization, curation, and preservation actions. Additionally, data stewardship is cognizant of social considerations such as data policy, frictions, regulations, openness, ethics, and privacy.

As data impacts academia, the workforce, and the society at large, understanding the relationships

between the various data approaches is vital. Reflecting on the convergence, tensions, and harmonies in the approaches when working with data may assist in creating an environment that allows for data to become more meaningful and impactful so that all actors involved can reap the benefits of our ever-increasing data-focused world.

At the Association of Information Science and Technology Annual Meeting held in 2019 in Melbourne, Australia, four researchers organized an interactive panel named "The Convergence of Computational and Social Approaches for Unveiling Meaningful and Valuable Data." The goal of the panel was to initiate a discussion between the experts of data acumen and data stewardship to examine the tensions and potential points of collaboration and harmony between the two approaches. The purpose of the panel was to consider how these approaches are intersecting in research, as well as how this convergence could potentially lead to a more comprehensive approach and, ultimately, add more meaning and value to data.

The panel consisted of three parts. The first part included a brief introduction to the panel topic and overall structure by the panel facilitator, Dr Daqing He, from the University of Pittsburgh. The second part included three presentations, one from each panelist, Dr Angela Murillo from Indiana University, Indianapolis, Dr Wei Jeng from the National Taiwan University, and Dr Renata Curty, formerly from the State University of Londrina. Part three included a question and answer session, which also allowed some discussions on related topics.

The key takeaways from the presentations and discussion include:

- The convergence, or at the least the reflection on the convergence, of different data approaches is timely as it impacts nearly all sectors and disciplines.
- While these approaches may lead to tension, they also should be complementary.
- The need for machine-actionable pathways to integrate data stewardship and acumen to minimize potential frictions.

2 Panel Introduction

Daqing He (Facilitator): "The Convergence of Computational and Social Perspectives for Unveiling Meaningful Data"

Dr. He's presentation started with a brief welcome note to the audience, introduction to the panel and structure, and an overview of the topic.

As described by He, the key purpose of the panel presentations was to raise a discussion about the convergence of computational and social approaches when researchers were working with data. In particular, each panelist shared their experience and examined how these two approaches shape or interact in their research activities that emblematized the topic. He noted the importance of understanding, engaging with, and, the need for both social and computational approaches for working with the data, given that data knowledge is needed across all disciplines.

Dr. He then introduced the definitions of two approaches, data acumen and data stewardship. While data acumen focuses on computational skills such as gathering, extracting, and modeling, data acumen takes on a more computational approach and focuses on the ability to draw more informed and better decision making based on data. Data stewardship focuses on the management of data across different stages of the data life cycle and focuses more on social and management skills such as policy, behavior, ethics, and privacy. Data stewardship addresses questions such as how to curate, store, preserve, and use data.

The concept of the data life cycle was also introduced to describe the current or present way the research community has understood data needs. He provided two data lifecycle models, the Harvard Medical School (HMS) Data Lifecycle[1] and the Digital Curation Centre (DCC)[2] Data Lifecycle, created by the research community as a way to understand data workflows.

Dr. He ended his presentation by suggesting that both practitioners and academics should consider a more integrated and comprehensive framework to deal with data issues and challenges.

3 Panel Presentations

3.1 Angela P. Murillo "Computational and Social Research in Agriculture"

Murillo's presentation discussed her preliminary research, which was a joint effort between her and a colleague in data science that stemmed from their mutual interest in using their research to improve the impact on earth sustainability. An industry partner had approached them and asked for a proposal to consider how to examine

1 https://datamanagement.hms.harvard.edu/hms-data-lifecycle
2 https://www.dcc.ac.uk/guidance/curation-lifecycle-model/

agricultural data throughout its entire life cycle, as well as recent changes in farming technology that helps to enhance data stewardship and data acumen activities.

Her proposed research examined how agricultural data are being collected, new technologies that have been created to assist farmers to gather field data, the various data stewardship activities being conducted by farmers, and how computational or data acumen practices can assist farmers in data-driven decision making.

Regarding data stewardship activities, the research team learned that farmers are using mobile tools to gather, organize, and manage data from both field and farm. During stewardship activities, lot of information are collected and managed, including seed types, treatment types, weather and soil conditions, irrigation information, geolocation information, as well as drone images.

