Journal of Structural Geology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Structural Geology

journal homepage: www.elsevier.com/locate/jsg

Collaboration, cyberinfrastructure, and cognitive science: The role of databases and dataguides in 21st century structural geology

Thomas F. Shipley[a,*], Basil Tikoff[b]

[a] Department of Psychology, Temple University, 1701 N 13th Street, Philadelphia, PA, 19122, USA
[b] Department of Geology, University of Wisconsin, Madison, 1215 W Dayton St, Madison, WI, 53706, USA

ARTICLE INFO

Keywords:
Spatial cognition
Cyberinfrastructure
Expert training

ABSTRACT

Structural geologists support their mind with tools, and these tools are increasingly computer based. The advent of Intelligent Systems will allow creation of research teams that combine the strengths of the human mind and computer processing to produce new research results. The efficacy of these approaches will require a solid grounding in cognitive science. Critical to this approach are databases, which are potentially transformative solely in their ability to allow access to data, in a primary form. Emerging more recently, however, is the concept of a dataguide, in which computer-aided analysis informs ongoing decisions about where and what data to collect. The creation of human and computer teams can expand the types of questions that can be addressed in structural geology and tectonics research, but it will take a community-based effort to understand the value of data to experts and how computers might aid an expert in the field.

1. Introduction

The pace of technological advances influences nearly every aspect of our lives, including the professional aspects of being a structural geologist. The transformation in the last 50 years has been profound: There were virtually no computer skills necessary in the 1960s to operate as a professional geologist. Now, it is difficult to imagine working without drafting programs, digital stereonet programs, web search engines, visualization platforms such as Google Earth, and Geographical Information Systems. What will the future of geology look like as technology advances our ability to collect and assimilate data? The rate of data collection is certain to increase with advances in instrumentation, such as using a cell phone to measure strike and dip (see Whitmeyer et al. (this volume) for current limitations) or a mobile agent that can autonomously carry an instrument package to collect data at new locations (Qian et al., 2017). Future data will primarily be digital and thus require digital systems to store, search, analyze, and share.

Digital databases are now essential for every field in the sciences - here we the term database, synonymously with data system, to include both storage and access. Structural geology – in addition to many other field-based research areas that do not collect instrumental data – has been slow to adapt to digital databases. To not utilize databases – as an individual or a community - is to invite scientists to ignore research that is not in a database (Chan et al., 2016). The more positive way to view the situation is that cyberinfrastructure can both increase the quality of what we already do, as well as facilitate new types of analyses and approaches that we have not imagined. The appropriate technologies (e.g., data systems, graph databases, digital field instruments and aligned metadata) are now available and efforts – such as StraboSpot (https://www.strabospot.org/) – are attempting to support this effort for the community.

The expert geologist has always supported their mind with tools. A Brunton geologic compass measures slopes more precisely than the human eye, and maps hold more data than can be held in mind at a given moment. Here we consider the transformations in practice and the new ways to support the expert mind that are afforded by digital data. First, we discuss current data use and data collection strategies. Second, we consider the relative strengths and weaknesses of the human and computer reasoning. Third, we combine these to consider how practice could change with access to data collected by more than one person. Fourth, we discuss how practice could change if computational tools were developed to augment the limitations of the human mind, and offer some suggestions on key areas for future research. We conclude by considering the nature of these new human – computer teams, how to support such teams, and highlight some key questions to focus on opportunities and obstacles to adoption of community-data based science.

The paper is intended to serve two readers, the structural geologist interested in thinking about how they can make best use of new classes

* Corresponding author.
E-mail addresses: tshipley@temple.edu (T.F. Shipley), basil@geology.wisc.edu (B. Tikoff).

https://doi.org/10.1016/j.jsg.2018.05.007
Received 26 December 2017; Received in revised form 12 March 2018; Accepted 10 May 2018

of tools to transform new field and laboratory practices - to free themselves from looking for digital replacements to field notebooks and think about the new intellectual opportunities offered by data systems. We also explicitly aim to recruit the members of the structural geology community who will take the steps needed to forge the interdisciplinary links with computer science to build the types of tools that can transform field science.

2. Collaboration

Many geologist have recognized the myriad opportunities available in digital data. Some have addressed the need for a new tool (e.g., Stereonets; Allmendinger et al., 2017), and others have worked to develop the capacity for geologists and computer scientists to work in interdisciplinary teams to develop new tools (e.g., Mookerjee et al., 2015a; b). For the most part, the perspective of these efforts is on building a tool that will allow an expert to advance their individual research program. Here we take a different perspective considering what are the limitations of the human mind in practicing science and how advances in digital data can create new research teams that combine the strengths of the human mind and computer processing to produce transformative research.

