The Future of Field Geology, Open Data Sharing and CyberTechnology in Earth Science

Marjorie A. Chan, Shanan E. Peters, Basil Tikoff

ABSTRACT

The central argument of this article is straightforward: ***Sedimentary field geologists must combine efforts and contribute to digital knowledge bases, or our data will simply be ignored***. If sedimentary geology data are not integrated into cyberinfrastructure initiatives, our progress will be impeded and our data and knowledge will become marginalized. Here we outline how the digital approach can move our community forward while simultaneously transforming the way we conduct science. Open data sharing will enable new collaborations, lead to new visualization developments, and enlarge our societal impact by increasing our ability to communicate our understanding of Earth system processes. We highlight current activities that can facilitate digital transitions, to allow us to fully capitalize on cybertechnology.

SEDIMENTARY DREAMS

What is the ideal future for sedimentary field geology? What if you could access all the original data for work that had been done on an outcrop, or even on the region at any spatial scale? What about accessing all the work done in allied fields (structural geology, geophysics, etc.) on that area or site? How about clicking a button and having any scientific paper that used data from the specific outcrop be immediately accessible? Web search engines, GPS, and visualization platforms, such as Google Earth, have certainly changed the way we find and locate information, but technology is on the cusp of being able to help us do so much more. Earth science combined with cyberinfrastructure can empower breakthroughs to allow us to meet the challenges of our science in transformative ways.

New technologies can help the field sedimentologist in two different but fundamentally important ways. First, they can completely change ***how*** we conduct fieldwork. Imagine being in the field with a new generation smart notebook or phone (with a very long battery life) that can sit in your pocket and automatically locate where you are. You can start talking about your observations while a voice-activated program records and conveniently puts your verbalized thoughts into a digital field system that can be easily queried while in the field and later accessed from any device or computer. Hands would be free to take samples and photos. It would be easy to click on your locality with the GPS coordinates or a map, and have access to any geological information related to that spot with the ability to zoom across multiple scales. This information includes maps, cross sections, stratigraphy, subsurface data, paleontological identifications, photos, sample information, age dating, mineral analyses, microscopic images, and other types of sample-based data. Interoperability and open data sharing would allow digital manipulations, comparisons, or visualizations across multiple data sets in the office or as you sit on the outcrop.

Second, technology can completely change ***what*** we work on in the field. What we choose to measure in the field is generally a result of what one person can carry and do with a paper notebook. When that limitation is removed – and one has direct access to the details of prior research, or assistance from airborne robotic scouts – one can start to pose new and different questions. Having access to more information in an interactive way might: a) change how much time we might spend at an outcrop, b) direct what kind or level of data or observations we would look for, and c) influence what we might sample. In short, it might help us prioritize fieldwork and data collection so as to maximize its scientific impact. Moreover, if previous research and metadata were automatically pushed to your device while in the field, it might be possible to generate hypotheses that are not otherwise formulated until a large amount of work has already been done. Interacting with what is known as we make new observations is not only time-saving, but would increase our knowledge base, and its discoverability, almost instantly.

SEDIMENTARY REALITY

How realistic are these scenarios for future sedimentary research? Although there is still a long way to go, incremental steps are bringing elements of this vision ever closer. In large part, rapid movement on this topic has been prompted by the U.S. National Science Foundation (NSF) EarthCube initiative, which is a collaboration of NSF's Division of

Advanced Cyberinfrastructure (ACI) and the Geosciences Directorate (GEO). The goal of EarthCube is to catalyze basic geoscience research, and maximize return on data acquisition and data repository infrastructure investments, by leveraging advances in information science and technology (Richard et al., 2014). In the formative governance stage to shape the community processes, EarthCube teams and leaders have articulated strategic visions and roadmaps for the future (www.earthcube.org). EarthCube's Council for Data Facilities serves in a coordinating role for existing and emerging geoscience data facilities having significant federal funding. In the initial conceptualization stage, many of the geoscience subdisciplines including the sedimentary community (Chan and Budd, 2013) held EarthCube workshops to discuss discipline-specific needs and priorities. Other field subdisciplines, such as the structure-tectonics, geochronology, and paleobiology communities expressed similar needs (all community EarthCube reports are at: earthcube.org/type-document/workshop-reports).