With respect to acumen activities, the researchers learned that there is an interest in ensuring appropriate data integration and transformations so that multiple data sets can be combined. Potential acumen activities include larger-scale return on investment studies related to the farm, the field, seed types, and field treatments as well as to look at yield reporting for different treatment types.

While there are many positive aspects to these efforts to consider agricultural data more comprehensively from both the data stewardship and data acumen approach, there are potential tensions such as farmers having to attend training sessions to learn and invest in a new technology. From a positive perspective, if these data are gathered year after year, long-term studies can be conducted on seed yields and the impact of various treatments, as well as on larger geographic areas.

Another tension to consider is that in collaborative research there could be focusing issues and constraints. Researchers who are more focused on social aspects of data may only consider factors such as user needs, designated community, and design interface. From the computational approaches, researchers may be more interested in the amount of data, type of data, and potential computational approaches. While these differences of focus can cause friction between both sides, it is important to note that these boundaries can and do merge, though the roles can be ambiguous. This ambiguity can occur if the same people are performing both acumen and stewardship roles.

Murillo ended her presentation with some questions to explore the convergence of data acumen and data stewardship in agriculture.

- As a data steward expert, what recommendations would you provide the farmers to ensure appropriate and data steward practices in regards to their farm work, such as sharing and reuse of these data, and long-term considerations?
- As a data acumen expert, what additional data would you consider pertinent to analysis? What suggestions do you have for the incorporation of computational methods to assist farmers, such as visualization and predictive modeling?

3.2 Wei Jeng: "Preserving the Research Workflow in Biology and Medicine"

Jeng's presentation explored a preliminary study of bio-medicine research practices. She presented early observations from the data of seventeen participants in Taiwan that focused on the problems between data acumen and stewardship.

Jeng described how researchers working in this area considered issues such as data openness and reuse, FAIR data, reproducibility, and data reuse. On a higher level, these researchers were looking for a better solution to support the digital scholarship for their research infrastructure, especially the data-related activities around the research life cycle that involved data acumen and data stewardship.

Jeng described online information and tools such as the data curation profile (DCP)[3] from UIUC and the UK's DCC for researchers' data curation practices. However, the academics in these resource centers focus on practices in Western society. And as she is working in Taiwan, scholars still need to examine researchers' data activities in a local context. For example, researchers in Taiwan may not solely depend on online information or the established DCP. The project's goal was to gather information from a local context to understand researchers' reproducibility practices, as well as challenges/obstacles of their data sharing.

The project targets Principal Investigators (PIs), the lead researchers for grant projects, in the biomedicine fields. The biomedicine field has very complicated and diverse data, including sequence data, lab test results. It includes doctors' handwriting and summary with diagnosis and medical images and other unstructured data, as well as Perl scripts and Python codes to gather DNA sequence data.

The project team conducted a design-thinking workshop, which is a common approach used in the Human–Computer Interaction (HCI) community to

3 http://datacurationprofiles.org/

stimulate or motivate people to co-define and cocreate a better solution to their problems. The project team consisted of seventeen PIs from the biomedicine field in Taiwan, funded by the Taiwan government. They were divided into five groups of three to four people per group. The workshop was about 3-hours-long, on average.

The design-thinking workshop followed the Double Diamond Framework, which is a design process developed by the British Design Council (2015) around fifteen years ago. It comprises four phases: *discover*, *define*, *develop*, and *deliver*. The first and the third phases are considered as "divergent" sessions that can help participants discover problems and develop more possible solutions. The second and the fourth phases are "convergent" phases that gather consensus from the *discover* and *develop* phases.

In particular, the participants first defined the problems raised in their given scenario, so that their thoughts could be stimulated and the challenges could be defined. The participants chose the most significant problem they would like to work on for the rest of the workshop; they will later try to come up with some possible divergent solutions and incentives. Then, participants worked through each phase of the Double Diamond Framework.