There are opportunities inherent in any scientific collaboration (Galison, 1997). The field of tectonics now commonly involves teams of experts from different disciplines (e.g., structural geology, geophysics, sedimentology, and geochronology) that come to understandings that were not possible when each discipline considered the problem in isolation. Such teams are constituted with knowledge overlap such that members agree on common goals and what counts as an answer, and complementary skills and knowledge to allow new approaches that would not occur from more traditional approaches. Similarly, designing an intelligent system to support a human-computer team requires understanding the strengths and weaknesses of the human. By understanding human strengths, we may avoid spending time trying to solve hard Computer Science problems that humans are proficient at and devote efforts to solving problems that humans are not-so-proficient at. Furthermore, understanding how humans reason about and practice science is necessary to design a science partner that can coordinate their efforts with the human in such a way as to minimize cognitive load on the human. New designs will require programs of work to address a range of fundamental cognitive science and computer science questions. Below we consider the relative strengths of human mind and computer algorithms, and how expert's work is influenced by initial hypotheses.

2.1. Modes of mapping

When an expert goes into a new field location, they are not a *tabula rasa*. They have initial hypotheses that come from some collection of prior information that might include published papers, aerial photos, other people's geological maps and narratives, and an understanding of the regional geological history.

Research has confirmed the importance of *a priori* information in pre-field planning and in-field decision making. Baker et al. (2016) studied expert and novice data collection in an unfamiliar field area when individuals were tasked with a goal of developing a geological map of the area. No geological interpretation for the area was provided. Although this approach would be unusual practice for an area where previous work had been done, this set-up allowed the researchers to study practice in the absence of a mental model developed by others, thus simulating a novel field area. Participants were shown examples of the rock types that were present and were provided with an aerial photo and topographic map of the field area. Expert's paths, reports, and constructed geologic maps all revealed the use of the aerial photo to establish an initial hypothesis. In contrast to novices, experts were more likely to ultimately develop a correct geological map if they had an initial hypothesis (even if the initial hypothesis of the structure in the area was incorrect), and expert's paths through the area were more efficient than novices, tending towards the areas that provided the highest quality information for discerning among possible interpretations.

How to take advantage of the various sources of information from the past that are available before entering the field is a critical part of the apprenticeship in a field science. Each mentor may vary in their explicit training on how initial models, or hypotheses, may guide data collection (Gilbert, 1886; Chamberlin, 1890). We have observed at least four distinct data collection approaches.

Reconnaissance mode is necessary when one is new to an area, and broad divisions of the geology are necessary. This is employed when the data collector does not know much about an area - still widely applicable in places like Alaska – and is akin to using a satellite photo to get a sense of large scale structures and the geomorphological patterns. In the field, this activity is typically done by finding, and climbing up to, the highest point or otherwise best views of the field area. From this vantage point one may develop a mental model of the area by visually estimating large scale features and preparing for navigation within the area. One important function of reconnaissance mode is data collection planning: Where do rocks outcrop, where are the good, or dangerous routes among stops, which outcrops might be most revealing about larger scale structures?

Sampling mode is employed when your goal is to get a specific specimen and just enough context to use it. This mode is now widely used in geochemistry, for example. One drawback to operating exclusively in this mode is that results are often misinterpreted because the context was not understood. The best use of this approach is typically done when a specialist in sample analysis works with a geologist who is well-acquainted with the field area to provide the context from previous larger scale work.

Mapping mode is survey-/field-camp-/quadrangle-style mapping. The mode involves making stations that contain a limited number of measurements and identifying contacts between units. The goal of covering as large an area as possible in a limited time results in a necessary tension between the number of measurements and the spacing between measurements – and a tendency to minimize the time at each station, and thus likely the details recorded. Generally, the locations of units between stations is interpolated from trends at outcrops and larger map patterns. A particularly valuable application of this mode is the construction of outcrop maps, which do all of the above and also display the extent of the outcrops, to preserve the data that was used for filling-in and aid future researchers finding key outcrops. Because of its pedagogical value (Whitmeyer et al., 2009), most geologists are taught basic mapping mode.

Problem-solving mode is collecting data guided by one or more mental models. This approach is most clearly aligned with the formalism of a multiple working hypotheses approach to science, where data is collected to discern among different potential models (Gilbert, 1886; Chamberlin, 1890). Most academic structural geologists work in this mode. In this case, the boundary of the field area is defined by the problem to be solved. There is significant variability about how to work in this mode, because both the problems and the people who work on them are very different from each other. Within problem-solving mode, there are discernible approaches. One common goal is to collect as many high-quality observations as possible in a field day. Different approaches emphasize in-depth querying of each stop vs. higher density observations vs. more observations over a larger area. The variations within this spectrum result from the personal preference of the practitioner and/or the nature of the question to be addressed.

2.1.1. Empirically based vs. theoretically motivated data collection

The modes described above represent approaches to disciplinary aims that can be located along a continuum from data-driven data collection to theory-driven data collection. For example, mapping mode could be solely empirically based data collection. In this approach, one does not know what he or she are going to find, and the geologist will work out a model (or adopt one from the literature) once they have an idea about what the rocks "are saying". Conversely theoretically motivated data collection generally requires working in problem-solving mode. This approach typically requires that the practitioner think they are in a good place to test a conceptual model. From a cognitive science perspective "data-driven" and "theory-driven" distinguish between "bottom-up" and "top-down" approaches where thinking is influenced by information from the world or from memory, respectively. These data collection strategies are sound scientific approaches, but it is important to recognize that they are also, ultimately, connected to what is going on in the mind of the practitioner.