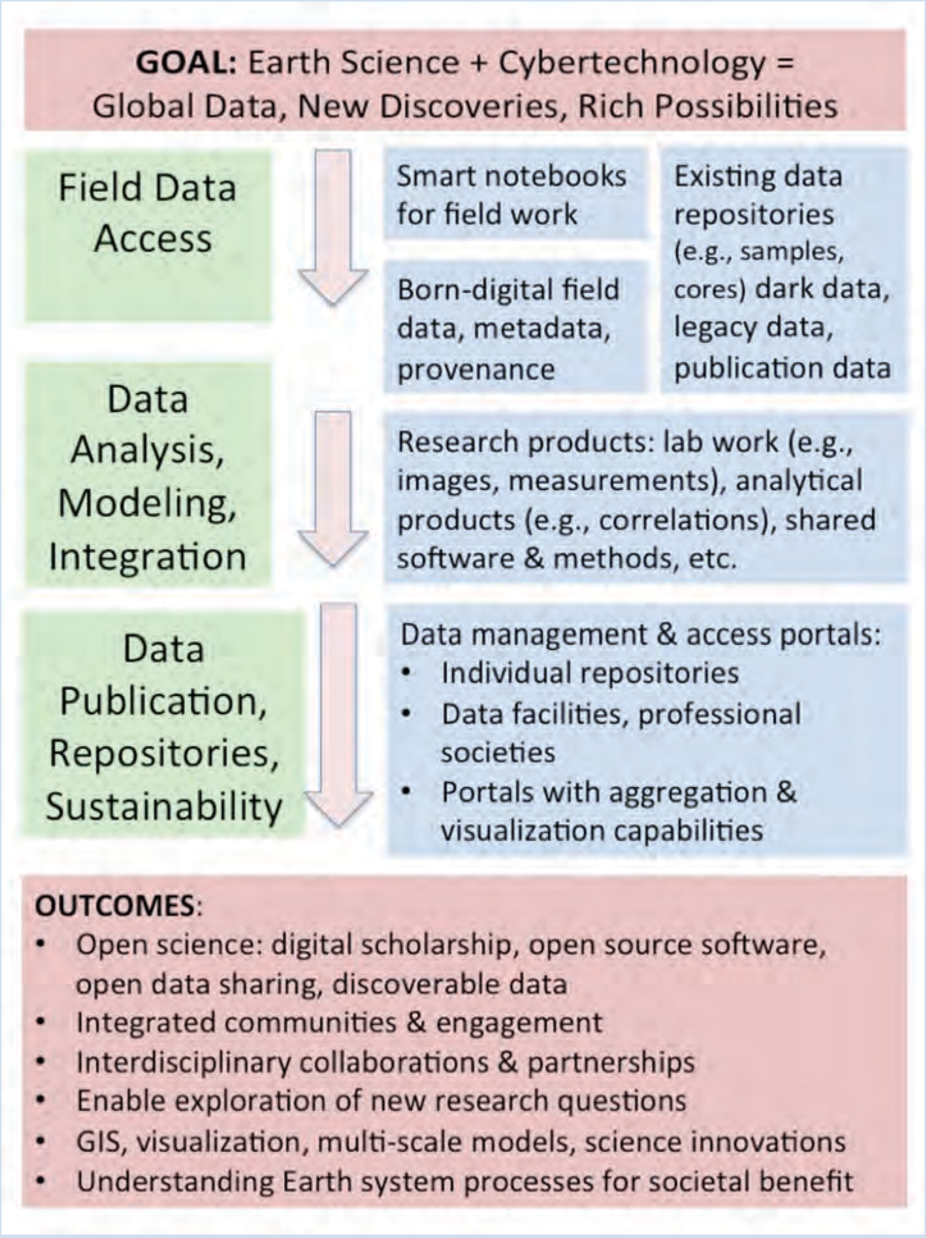


Figure 1: Potential workflow for capturing input field data, research products, and publishing outcomes for sedimentary field geology.