First, in the *discover* phase, our participants discussed the problems of their data sharing and reproducibility process. To achieve this, the DCP was used as a persona. The participants read those personas, thought of their situation, and listed out their challenges. In the second phase, define, the participants put aside the discussion regarding problems and challenges from the DCP persona. They were asked to come up with their challenges, and at the same time, listened to their peers in the same group, which helped the research team discover more challenges. At the end of the *define* phase, the participants voted for the most important challenge. In the third phase, *develop*, the research team encouraged people to give different answers or solutions to the challenge. Participants were asked to discuss the tasks toward data sharing, the challenges, difficulties, the benefits, and positive things associated with those tasks. The research team provided solution cards that provided pre-existing solutions, and the participants were also provided empty cards so that they can write down their solutions. In the last phase, *deliver*, participants discussed and shared their ideas and solutions and each group agreed to a relatively better solution.

Jeng provided observations of two scenarios that occurred during the workshop. In the first scenario she discussed how that PIs in the bio-medicine field stated and faced the difficulties in documenting each and every decision or the rationales systematically in a competitive lab culture. She also discussed about the competitive nature of lab culture that existed among the group members and informed that in one lab's group meeting, each one tried to pixelate the critical results because they did not want others in the same research group to have the information. Because of this competitive culture, the lab groups will try to conceal the data until the results are published. This is something Jeng found surprising and concluded that the data acumen is dominating the whole culture of the lab as they try to memorize or note down some really important decisions or data. They might never "go digital" because they are afraid that important information will get leaked out.

In the second scenario, the importance of data stewardship was explored. The PIs discussed how they had two paths, Path 1 and Path 2, that could be described as a "fork in the road." However, as they had poor documentation, they could not go back to the fork and make a different decision. Their main point was that poor data documentation impacted their decision-making ability.

Jeng ended her presentation with several questions to help explore the tension between data acumen and data stewardship.

- How can we document everything in a highly competitive lab culture?
- How can we freely access or even travel back and forth on every decision that we have made?

3.3 Renata G. Curty: "The Role of Paradata in Digital Archaeology"

Curty discussed data acumen and stewardship, tensions, and convergence from the perspective of digital archaeology and her ongoing study is at the very early stages. The main question of her talk was, "how can we bridge social and computational approaches to improve transparency and reliability in digital archaeology?"

Curty presented the various types of data in archaeology, which included tangible and intangible data that were important throughout the processes of data collection, data documentation, and sharing. Curty then introduced the idea of paradata that emerged in the late 1990s with the increase of web surveys and the data generated automatically by self-administered surveys from respondents interacting with survey instruments and is not often discussed in LIS field.

Curty explained how discussions of paradata have expanded to other fields and disciplines. But, she

emphasized that there are different perspectives of what paradata entails and some confusion remained with contextual metadata. Paradata is not reduced to provenance, but it also accounts for the reason behind the whole interpretation process. In the case of archaeology, the interpretation process starts as soon as data is being collected in the field. So, it accounts for all the process and the reasoning behind some of the data gathering and data collection that happens in the field work.

A definition of paradata we can apply to the field of archaeology is the one provided by the London Charter (2009): "Data about human processes through which artifacts or collections were processed or interpreted." This allows for the idea of interuse of data, which differs from many data life cycles that usually have use and reuse at the very end of the process. Since in archaeological research, in many cases, artifacts that tell a story and a relationship among them are grouped together, that can be established with data gathering; so paradata plays an important role throughout the whole data life cycle. There are situations where the descriptions of the data start within the structuring of the dataset, and these are used as evidence to interpret an artifact or comment on methodological premises in a research publication.

Curty also presented the two dimensions of paradata. The first, the intrinsic paradata, includes the architecture of the model in relation to digital archaeology and the modeling process, including choices that were made about instruments, tools, applications, and how they were used. The second, extrinsic paradata, provides the description of critical interpretation of sources used by the researchers and presents a certain degree of reliability and objectivity about the process, including geometry, location and position, date and age, and so forth.