Both approaches require empirical data collection and a conceptual model; the difference is which of the two, data or model, has primacy. Mapping mode – relative to problem-solving mode - allows a weaker commitment to prior expectations about what will be found. Consequently, the expert has more latitude to wander in mapping mode, using exploratory procedures, such as walking a rough grid, in an area to develop a sense of the outcrops, and what observations might be available. Theoretically driven problem-solving mode, in contrast, is the product of strong expectations about what will be found so that a few high-quality observations may yield significant new insights (such as those that could discriminate among theories). In many such cases observations would not have been made without the expectation to guide searching. In practice experts might vacillate between the two modes, as one observation triggers a revision of model and thus new plans are required to explore a new line of inquiry.

Notice that observation of current practice does not unambiguously identify one approach as superior to another: They have different strengths and weaknesses. Likely the strengths and weaknesses interact with both the skills of the observer and the context of the problem. But, a fundamental problem for supporting field science and field data collection decisions is that we do not have a good metric to measure the value of any type of field data collection practice. Similarly, when students develop data collection strategies and begin to learn to coordinate models and data, it would be helpful to have measures of good and poor practices as skills develop. For example, we do not have clear evidence to guide educator's practice of providing a subset of information on an area where students will be training. The work of Baker et al. (2016) identifies a number of variables that are potentially important indicators of developing data collection decision making skills.

All science is inherently a collaboration with the past, but we are on the cusp of a transformation in this practice as we deepen collaborations mediated by computers. Computers allow access to much more of the past at any given moment in the field than ever before possible in the form of *databases*. Further, computer aided analysis can serve as a *dataguide* to enhance the use of past data to inform ongoing decisions about where and what data to collect. Fig. 1 illustrates the ways databases and dataguides might influence the familiar workflow of a geologist. Together, human-computer teams can transform model development supporting the cycle of prediction, observation, and revision that constitutes science. To invest in development of computational resources in a strategic manner requires some understanding of the value of data to experts and ultimately how computers might aid an expert.

2.2. Relative strengths of the human mind and computers

To understand the cognitive challenges and opportunities of human-computer teams consider what the human mind does well, and not-so-well, and how with the current state of the art computers might support cognitive weaknesses. Broadly speaking, humans are very good at tasks that require cognitive flexibility and poor at tasks that involve optimally combining a large number of variables. In contrast, the computer's strengths and weaknesses are the complement: Computers are inflexible, unable to apply competence in one domain to any other, but excel in rendering precise calculations with any number of variables.

Humans skill in flexibly solving problems is important for handling novel problems, and for making reasonable decisions under uncertainty due to little data (although not always with high accuracy in geology, see Bond et al., 2007). Further, humans use knowledge acquired in one domain to solve problems in another. This skill highlights a strength of human cognition: Analogical reasoning. Humans can solve problems by analogy by mapping what they know about something onto a new problem to generate a novel solution (Gentner, 1983). The generation of new hypotheses to explain conflicting data is an example where humans can work from analogy to develop new mental models of the world (e.g., Chi, 2008). This approach is well exhibited by the work of Gilbert (1877), whose account of laccoliths in the Henry mountain formation relied in part on an analogy to volcanoes. Although recent advances in qualitative spatial reasoning in Artificial Intelligence (AI) has led to impressive accomplishments in analogical reasoning, such as recognizing pattern progressions in Ravens progressive matrices (Lovett and Forbus, 2017), the application to general reasoning and novel problems is still in the future. Similarly, while humans are good at pattern recognition (as documented by Chase and Simon, 1973 for chess experts), computers are rapidly advancing in this area (Armengol et al., 2017). However, the notable human skill is to extract patterns from a small number of examples, where in contrast computers need many examples to learn patterns. Another domain where humans notably exceed current machine ability is in locomotion. Humans, like all mobile animals, are highly skilled in adapting their movements to varied terrains.

Conversely, humans are also notably poor at well-structured problems that require combining multiple variables. For example, humans deviate in predictable ways from Bayesian solutions to estimating the likelihood of events with differing base rates (Kahneman, 2000). When confronted with such problems, even experts in statistics, employ heuristics (Tversky and Kahneman, 1974). While heuristics may represent satisficing (searching for an *acceptable* solution but not necessarily the *optimal* solution) given the constraints of human memory and time, they yield predictable errors and biases that can interfere with scientific progress (Hergovich et al., 2010).