NSF is currently funding pilot projects (Gil et al. 2014) called building blocks (BBs) and research coordination networks (RCNs) to lead up to an implementation stage for EarthCube components. Paleobiologists have already organized and built on the momentum of a number of existing data repositories (e.g., Neotoma and the Paleobiology Database), and these are now coordinating development and community organization activities more closely than ever as a result of EarthCube. However, the two subdisciplines of sedimentary geology and structure/tectonics typically collect individual, personalized data in field notebooks and currently have no major shared data repositories. But, even this is changing. The structural geology and tectonics community is currently developing a data system for inputting field data (named Strabo). ***Sedimentary field geologists must also figure out a way to get our data into a digital format or our data will simply be ignored***. Our data are critical to establishing an environmental and temporal framework for many other Earth science subdisciplines. Certainly, one way forward is for these field-based disciplines with similar needs to partner together to create field observation data repositories. Collectively, we will be able to do revolutionary types of data analysis once we have a comprehensive pathway for field geology data that spans data collection, data integration and analysis, and data publication in sustainable repositories (Fig. 1).

For an effort like EarthCube to succeed, the social hurdles of bringing scientists and technologists together must be overcome. Communication and collaboration are important to unite the diverse stakeholders involved in EarthCube. The cyber community has a strong interest in working with geosciences data (see Gil and Pierce 2015). A fast growing group with a similar vision to use cyber technology for Earth science problems is the Federation of Earth Science Information Partners (ESIP Federation) - an open, networked community that aims to bring together science, data and information technology practitioners. While ESIP addresses some Earth science research, it does not yet reach into all of the geoscience disciplines.

To get the best outcomes for the Earth sciences, it is critical that scientists communicate and work with the information system specialists who will design products that reflect how scientists collect and ultimately use their data. An EarthCube RCN for field science fostered interdisciplinary communications and field-oriented exchange between scientists and technologists, and revealed the need for more comprehensive field hardware and software, standardization efforts for field metadata, and comprehensive data repositories for field data (Mookerjee et al. 2015a, b). A critical message is that scientists need to be heavily involved in the articulation of needs and in the design of technology so that the cyber-tools are developed to meet scientists' requirements, as opposed to scientists being expected to adapt to new technologies that do not fit the way we conduct science.

GLOBAL OPEN DATA SHARING

In our virtual world, we want to go to our computers and have instant access to all known information. That goal is achievable if we can start to archive data in revolutionary ways, akin to digital books, but with deeper levels of contextualization and machine understanding of the contents of those documents. We could access not only past journal articles, but past data tables as well without having to laboriously hand type in the data ourselves. GeoDeepDive (geodeepdive.org) is an EarthCube building block project focusing on building that type of reliable, scalable infrastructure to support geoscientists, and other disciplines. The approach is to find and extract data and information that are currently buried in the text, tables, and figures of published articles and reports. Early work by this team suggests that machine-reading approaches to knowledge-base creation can produce useful databases with quality rivaling that of human experts (Peters et al. 2014), but access to documents remains a critical challenge due to existing publisher licensing agreements and other basic access limitations. These challenges are slowly being overcome, and to date, the GeoDeepDive team has been working with Elsevier, AGU and Wiley, and the USGS to incorporate their content. The current library consists of nearly 800,000 documents and is growing in size at a rapid clip. The data in these documents are sometimes called dark data, and to help bring dark data to light, GeoDeepDive pre-processes all documents using a variety of software tools, including natural language parsing (NLP) and document layout-focused optical character recognition tools. Thus, in addition to doing simple full-text string searches (comparable to what one achieves when doing a search on digital books), it is possible to analyze the linguistic usage of and relationships between terms, how they are used in relation to figures and tables, and to write software applications to extract structured data from many thousands of documents simultaneously.

Another class of "dark data" is the legacy data produced by field geologists who have retired or passed on. These data - including that contained within field books, thin section, notes, maps, sketches of interpretations, etc. - are typically lost to future generations, reflecting a waste of both human and financial resources. For truly important datasets from inactive researchers, current dedicated community members must prioritize the value of the legacy data based on potential scientific impact so it can be translated to a usable form. In the future, if we leverage the capabilities of cybertechnology routinely in our data collection (Fig. 1), we can capture field data "born-digital" while conducting fieldwork and making observations in real time. Discussions about field-based data systems have revolved around the concept of integrating data collection into the workflow of the scientist. This approach facilitates the preservation of data for future generations.