Regarding data stewardship, there are different tools that are used by archaeology and other researchers in other fields. There is a chance to empower reflexive thinking to the field of archaeology, by bringing more perspectives to the table, to minimize the gap between data collection interpretation, and also to ensure that paradata is better documented. The reasoning process can be registered and annotated on-site using tools such as Evernote and Jupyter Notebooks, whereas the digital part of archaeology involves the data acumen perspective. For the computational part of archaeology, there are different existing tools for photogrammetry such as COLMAP, MicMac, and RealityCapture. The tension between data stewardship and acumen in archaeology resides, in part, in the translation of cultural heritage with transparency and incorporating context, in a fairly comprehensive way, to the digital world. Along these lines, another question we can further discuss is: how to translate cultural heritage into the virtual world in a more accurate and transparent manner while preserving its contextual richness?

Curty presented Apollonio and Giovannini's (2015) study that discusses cultural heritage and how to integrate context in 3D modeling in a more transparent way. Based on the assumption that paradata is crucial to ensure scientific transparency of any virtual archaeology project, the authors argue that as metadata, the paradata should be complete, clear, concise, and easily available, as well as standardized. Not only does paradata allow researchers to track the entire modeling process, but also it is a valuable resource to allow better understanding and interpretation of data objects.

Curty described the tensions between the data documentation and the computational translation to digital artifacts. The first tension is a common sentiment that subjectivity should be viewed as an enemy of archaeologists, but in reality one needs to understand that subjectivity is ingrained in the research process and difficult to avoid in the social sciences. There is also some discussion about how some archaeologists have been overly excited about new technologies, and fetishizing its capabilities in search for a more appealing way of doing archaeology, which is directly related to the data acumen approach. But there should be a good combination of both worlds to allow that these digital representations will incorporate the contextual aspects behind them. Third, for digital archaeology, a person or a group of people will work or use the computer, and therefore subjectivity is involved in this process too. So it is important to discuss ways to make digital archaeology more transparent, reproducible, and accurate. It should not be a matter of disregarding subjectivity, but to document fully all the reasoning behind interpretations and well-documented paradata may be one step forward to help in that process.

Curtly concluded with how she is currently examining repositories in archaeology and also data journals, checking how policies cover the issue of paradata, and how they recommend archaeologists to document the reasoning behind some decisions during the data collection stage, along with the contextual metadata and paradata.

4 Questions and Discussions

Question: *This is more of a comment than a question. I've done a lot of work and research within archaeology. The first thing I want to say is that tangible and intangible are important, far more important in an archaeology context.*

I think that this brings me to my next point, objectivity doesn't exist. What I'm saying is much more complex and I am making your job more difficult, because it is about recognizing that there is not one correct record, there is not one current system that is going to work for everybody in the same field, let alone for everyone interested in the area of archaeology. There is a need to recognize the complexity. For example different archaeologists, all working on the site, they do things differently, not because they don't understand the system, not because they are being difficult, because the systems we have created do not reflect the complexity they are working with. So we should not say that in terms of objectivity or subjectivity, we need this in terms of the intersubjective expertise as we have to recognize the people working in the field of the experts. What I am suggesting is the concept of intersubjectivity is a much more useful theoretical lens through which to engage with what you are interested in than the Cartesian objectivism vs. subjectivism subject.

Answer: The exact point I want to suggest is, like the way people work in information science, researchers in the field have to use data-oriented activities to help their work. But we cannot assume that they understand all the things about data, and therefore, we just tell them, "here is a tool, use it!" We have come to conceptualize all the tools and all the activities, and this is the tension that we discussed, that we need to find a way to engage what we call the social aspect and also the actual computational aspect of it.

Question: *The three speakers gave very interesting presentations. My son is an archaeologist, and his party just came back from an excavation in Egypt, and the ministry of antiquities, where he was creating digital records of artifacts, told that digital records had no legal and official validity and that he had to document physically about antiquities. Is that something about the conservative nature of archaeologists? And how does that affect the information at all?*

Answer: There is a full section in my paper about drawings for field archaeologists as a part of information. So I am completely unsurprised. I can give a long answer if needed. One conflict is the risk, but it is not conservatism. They argue that drawing provides different information from the digital. But I know they never say that the digital does not have legal validity. But they do say that digital does not replace drawing. You have to have both.

Answer: My work with anthropologists has been similar. Their drawings and the use of drawings are fairly well-documented in Italian literature and cultural heritage. Sometimes they just think that the drawings will have more details about the artifact. The simple answer is it allows them to draw the details. I think it is important, whereas the digital tools include a lot that is unimportant.