In distinct contrast, the domain of the multivariate is where computers excel. Computers can readily keep track of a vast number of variables and optimally combine them. The quality of computer reasoning in well-formulated problems is independent of the number of variables – the outcome may take longer to get to, but the outcome is no less accurate for a million variables than for three. This situation means computers are good at "needle-in-a-hay-stack" problems. At a minimum, computers are better than humans at finding optimal answers to complex searches, although computers are not necessarily always good at satisficing (Lin et al., 2015). To state the obvious, this situation is why we rely on computer assistance in calculating statistics and not our own intuitions about what is significant or likely due to chance. Even in cases where computers are solving complex problems, they are doing so in ways that yield narrowly focused solutions. IBM's Watson can provide usable (even valuable) support for oncologists. It has done so by incorporating hundreds of thousands of pages of research reports to achieve diagnostic skill that is trusted because there is high concordance with experts (Somashekhar et al., 2017). Watson is adept at handling rare diseases for non-expert doctors, in clinics where those diseases are hardly ever seen. However, for all of Watson's skill, its current ability would not transfer to other domains. For instance, it is not currently designed to diagnose other diseases – much less learn to play a new game or drive a car.

We do not anticipate that in the near future computers will be replacing experts. As Eric Schmidt, CEO of Google noted in 2013, "… humans will continue to do what we do well, and that computers will continue to do what they do very well, and the two will coexist, but in different spaces …. the separation of powers means that computers will

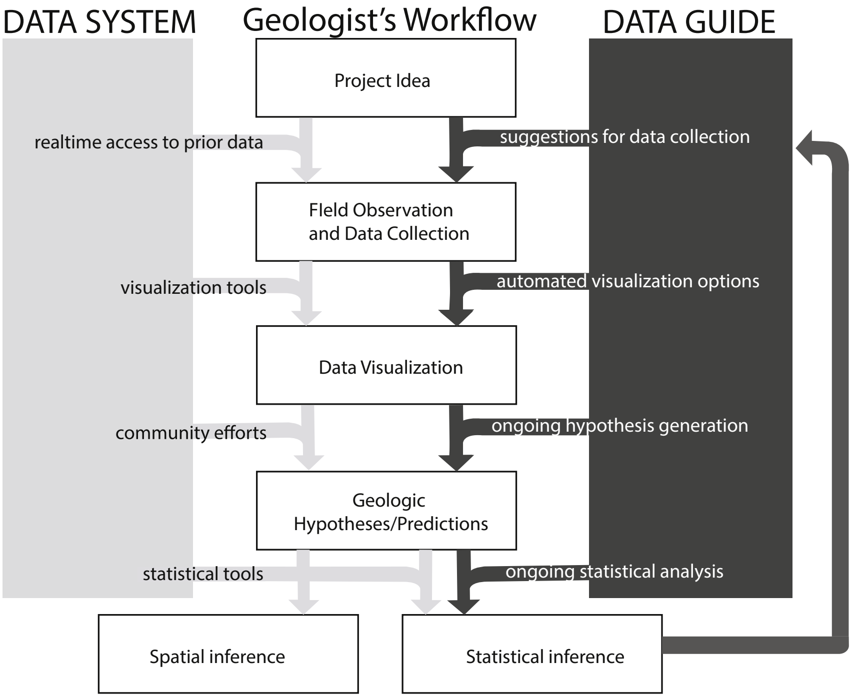

Fig. 1. How a geologist's workflow (center) may be influenced by data systems (left column), which are the current data, models and tools (e.g., visualization and analytic methods) of the research community, and dataguides (right column), which are software routines that are built from community tools to actively support each step of an individual's workflow.

sit around and help you. They'll serve as your assistants." Working in cybernetic teams recognizes the value of the two partners. We advocate research investment towards developing cyberinfrastructure that focuses on developing tools that can support the geologist's mind – their most important tool – in two ways: 1) Increasing the amount of data available to the human mind; and 2) Increasing the quality of decisions when the human seeks to optimize collection of data.

3. Sharing and combining data (in the field)

Bringing digital data into the field may seem to be simply a logistical solution to the limitations of paper and backpacks (e.g., thousands of high resolution thin sections in place of tens of photocopies of thin sections). In fact, the advent of databases for field data frees humans from their memory limitations, just as maps and photocopies do, but on a scale that may itself be transformative. Fig. 1 illustrates how community data and models (illustrated on the left as parts of the data system) might influence the familiar individual workflow in the center column. Consider the scientific potential of having a record of other's data available while in the field ("Realtime access to prior data" of Fig. 1). Having data present on location, in the midst of on-site data collection decisions, would change the nature of scientific practice and debate in three fundamental ways:

3.1. Changes in model testing

The current practice of sharing data in an interpreted form in journals, or geological maps, removes access to the raw data. This practice is changing: De Paor and Whitmeyer (2009) and Pavlis and Mason (2017) discuss the weaknesses of current practices and how digital data opportunities change field workflows. Most experts only have access to their own un-interpreted data. Access to data *in situ* allows experts opportunities to build on, reinterpret, and when needed, correct data. Evaluation of new interpretations are now done largely on the basis of internal consistency (does a proposed model account for the existing data). While there may be aspects of a field area that would clearly contradict a specific interpretation, unless that aspect were recorded, the limitations of human memory are such that it is difficult to evaluate whether or not past experiences are consistent with a proposed theoretical interpretation. Thus, access to other researcher's models in the field would allow the embedded expert to test that model – the availability of the model could guide an expert to attend to previously unnoticed (and thus unrecorded) aspects of a field area.