There are many important differences between the print literature of the past and the value of open data sharing that is the vision for the future (Table 1). The strengths of digital scholarship are gaining international attention, particularly to promote transparency and open science (Alberts et al. 2015; Nosek et al. 2015). EarthCube's Geoscience Papers of the Future Initiative has trained hundreds of scientists to publish articles following best practices to use unique identifiers and citations to document data, samples, software, and provenance (Gil 2015). Additionally, the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS, www.copdess.org) aims to promote common policies and procedures for the publication and citation of data across Earth Science journals.

The expanding information and literature on digital networks, e-infrastructures and technologies, and the uses of big data is vast and overwhelming. However it is clear that open data sharing has important potential benefits that include economics, societal expectations and resources, input to decision making, and education and research innovations. These opportunities necessitate increased interdisciplinary and interagency collaborations. An open source, open data sharing approach will require intense community involvement to ensure common standards, interoperability, data traceability, quality control, preservation and storage,

Traditional Print Literature	Digital Networked Information
Physical, fixed, static, rigid	Virtual, interactive, dynamic, flexible, scalable, iterative
Geographically local, limited content	Global, unlimited content & multimedia
Referenced, key words	Easily linked with multiple identifiers, samples, methods
Centralized production, linear access	Distributed integrated production, non-linear access & discovery
Restricted formatting & tools	Multiple options, overlays, 3D - 4D visualization formats & tools
Cumbersome copying, retyping	Simple copying, identical replication
Slow knowledge diffusion & dissemination	Accelerated, rapid knowledge diffusion & dissemination

Table 1: Data sharing differences of the past at left with the present/future at right, (modified after Uhlir 2006; Uhlir and CODATA 2015).

and a host of other data principles being considered by national to international entities spanning all the sciences. As an example, adoption of data principles has been extensively considered by the 101-nation Group on Earth Observations (GEO, www.earthobservations.org). Additionally, countries in the Belmont Forum (www.belmontforum.org) are united in their effort to build international knowledge and support human action and adaption to global environmental change, which requires leadership in e-infrastructure. The Belmont Forum has formally adopted a data policy to implement standardized data management. Despite challenges, open data sharing must and will happen globally.

Planning for the future needs to start happening now. In order for geologists to access America's great databases of the past, present, and future, where would these databases reside? The data repositories need to be stable, long term, accessible, and sustainable. In some cases, we may be able to construct customized software applications to allow our data to go into existing databases. Another alternative might be to examine whether professional organizations could embrace a new role in being the ultimate long-term repository for databases. Professional societies are seeing their roles change as more journals and publications go online and open access policies proliferate. Those societies are thus asking how they can stay relevant. Assisting in the formation of long-term repositories is one potential avenue. There are many positive aspects of engaging and partnering with our professional societies, including the depth of their membership, their national to international reach, and their internal partnerships and affiliations that share field geology as well as geoinformation needs. The role would be akin to our ultimate virtual library of not only journal articles, but also networking the raw geologic databases, and rich metadata. We have already seen what kinds of struggles can occur between scientists generating the data, and privately operated, for-profit journals that own the copyright to the data and images that find their way to print. Funding agencies are interested in helping start initiatives, but need the projects to be sustainable as federal agencies cannot generally commit to long-term funding. While this issue is unresolved, it will take visionary leaders to find long-term sustainable solutions for the future.