Question: *Is there a generational issue there as well? In terms of not all people like me, but young millennials who use technology? Because what they were saying that we are doing is taking photographs and tracing the picture, the younger archaeologist were pushing back against it. So actually, you know, this is the problem, the physical is something which is the authority. So just a comment, not a question. But thank you to all the presenters.*

Answer: Drawing has a way of capturing more details and observations because the drawer is forced to examine the object in more detail. It is similar in geology, but the use of the drawing provides a full record, and, of course, there is also the use of photography in geology.

Answer: So it is interesting that you say that, because I was quite convinced having started my study with an attitude of why in the twenty-first century are they drawing? My theory was that the documentary arguments were really exploitation to some extent. My thought was the act of drawing itself fosters an understanding of what they are working on.

Question: *I work with medicine and surgeons. So I talked to surgeons, I hear constantly that it is just too inconvenient and too time-consuming for data documentation, and there is a culture of time pressure. You can always find a medical emergency, which they would much rather be doing. Additionally, there is software and hardware for them to use, but it is not the best at capturing. Could you comment on this?*

Answer: It is really hard to balance. In the chemistry lab, we can see the same thing. The chemists that I talked to at their lab were working with really toxic liquid. So, eventually, I really wanted to recommend that they can document some protocols using the iPad or something. And when I got into the lab, I realized it was impossible. They were wearing gloves, so it's really hard for them to use an iPad. So it is really hard to ask them to document things.

Answer: In a previous study, I observed the workflow data from a Viral Vector Core lab. I found that they were so busy that they ended up putting in metadata after the completion of work. Everything would happen there. But there, they would end up documenting all of the steps that happened throughout the day for the data workflow. And they realized they were missing vital information. One solution was to capture throughout the entire process, which was a long process. When they came to us, they knew they were missing metadata throughout the process, and they were losing financially because someone would have to go back and put the metadata in, and often they

were the research scientists. And so, we found ways to automate the process and look for particular points in the workflow that were not captured properly.

Answer: The data creation processes themselves are a social process. When you start to interrogate or question, so exactly where does the data or metadata come from anyway? You realize that those human processes are behind it. For example, in real life, websites are capturing everything that people do. In the case of a citizen science project, when you start to probe, you realize, well, in fact, actually people record things differently from what we would think for the same activity. And so the data are not objectively capturing participants' experiences. Sometimes this comes out of the interaction between the system that we created and participants. So it really does force you to look at the data.

Answer: Also, my investigation proposes and concludes that the process of information creates a social interactive process. Information practitioners consult information organizations in order to implement information systems when we see a greater chance to record information in that workflow. Our information-process and data-process creation are more social and probably has the greatest degree of complexity than we probably think when we design information systems that actually contemplate the nature of the information process.

Question: *In the presentation about farmers, I was thinking about information sharing among the farmers does not create an economic disincentive because the knowledge shared does not create an economic difference. I actually put my weather station data into a relational database, but the weather as a community aggregation of data actually gives a much more fine-grained notion of weather then they are capable of maintaining themselves. Because weather patterns can create microclimates that have impacts on individual fields.*

Answer: The thing that was really interesting about how they are gathering data by field. Weather conditions could be included, just like drawing images so we can get really detailed images and data, and you end up capturing more information that could be useful.

Question: *At the City College of New York, one of the things that we found was that they were very interested in the quality of the data and/or what kind of state it was in was a computational factor. But there are also social considerations in terms of what they would share or not share with other people within the project and beyond the project. So I think this is another example of where this convergence will be at the end of this data life cycle. In terms of what now happens to the data. Do we make it open? Do we keep it closed? What rules are there around that with funded research? And so it is sort of a challenge.*

Answer: Your question made me recall something from the workshop. The PIs said that their research assistant usually refused to do data documentation because they think they have something more important to do than documentation. They hired a research assistant with analytic skills in the biomedicine domain. They really do not know how to do the data management or the curation stuff. So it is a real problem that they recognized the data stewardship in their project, but at the same time, they also confirm that they couldn't hire a person or an expert to help them with this.