3.2. Changes in model development

Perhaps the most transformative opportunity provided by access to data in the field is the affordance for developing new models. Providing raw data to a broader community opens the opportunity for more scientists to draw on their areas of expertise. Each person may consider the patterns evident in the data and align them with models in their domain of training. This approach lowers the logistical barriers to interdisciplinary theory trading (Galison, 1997), and may allow consideration of application of models in one domain in another. Providing data in the field could be an especially powerful way to advance science as the field area in which the expert is embedded can help constrain and inform new models. Furthermore, a field area may offer clues to interpretation of other field areas. Accessing data in the field allows experts to search for other areas of the world that express patterns that are similar to the ones evident in the area they are currently investigating. Although current algorithms for searching for data patterns are rudimentary, at best, making data sets publically available should effectively feed a positive feedback loop of increasing search algorithm quality and data quantity. Allowing new types of visualization ("visualization tools" of Fig. 1) is another way to facilitate new model development.

Undergraduate field education in mapping often includes heuristic advice to increase the efficiency of mapping mode data collection. The practice of walking along strike and perpendicular to strike highlights the importance of continuities and discontinuities in the geometric patterns to inferring the geological structures present in an area. Such practices recognize the primacy of working from a geometric interpretation towards kinematic and dynamic models of geological processes (Shipley and Tikoff, 2016). Developing kinematic and dynamic models had until recent advances in computation required significant mathematical sophistication, at least in structural geology. The advent of digital resources in the field offer the opportunity to more closely integrate geometry, kinematics and dynamics with data collection.

3.3. Changes in data sampling

Access to raw data would fundamentally change even the basic decisions about where to collect data. The expert would no longer be on their own, they could have the past history of data collection in any area (Fig. 1). This approach affords several important opportunities: 1. Greater efficiency by not collecting redundant data. Although some data collection leaves lasting traces, many do not, so an expert cannot know by looking at an outcrop what data has been recorded there. 2. Strategic coverage by collecting data that complements existing data in areas that have not been covered or where there is inconsistency in the extant data. 3. Developing data sets that allow statistical analysis by collecting multiple observations for the same information in the same location – this approach then allows assessment of prior assumptions about the representativeness of data in a region. 4. Increased quality assurance by allowing the community to correct erroneous data in the manner of Wikipedia and OpenStreetMap. Such new community practices will be important to support trust in the digital data.

Databases can also facilitate a community effort, such as the development and testing of a community model ("community efforts" of Fig. 1). This approach is more common in scientific communities in which equipment or data is very expensive, but it could be equally well applied to community-based structural geology objectives.

Our predictions about the opportunities are largely theoretical, based on basic understanding of the strengths and limitations of the human mind and observations of current practices. The value of high resolution digital data is already evident in resolving some field problems (see Pavlis and Mason, 2017), but how it can best be used is an important question to address as such technologies become more widely adopted. The absence of evidence to guide research planning points to a critical need for evidence to inform field-database design decisions and future human-computer team field practices.

4. Cyberinfrastructure as a field partner

Section 3 highlights the advantage of supporting the human expert's memory with a computer to provide greater data access; this section considers the potential active roles a computer could play in a data collection and interpretation team that combines the strength of the flexible human mind (recognizing patterns from small number of examples, creating new solutions to problems by analogy to unrelated domains) and the strength of a computer mind (extracting patterns from massive data, immune to bias, using probability-based optimal strategies, such as those grounded in Bayesian statistics). We refer to the broad category of computer generated advice for data collection in the field as a *dataguide*. Fig. 1 illustrates how a dataguide, on the right side, might influence the familiar individual workflow in the center column. A critical piece of the architecture of a dataguide will be the reasoning based on scientific values so that expert and machine can work together to optimize the scientific value of their teamwork. Here we focus on advice on optimal strategies for data collection drawn from prior data, another important area for work would be advice based on extraction of patterns from data, and more broadly research on intelligent systems (e.g., Gil et al., in press).

We draw in part from the research literature on data collection by autonomous agents. Robots collect data in a variety of environments inhospitable to humans on Earth (e.g., Binney et al., 2010) and Mars (e.g., Woods et al., 2009). Some of the computer science design work in these areas should serve as a foundation for developing dataguides. We conceive of the human-computer teams as similar to human-robot teams, without the need to solve the difficult mobility and instrument deployment challenges that need to be overcome for a robot to autonomously collect useful data. How might a computer guide a human to collect data they would not normally collect as a matter of standard expert field practice? The computer must understand the scientific goals and the state of extant data and models.