ON THE WAY TO MAKING PROGRESS

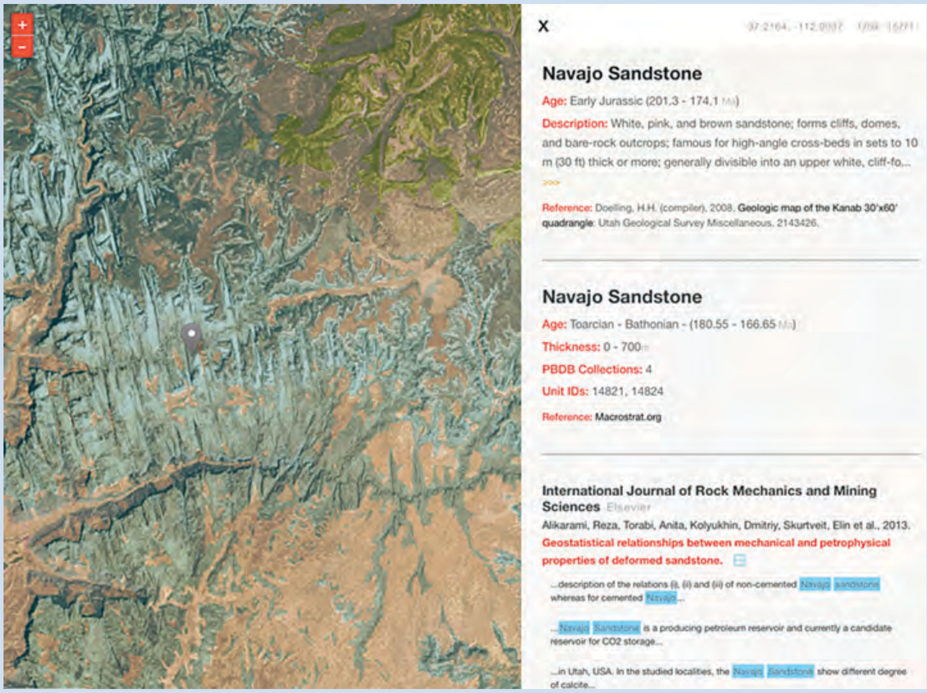
Three current examples of geological approaches that utilize cyber tools illustrate where our community is today, and how data sharing and cybertechnology can work. These examples show potential ways for our sedimentary community to move forward.

1. The System of Earth Sample Registration (SESAR)

SESAR (www.geosamples.org) is a sample registry that distributes and catalogs sample metadata and allows users to register IGSNs (International Geo Sample Number). Governed by an international implementation organization (the IGSN e.V.; www.igsn.org), the IGSN is a persistent and globally unique identifier for sites and specimens. Sample types can range from deep sea to ice cores, to rock, mineral, and fossil specimens, to synthetic specimens, to water samples and more. The use of the IGSNs in publications (Hanson 2016) can connect physical samples and sample collections across the Earth sciences with digital data infrastructures, thus improving the discovery, access, sharing, analysis, and curation of physical samples, as well as the data associated with them. Additionally, the EarthCube RCN iSamples (Lehnert et al. 2015) gathers together a broad range of stakeholders who use, curate, and access all kinds of samples to define and address the needs and challenges of digital sample management and to develop a set of community-endorsed best practices (e.g., the use of the IGSN) and standards that draw upon existing and emerging efforts both within and outside of EarthCube.

2. Macrostrat: Leveraging Existing Sedimentary Knowledge for a Data-Rich Starting Point

Sedimentary geology does yet not have a centralized data repository, but there nonetheless exists a large amount of useful published data and knowledge (of all types and qualities) on the distribution of sediments and sedimentary processes in space and time. Macrostrat's primary purpose is to integrate this existing information (Fig. 2), including regional geological columns and geologic maps, in order to facilitate the quantitative analyses that are necessary for testing a wide range of hypotheses and for calibrating models of Earth systems (e.g., Peters 2006; Finnegan et al. 2012; Halevy et al. 2012; Peters and Gaines 2012; Peters et al. 2013; Heavens 2015). Macrostrat currently has a chronostratigraphic inventory of > 33,000 surface and subsurface rock units that are also linked to more than 2.5 million geologic map polygons, tens of thousands of fossil collections (paleobiodb.org), paleocurrent measurements (Brand et al. 2015), and nearly 200,000 geochemical measurements from the USGS national geochemistry database. Building on Macrostrat's API, another program named Flyover Country (fc.unm.edu) focuses on geological discovery from a plane window and an iOS app called Mancos provides mobile users mobile access to maps, columns, and fossil collections.



Navajo Sandstone

Age: Early Jurassic (201.3 - 174.1 Ma)

Description: White, pink, and brown sandstone; forms cliffs, domes, and bare-rock outcrops; famous for high-angle cross-beds in sets to 10 m (30 ft) thick or more; generally divisible into an upper white, cliff-fo...

>>>

Reference: Doelling, H.H. (compiler), 2008, **Geologic map of the Kanab 30'x60' quadrangle**: Utah Geological Survey Miscellaneous, 2143426.

Navajo Sandstone

Age: Toarcian - Bathonian - (180.55 - 166.65 Ma)

Thickness: 0 - 700 m

PBDB Collections: 4

Unit IDs: 14821, 14824

Reference: Macrostrat.org

International Journal of Rock Mechanics and Mining Sciences Elsevier

Alikarami, Reza, Torabi, Anita, Kolyukhin, Dmitriy, Skurtveit, Elin et al., 2013.