Question: *I would like to hear if you have any thoughts about the distinction between data acumen and stewardship, and about these two different scenarios and how we can work together.*

Answer: A lot of teams did not document the data until during collection, and a lot of people work together because they wanted to get the work finished before anybody else did. After the completion of work, they will go back and document rather than documenting it at the time they collect. Of course, this can reduce the quality of the data and cause problems.

Answer: I think I agree with you. However, information professionals have unreasonable expectations around the role of systems in real conditions. For example, expecting iPad to work in Scotland, where it is raining all the time.

Answer: I think it is important for organizations to employ or assign the job to someone. People see information management is important and expect certain high-reliability documentation. In other cases, it is probably not. It is a matter of what gets rewarded.

Answer: There is a huge amount of data being collected in a small business, and we do not have the resources to employ or assign one person for the job. It is necessary to get the data organized. And it is still our job to find ways to make that information retrievable. I would compare that perspective with every lab manager at universities.

Question: *What would you say are the biggest challenges to combine both approaches in research projects? What would be possible solutions to overcome such challenges? We would like to hear if you have suggestions based on past experiences and also what would be suggestions for us to build a stronger community bringing expertise and these different views and also how we could make cases or how we could take advantage of this conference, not taking advantage in the bad way, but how we could use this as a base for discussion and for prolific discussion in social and computational approaches.*

Answer: Well, you make a critical point. If you think about a scientific paper as being a valid assertion, the authorization of legal validity is basically the problem for research data. And so we do not credit it.

Answer: One experience I observed is that medical and life sciences are different domains, and the behaviors are very different. For example, some domains are very open and have a lot of data standards, some others not so much. I have also noticed most often the motivation comes from within, the sharing of resources and data.

Question: *Some of you mentioned that one of the challenges is that this is just time-consuming. Then, we had some other people talking about issues related to the data credibility. What would be potential solutions for these issues?*

Answer: It became very clear to me that the idea that you will just make something with quick (answer) boxes and really fast does not work on because it does not fit in with the multiple and various interests that researchers may have. So, while they want it easy, they also want to be specific in the right kind of ways. So it needs to either be much customized or have lots of options in that way. It is in that specificity that researchers can see the point. If it is a general system, it might be easy for me [as the creator of the system], but not useful to them.

Answer: I think for us in the information science community, the constant conversation with the researchers is essential.

Answer: This makes me wonder about the potential for newer interface modalities, like voice recognition and watch or video recorder in the work site.

Answer: That's my perspective, sometimes on the business side of things, and we have proposed to use automated agents so that people do not have to fill three hundred fields of metadata.

Answer: I organized a conference of artificial intelligence for data discovery. For this purpose, I tried to see if there are any automation methods I can use, for example, to store metadata. But, so far, there is no consensus and that's why we are discussing here. Still, you have to do a lot of hard work in advance.

References

Apollonio, F. I., & Giovannini, E. C. (2015). A paradata documentation methodology for the uncertainty visualization in digital reconstruction of CH artifacts. *SCIentific RESearch and Information Technology Ricerca Scientifica e Tecnologie dell'Informazione, 5*(1), 1–24.

Council, D. (2015). *What is the framework for innovation? Design councils' evolved double diamond* [Internet]. London: Design Council; Retrieved from https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond

Cox, A. M., & Tam, W. W. T. (2018). A critical analysis of lifecycle models of the research process and research data management. *Aslib Journal of Information Management*, *70*(2), 142–157.

DCC Curation Lifecycle Model. (2020). Retrieved April 28, 2020, from http://www.dcc.ac.uk/guidance/curation-lifecycle-model/

Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1), 134–140.

Lau, C. H. (2019, January 10). *5 steps of a data science project lifecycle*. Toward data science. Retrieved April 28, 2020, from https://towardsdatascience.com/5-steps-of-a-data-science-project-lifecycle-26c50372b492

London Charter. (2009). *Glossary. London Charter for the computer-based visualization of cultural heritage.* Retrieved from http://www.londoncharter.org/glossary.html

National Academies of Sciences, Engineering, and Medicine (U.S.) (2018). *Data science for undergraduates: Opportunities and options*. National Academies Press: eBook Collection. https://doi.org/10.17226/25104