Candela et al. (2017) outline the concept of a hypothesis map to support communication between expert and a robot about data and hypotheses. The structure allowed a robot to improve maps of rock-type by sampling in areas that would help improve geological maps that had been built from sparse observations with rock unit boundaries interpolated by eye. Such a structure could as easily guide a human as a robot to important areas in a landscape to collect data. To serve as a dataguide for structural geology experts, the system would need to evaluate Bayesian priors for all measurements of interest (Fig. 1). A simple way to achieve this goal would be to interpolate among existing data points in a region. This approach highlights the importance of establishing functions for spatial fitting of data (Pebesma, 2004), and the broad importance of developing statistical models of structural geology data (e.g., Davis and Titus, 2017).

Thinking about how best to support an expert emphasizes the importance of conceptualizing expert practice as a form of learning (Shipley and Tikoff, 2017). Thus, dataguides could support learning in each of the modes of practice with their differing goals. For example, a dataguide for reconnaissance mode should be able to provide the computably best location that combines cost (e.g., danger and time) to get to a location with its benefit (e.g., viewshed – the area visible from a specific location). In other modes, a dataguide could check incoming data for consistency and identify places where observations may be erroneous or variable and where there would be significant value to additional data. For example, in problem solving mode, a computer might aid in developing the optimal balance between sampling density and area for a given time frame given the variability and trends in the incoming data. Ultimately dataguides may include models of an area and be able to search large data sets for data that is inconsistent with a proposed model, or even propose alterative models (for examples of such a dataguide in molecular biology see Karp, 1991). Finally, we note that dataguides may also serve as potential supports for student learning in the field where instructor contact is often limited by the logistics of field space and students' independent paths.

5. Discussion

Just as an expert may provide advice to another expert when they head to a specific field area for the first time, so too may a computer provide advice based on past and incoming data. This team could improve expert field-decision making with computational resources to bring new data to the field and combine multiple variables to improve the quality of data that can be collected under conditions of limited resources. Consider three key questions for this future:

1) What is the path to developing cybernetic teams? Our experience suggests that effective science in this area will arise from interdisciplinary teams that combine geology, cognitive science, and computational science. The historian of science Peter Galison (1997) has investigated moments of technological transition in physics, including the Manhattan project. He observes that these moments involve experts from two disciplines, often theoretical and applied sciences, to come together to work towards a common goal (e.g., functional land-based radar systems). This coming together formed a trading zone, which was a physically common space (the experts literally worked side by side) and an analogically common space (the experts agreed on goals and what counted as progress). Such trading zones have a number of important properties from the perspective of advancing science. A trading zone lowers the energetic barriers to theoretical change so that progress can be made in the absence of a Kuhnian revolution (Kuhn, 1962). Furthermore, the trading zone is a practice that structures interdisciplinary interactions to allow progress. Finally, a trading zone effectively tests theories by probing them to offer direction in design decisions. Such a practice is critical to addressing *convergence science combining social and natural sciences to advance the human technology frontier* (U.S.

National Science Foundation Big Ideas report, 2016) and may transform both disciplinary geology and cognitive and computer sciences.

2) What are the key questions/hurdles/challenges? The value of the database will be determined by the richness of the data it contains and the ease in extracting useful data. Thus, important work needs to go into developing a system that is easy to input, output, and search for data. Supporting community adoption is not separate but should be integral to efforts to build the database to increase the likelihood of adoption. Comparable considerations are needed for dataguides. For example, Wiltshire et al. (2013) argues that to be effective, the technological agents must be social. Beyond providing information, a dataguide may need to have some rudimentary social skills (e.g., provide precise and accurate communication, detect and try to correct communication errors, and take into consideration all members of the team in plans), in order to be adopted and support extending science practice. We note that digital changes bring in new types and standards of evidence that are inherently statistical, professional development and education will need to provide these new analytic skills.
3) What new science questions become possible? We note informally that insights that occur in the field are predominantly small insights, limited in theoretical important, because they tend to be spatially local (e.g., recognition of relationships among observations in a local field area). The more important insights, recognized as advances by the community, tend to occur at home. We hypothesize that this situation reflects more than just incubation time, but the current state of regional information available in the field, and opportunity to develop ideas with peers. With a rich database that includes additional critical pieces, such as papers interpreting similar areas in the world, or data not usually found in the field (e.g., seismic sections), insights become possible in the field with the accompanying opportunity to immediately check them *in situ*. Thus, the rate of model development moves from a one cycle per field seasons to multiple cycles in a single season.

5.1. Digital ecosystems: what is possible (and coming)

The structural geology community – because of the lack of a digital database – is not yet fully engaged in the possibilities of a digital data system. However, community efforts, including StraboSpot and EarthCube end-user meetings, seek to integrate the field geology community into a broader effort to support cyberinfrastructure (Gil et al., 2014; Gil et al., in press). In addition to combining data across labs, a database can alter familiar community practices by introducing the possibility for new collaborations where a researcher could watch data coming into a database as it is being collected and provide guidance and suggestions to the individual in the field. Further, a database could have an *expert on call* for a given region to answer questions for visitors. Having digital records of expert's generating hypotheses from new data and ways to test them could be used to aid training of, and better understand the learning opportunities for, advanced students, and such records could also aid cognitive science as it seeks to develop ways to support disciplinary practice. The field investigator might also query the community if they encounter an interesting outcrop they could ask, where are more of "these?" In the near term, human experts in the area might have to answer these queries, but eventually a dataguide might provide an answer from satellite images or other data. The community is working to develop databases to make available the community's intellectual tools to a researcher in the field. There are active efforts to develop the pieces of the data system highlighted in Fig. 1. Here we offer a vision of future research where community tools are leveraged by AI to transform science.