Geostatistical relationships between mechanical and petrophysical properties of deformed sandstone.

...description of the relations (i), (ii) and (iii) of non-cemented Navajo sandstone whereas for cemented Navajo...

...Navajo Sandstone is a producing petroleum reservoir and currently a candidate reservoir for CO2 storage...

...in Utah, USA. In the studied localities, the Navajo Sandstone show different degree of calcite...

Figure 2: A screen capture of Macrostrat's geologic map interface (draped over an aerial image) shows Jurassic Navajo Sandstone exposures northwest of Kanab, Utah (close by the cover image). The image area is ~ 13 km across. Click-interactions utilize a point-based query system that gives access to: 1) bedrock unit descriptions from original map sources; 2) corresponding Macrostrat units and modeled ages; 3) literature linked to that geologic unit on the basis of rock unit nomenclature and age; and 4) usage snippets of that rock unit name from the full document text (via GeoDeepDive).

In the field, Macrostrat can help by providing broad context and rapid connections to existing knowledge and the literature. A mobile application called Rockd, currently in beta testing, is focused on facilitating the digital archiving of outcrop-based descriptions, photos, and measurements that are readily linked (with intuitive user guidance of geographic location-aware data streams) to existing Macrostrat rock units and other entities in the database (e.g., geologic map polygons, stratigraphic nomenclatural hierarchies). The connection between field data and the broader knowledge base will continue to deepen automatically as the field location data improve and as new data accrue in the literature and in Macrostrat. This digital platform can currently accommodate a variety of scientific and educational uses. However, in the future it will be able to integrate other datasets and thereby accelerate application developments focused on acquiring new or differently structured sedimentary data.

The Macrostrat infrastructure is a good starting point only because it already has a wide range of basic data and a programmatic interface for accessing them, but it currently lacks community involvement in the critical processes of curation of existing content and the generation of new field-based data summaries from other geographic regions. Rockd will be individually managed and has tools to facilitate community involvement, but there is a need for even more substantive participation. In particular, engaging regional sedimentary geologists who have gained deep knowledge through extensive fieldwork is critical to the future of Macrostrat or to any other pursuit that aims to archive sedimentary data.

The potential scientific impact and value of community participation in enhancing existing information in Macrostrat and expanding its reach geographically is hard to overstate. No matter how sophisticated or complete a database like Macrostrat becomes, it is never perfect and the work is never done. More importantly, the work of regional geologists - generating new age constraints, solving regional structural problems by mapping contact relationships, and refining our understanding of the origin and meaning of the stratigraphic record - must be more completely represented as part of the investigative process itself if any real progress will be made in bringing sedimentary geology to the information age.

3. Strabo for Structural Geology and Tectonics (SG&T)

The structural geology and tectonics community was, until very recently, in the same situation as the field-based sedimentology community - without any existing data repository. The main reason for this deficiency is that field-based structural geology data are complex, including a wide range of temporal and spatial scales (across multiple orders of magnitude), complex three-dimensional geometries, and the necessity of making temporal inferences from spatial observations.

A NSF-funded cyberinfrastructure project focuses on a data system known as Strabo (Walker et al., 2015). A key breakthrough was the development of the "Spot" concept, which allows tracking of hierarchical and spatial relations between structures at all scales, e.g., linking map scale, field mesoscale, and laboratory scale data. A Spot can be a single measurement, a group of measurements, or a relationship shared between numerous other Spots (e.g., cross-cutting relations). The Strabo data system is platform independent from mobile device to desktop, and can accommodate other digital data types (e.g., ArcGIS) to enable collection and sharing of data from field to laboratory applications.

The SG&T community was invited to engage in the development of Strabo. The community provided input through town-hall meetings and workshops to develop community standards. Upcoming field workshops will test Strabo in three pilot areas designed to illustrate the capabilities of the database. In addition, the Strabo data system will be used during the University of Kansas field camp. This is a critical step to see how the next generation of field geologists will use the system and will provide valuable feedback from a group with extensive experience with mobile devices and applications.