Acknowledgements

The work was made possible through the NSF-sponsored SILC (Spatial Intelligence Learning Center: SBE-0541957 and SBE-1041707), both directly (via funding) and indirectly, via the interdisciplinary collaboration that is at the heart of SILC. Additional interdisciplinary collaborations that contributed to this work were funded by NSF Science of Learning Collaborative Network grant (1640800) and an NSF National Robotics Initiative grant (1734365). BT was supported by a NSF grants to develop StraboSpot (EAR 1347285; ICER-1639549) and acknowledges intellectual input from the EarthCube IS-GEO group. Terry Pavlis and Chris Andronicos are thanked for brainstorming the different mapping modes used by structural geologists. We thank Maureen Kahn for help with conceptualization and realization of Fig. 1. Terry Pavlis and Steve Whitmeyer are thanked for reading an earlier draft of this paper and providing numerous helpful suggestions.

References

Allmendinger, R.W., Siron, C.R., Scott, C.P., 2017. Structural data collection with mobile devices: accuracy, redundancy, and best practices. J. Struct. Geology 102, 98–112.
Special issue on pattern recognition techniques in data mining. Armengol, E., Boixader, D., Grimaldo, F. (Eds.), Pattern Recogn. Lett. 93, 1–202. http://dx.doi.org/10.1016/j.patrec.2017.02.014.
Binney, J., Krause, A., Sukhatme, G.S., 2010. Informative path planning for an autonomous underwater vehicle. In: Proceeding of the IEEE International Conference on Robotics and Automation, pp. 4791–4796.
Bond, C.E., Gibbs, A.D., Shipton, Z.K., Jones, S., 2007. What do you think this Is? "Conceptual uncertainty" in geoscience interpretation. GSA Today 17, 4.
Baker, K.M., Johnson, A.C., Callahan, C.N., Petcovic, H.L., 2016. Use of cartographic images by expert and novice field geologists in planning fieldwork routes. Cartogr. Geogr. Inf. Sci. 43, 176–187.
Candela, A., Thompson, D., Dobrea, E.N., Wettergreen, D., 2017. Planetary robotic exploration driven by science hypotheses for geologic mapping. In: Conference Paper, IEEE/RSJ International Conference on Intelligent Robots and Systems.
Chamberlin, T.C., 1890. The method of multiple working hypotheses. Science 15, 92–96.
Chan, M.A., Peters, S.E., Tikoff, B., 2016. The future of field geology, open data sharing and CyberTechnology in Earth science. SEPM Sediment. Rec. 14, 4–10. http://dx.doi.org/10.2110/sedred.2016.1.
Chase, W.G., Simon, H.A., 1973. Perception in chess. Cogn. Psychol. 4 (1), 55–81.
Chi, M.T.H., 2008. Three types of conceptual change: belief revision, mental model transformation, and categorical shift. In: Vosniadou, S. (Ed.), International Handbook of Research on Conceptual Change, pp. 61–82.
Davis, J.R., Titus, S.J., 2017. Modern methods of analysis for three-dimensional orientational data. J. Struct. Geol. 96, 65–89. http://dx.doi.org/10.1016/j.jsg.2017.01.002.
p. 65-89, doi:10.1016/j.jsg.2017.01.002.De Paor, D.J., Whitmeyer, S., 2009. Innovation and obsolescence in geoscience field courses: past experiences and proposals for the future. Special Pap. Geol. Soc. Am. 461, 45–56. http://dx.doi.org/10.1130/2009.2461(05).
Galison, P., 1997. Image & Logic: a Material Culture of Microphysics. The University of Chicago Press, Chicago, Illinois.
Gentner, D., 1983. Structure-mapping: a theoretical framework for analogy. Cognit. Sci. 7, 155–170. http://dx.doi.org/10.1207/s15516709cog0702_3.
EarthCube: Past, Present, and Future. EarthCube Project Report. In: Gil, Y., Chan, M., Gomez, B., Caron, B. (Eds.), EC-2014-3. Retrieved from: https://www.earthcube.org/sites/default/files/doc-repository/EC-PPF-v21-EXCERPT-FINAL_0.pdf.
Gil, Y., 32 others, 2018. Intelligent Systems for Geosciences: an Essential Research Agenda. Communications of Association for Computing Machinery (in press).
Gilbert, G.K., 1877. Report on the Geology of the Henry Mountains. United States Geological Survey Monograph, Washington, D.C.
Gilbert, G.K., 1886. The inculcation of scientific method by example, with an illustration drawn from the quaternary geology of Utah. Am. J. Sci. 31, 284–299.
Hergovich, A., Schott, R., Burger, C., 2010. Biased evaluation of abstracts depending on topic and conclusion: further evidence of a confirmation bias within scientific psychology. Curr. Psychol. 29, 188–209. http://dx.doi.org/10.1007/s12144-010-9087-5.
Kahneman, D., 2000. Evaluation by moments, past and future. In: Kahneman, D., Tversky, A. (Eds.), Choices, Values and Frames, pp. 693–708.
Karp, P., 1991. Artificial intelligence methods for theory representation and hypothesis formation. Bioinformatics 7, 301–308. http://dx.doi.org/10.1093/bioinformatics/7.3.301.
Kuhn, T.S., 1962. The Structure of Scientific Revolutions. The University of Chicago Press, Chicago, Illinois.
Lin, C.H., Kolobov, A., Kamar, E., Horvitz, E., 2015. Metareasoning for planning under uncertainty computer science. In: Proceedings of the 24th International Conference on Artificial Intelligence, pp. 1601–1609.
Lovett, A., Forbus, K., 2017. Modeling visual problem solving as analogical reasoning. Psychol. Rev. 124, 60–90. http://dx.doi.org/10.1037/rev0000039.
Mookerjee, M., Vieira, D., Chan, M.A., Gil, Y., Goodwin, C., Shipley, T.F., Tikoff, B.,

2015a. We Need to Talk: facilitating communication between field-based geoscience and cyberinfrastructure communities. GSA Today 25, 34–35.

Mookerjee, M., Vieira, D., Chan, M.A., Gil, Y., Pavlis, T.L., Spear, F.S., Tikoff, B., 2015b. Field data management: integrating cyberscience and geoscience. EOS 96 (20), 18–21.

National Science Foundation, 2016. 10 Big Ideas for Future NSF Investments. Special Report. Retrieved from: https://www.nsf.gov/about/congress/reports/nsf_big_ideas.pdf.

Pavlis, T.L., Mason, K.A., 2017. The new world of 3D geologic mapping. GSA Today 27, 4–10. http://dx.doi.org/10.1130/GSATG313A.1.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosciences 30, 683–691. http://dx.doi.org/10.1016/j.cageo.2004.03.012.

Qian, F., Jerolmack, D., Lancaster, N., Nikolich, G., Reverdy, P., Roberts, S., Shipley, T.F., Van Pelt, R.S., Zobeck, T.M., Koditschek, D.E., 2017. Ground robotic measurement of Aeolian processes. Aeolian Res. 27, 1–11 (August).

Schmidt, E., 2013. What Does a Networked Future Look like? Yale Insights. Retrieved from: http://insights.som.yale.edu/insights/what-does-networked-future-look-like.

Shipley, T.F., Tikoff, B., 2017. The role of geoscience education research in the consilience between science of the mind and science of the natural world. J. Geoscience Educ. 65, 393–398. http://dx.doi.org/10.5408/16-252.1.

Shipley, T.F., Tikoff, B., 2016. Linking cognitive science and disciplinary geoscience practice: the importance of the conceptual model. In: Krantz, R.W., Ormand, C.J., Freeman, B. (Eds.), 3D Structural Interpretation: Earth, Mind, and Machine. vol. 6. American Association of Petroleum Geologists memoir 111, pp. 219–237. http://dx.doi.org/10.1306/13561994m1113677. Hedberg Series number.

Somashekhar, S.P., Sepúlveda, M.J., Norden, A.D., Rauthan, A., Arun, K., Patil, P., Ethadka, R.Y., Kumar, R.C., 2017. Early experience with IBM Watson for Oncology (WFO) cognitive computing system for lung and colorectal cancer treatment. J. Clin. Oncol. 35http://dx.doi.org/10.1200/JCO.2017.35.15_suppl.8527. 8527-8527.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185, 1124–1131. http://dx.doi.org/10.1126/science.185.4157.1124.

Wiltshire, T.J., Smith, D.C., Keebler, J.R., 2013. Cybernetic teams: towards the implementation of team heuristics in HRI. In: Shumaker, R. (Ed.), e. Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments. VAMR 2013. Lecture Notes in Computer Scienc, vol. 8021 Springer, Berlin, Heidelberg.

Whitmeyer, S.J., Mogk, D.W., Pyle, E.J., 2009. An introduction to historical perspectives in and modern approaches to field geology education. Special Paper 461 In: Whitmeyer, S.J., Mogk, D.W., Pyle, E.J. (Eds.), Field Geology Education: Historical Perspectives and Modern Approaches. Geological Society of America, Boulder, Colorado. http://dx.doi.org/10.1130/2009.2461(00. vols. ii–ix.

Woods, M., Shaw, A., Barnes, D., Price, D., Long, D., Pullan, D., 2009. Autonomous science for an exomars roverlike mission. J. Field Robotics 26358–26390. http://dx.doi.org/10.1002/rob.20289.