The Strabo effort demonstrates how another community rapidly developed the ability to report data digitally. Strabo is developing both the interface as well as the backend database to serve the community. The database further is service oriented, so any person, group, or other effort can seamlessly interact with Strabo to extract or discover data and content. Further, many of the tools necessary for structural geology are applicable to sedimentology, and the Strabo data system may be able to expand to include sedimentological data. There are significant advantages to having different types of field data included in the same data system, as it encourages valuable integration across multiple subdisciplines.

SUMMARY

The field data we acquire is unambiguously scientifically and societally relevant. As individuals and collectively as a community, sedimentary geologists need to be responsible stewards of the rich data we have generated in the past and that we will continue to generate in the future. If we integrate the goals of merging Earth science and cyberinfrastructure, along with the requisite technical skills to utilize ever-growing digital data resources into our pedagogy, this approach will provide training and will help to open the minds of the present generation of students to new lines of inquiry.

Change is happening very quickly. We must start getting our field data into digital repositories because **we cannot afford to have our data overlooked or ignored**. Our professional societies are positioned to provide a foundation and possibly a repository for geologic databases. More importantly, all of our field data have the potential to contribute to many questions, including those that are often much bigger than we originally set out to address. For example, consider how much less we would know about the history of biodiversity, the severity of global mass extinctions, or the biological impacts of global climate change if paleobiologists had never constructed Neotoma or the Paleobiology Database. What outstanding questions are we failing to be address because sedimentary geology lacks such integrative database efforts? There is always inertia to overcome when attempting to modify our scientific workflows, but it can easily become second nature, in the same way we have embraced GPS coordinates over an old survey system of township and range. We need sedimentary leaders to initiate conversations, workshops, and proposals to get our community on the right track to digital integration. While the work ahead can seem daunting and intimidating, potential collaborations and outcomes for the next generation promise to be very rewarding.

ACKNOWLEDGEMENTS

We thank D. Budd and S. Driese for their constructive reviews, D. Walker for his input, and Y. Gil for her thorough review and important suggestions. Many of the ideas presented here are an outgrowth of numerous conversations and collaborative efforts linked to the National Science Foundation's EarthCube initative. We thank the EarthCube community for their many contributions.

REFERENCES

ALBERTS, B. CICERONE, R.J., FIENBERG, S.E., KAMB, A., MCNUTT, M., NEREM, R.M., SCHEKMAN, R., SHIFFRIN, R., STODDEN, V., SURESH, S., ZUBER, M.T., POPE, B.K., JAMIESON, K.H., 2015. Self-correction in science at work. Science 26: 1420-1422. doi:10.1126/science.aab3847

BRAND, L., WANG, M., AND CHADWICK, A., 2015. Global database of paleocurrent trends through the Phanerozoic and Precambrian. Scientific Data 150025 doi:10.1038/sdata.2015.25.

CHAN, M.A., AND BUDD, D.A., 2013. EarthCube and Sedimentary Geology for the Future: SEPM the Sedimentary Record 11(2): 14.

FINNEGAN, S. HEIM, N.A., PETERS, S.E., AND FISCHER, W.W., 2012. Climate change and the selective signature of the Late Ordovician mass extinction. Proceedings of the National Academy of Sciences U.S.A. 109(18):6829–6834. doi:10.1073/pnas.1117039109.

GIL, Y. (ed.) 2015. The Geoscience Paper of the Future: OntoSoft Training. figshare, 2015. Doi:/10.6084/m9.figshare.1586773

GIL, Y., CHAN, M., GOMEZ, B., AND CARON, B., (eds.) 2014. EarthCube: Past, Present, and Future: EarthCube Project Report EC2014-3, December 2014, 108 p.

GIL, Y., and PIERCE, S.A. (eds.) 2015. Final Report of the 2015 NSF Workshop on Information and Intelligent Systems for Geosciences. NSF IIS collection at the ACM Digital Library: http://dl.acm.org/collection.cfm?id=C13 and from http://is-geo.org/home/report/

HALEVY, I., PETERS, S.E., AND FISCHER, W.W., 2012. Sulfate burial constraints on the Phanerozoic sulfur cycle. Science 337:331-334. doi:10.1126/science.1220224.

HANSON, B., 2016. AGU opens its journals to author identifiers, Eos, 97, doi:10.1029/2016EO043183. Published on 7 January 2016.

HEAVENS, N.G. 2015. Injecting climate modeling into deep time studies: ideas for nearly every project. The Sedimentary Record, v. 13(4), p. 4-10.

LEHNERT, K. A., BOWRING, S., CUTCHER-GERSHENFELD, J., DENSLOW, M., GIL, Y., HALLETT, B. W., LENHARDT, W. C., MOLINEUX, A., MOORE, C. J., NOREN, A. J., 2015, The internet of samples in the Earth sciences (iSamples): Geological Society of America Abstracts with Programs 46(6):700.

MOOKERJEE, M, VIEIRA, D., CHAN, M. GIL, Y., PAVLIS, T.L., SPEAR, F.S., AND TIKOFF, B., 2015. Field Data Management: Integrating Cyberscience and Geoscience Perspective: EOS Trans. AGU 96(20):18-21.

MOOKERJEE, M, VIEIRA, D., CHAN, M., GIL, Y., GOODWIN, C, SHIPLEY, T., AND TIKOFF, B., 2015. We Need to Talk: Facilitating communication between field-based geoscience and cyberinfrastructure communities: GSA Today 25:34-35.

NOSEK, B.A., ALTER, G., BANKS, G.C., BORSBOOM, D., BOWMAN, S.D., BRECKLER, S.J., BUCK, S., CHAMBERS, C.D., CHIN, G., CHRISTENSEN, G., CONTESTABILE, M., DAFOE, A., EICH, E., FREESE, J., GLENNERSTER, R., GOROFF, D., GREEN, D.P., HESSE, B., HUMPHREYS, M., ISHIYAMA, J., KARLAN, D., KRAUT, A., LUPIA, A., MABRY, P., MADON, T., MALHOTRA, N., MAYO-WILSON, E., MCNUTT, M., MIGUEL, E., LEVY PALUCK, E., SIMONSOHN, U., SODERBERG, C., SPELLMAN, B.A., TURITTO, J., VANDENBOS, G., VAZIRE, S., WAGENMAKERS, E.J., WILSON, R., YARKONI, T., 2015. Promoting an open research culture. Science 26: 1422-1425. doi:10.1126/science.aab2374

PETERS, S.E., 2006. Macrostratigraphy of North America. Journal of Geology 114:391-412.

PETERS, S.E. AND GAINES, R.R., 2012. Formation of the 'Great Unconformity' as a trigger for the Cambrian Explosion. Nature 484:363-366. doi:10.1038/nature10969.

PETERS, S.E., KELLY, D.C. AND FRAASS, A., 2013. Oceanographic controls on the diversity and extinction of planktonic foraminifera. Nature 493:398-401. doi:10.1038/nature11815.

PETERS, S.E., ZHANG, C., LIVNY, M. AND RE, C., 2014. A machine reading system for assembling synthetic paleontological databases. PLoS One 9(12) e113523.

RICHARD, S. M., PEARTHREE, G., AUFDENKAMPE, A. K., CUTCHER-GERSHENFELD, J., DANIELS, M., GOMEZ, B., KINKADE, D., AND PERCIVALL, G., 2014. Community-Developed Geoscience Cyberinfrastructure, Eos Trans. AGU, 95(20):165.

UHLIR, P.F., 2006. The Emerging Role of Open Repositories as a Fundamental Component of the Public Research Infrastructure: in Open Access: Open Problems, G. Sica, ed., Polimetrica, p. 62 (2006), available at http://eprints.rclis.org/9656/1/OpenAccess.pdf.

UHLIR, P.F., AND CODATA, 2015. The Value of Open Data Sharing: A White Paper for the Group on Earth Observations: Committee on Data for Science and Technology Paris, France, Plenary Document 9:7-8. http://www.earthobservations.org/documents/geo_xii/GEO-XII_09_The%20Value%20of%20Open%20Data%20Sharing.pdf

WALKER, J.D., TIKOFF, B., NEWMAN, J., GOOD, J., ASH, J., MICHELS, Z.D., ANDREW, J.E., WILLIAMS, R.T., MOLLER, A., AND RICHARD, S.M., 2015. Cyberinfrastructure for Structural Geology and Tectonics, Geological Society of America Abstracts with Programs 47(7):820.

Accepted February 2016