



# 4D

Deep-time Data  
Driven Discovery

## **4D Workshop: Deep-time Data Driven Discovery and the Evolution of Earth**

4-6 June 2018  
Washington, DC

Abstract Volume

[www.4d-workshop.net](http://www.4d-workshop.net)

<b>Abstract Title</b>	Retrieval of Martian Surface Single Scattering Albedos from Mixed Solar and Thermal Hyperspectral Imaging Data
<b>Presenting Author</b>	Raymond E. Arvidson
<b>Institution/Affiliation</b>	Washington University in Saint Louis
<b>E-mail</b>	arvidson@wunder.wustl.edu
<b>Abstract Text</b>	<p>Retrieval of surface single scattering albedos (SSA) from Mars Reconnaissance Orbiter CRISM (~0.4 to 3.9 micrometers) and Mars Express OMEGA (~0.4 to 4.9 micrometers) imaging spectrometer data demands modeling of atmospheric gases and aerosols, together with use of appropriate surface scattering and thermal emission functions. Wavelengths longer than ~2.7 micrometers exhibit both solar bidirectional and thermal directional-hemispherical terms, given typical afternoon surface temperatures of ~240 to 290 K for equatorial to mid-latitudes, when observations have good S/N. We use the Hapke Function for the surface, and DISORT-based calculations and resultant look-up tables tuned to particular scenes to model radiance on sensor, including relevant atmospheric terms. Retrieval of kinetic surface temperatures and SSA spectra at long wavelengths is done using neural networks trained with a large number of laboratory spectra, and DISORT-based simulated radiances specific for given scenes. The actual radiance data are regularized using log maximum likelihood techniques that retrieve the best estimates of signal in the presence of Poisson-dominated noise. This includes CRISM along-track over-sampled observation maps projected at 12 m/pixel relative to the native 18 m/pixel resolution. Retrieved temperature maps and SSA spectra have been validated by comparison to Curiosity rover ground truth observations, including retrieval of mineral abundances and grain sizes using nonlinear modeling of SSA spectra, and comparisons to rover-based mineral inferences. Temperature retrievals were validated by comparison to rover-based radiometer measurements and local-scale thermal models.</p>

<b>Abstract Title</b>	Carbonates before skeletons: A database approach
<b>Presenting Author</b>	Kristin Bergmann
<b>Institution/Affiliation</b>	Massachusetts Institute of Technology
<b>E-mail</b>	kdberg@mit.edu
<b>Additional Authors</b>	Cantine, MD, Knoll, AH
<b>Abstract Text</b>	<p>Carbonate minerals have precipitated from seawater for the last 3.8 billion years, but where and how they precipitate has changed through geologic time. The earliest carbonates precipitated on the seafloor as crystal fans until ocean oxygenation, coupled with aerobic microbial respiration, made the sediment-water interface caustic for carbonate sedimentation (Bergmann et al., 2013; Higgins et al., 2009). The locus of carbonate precipitation and the dominant carbonate sediments can be used as a high-resolution proxy, in both space and time, for oxygenation and seawater chemistry.</p> <p>Geobiologists have successfully used large datasets to track fluctuations in Earth's chemical and biological cycles. Here we add to the growing large datasets to query the early earth by compiling a high-resolution database of global carbonate sedimentation. This dataset of Earth's 3.8 billion to ~500 million years old carbonate rocks are our best proxy for carbonate sedimentation in deep time. The Catalogue of Carbonate Sedimentology and Stratigraphy (C2S2) currently contains 144 formations, digitized at the meter scale and classified by environment of deposition. Lithofacies details are recorded for each platform, including a range of microbial fabrics, mineralogy, depositional environmental, age and location.</p> <p>C2S2, a temporal-spatial compilation of trends in sediments and fossils, represents a research tool not previously available to geobiologists. We can pinpoint the depth-dependent timing of oxygenation and the transition from anaerobic to aerobic respiration at the seafloor. C2S2 tracks dolomitization through time which can be correlated with other proxies for changing seawater chemistry. Applying C2S2 to ecologic questions will allow us to better understand the habitats critical to evolution.</p>

<b>Abstract Title</b>	Phinch: An interactive, exploratory data visualization framework for -Omic datasets
<b>Presenting Author</b>	Holly Bik
<b>Institution/Affiliation</b>	University of California, Riverside
<b>E-mail</b>	holly.bik@ucr.edu
<b>Additional Authors</b>	Pitch Interactive
<b>Abstract Text</b>	<p>The sheer volume of data produced from high-throughput sequencing platforms (e.g. Illumina HiSeq/MiSeq) requires new paradigms for effective data analysis. Scientific visualization represents an innovative method towards tackling current bottlenecks; in addition to giving researchers a unique approach for exploring large datasets, it stands to empower biologists with the ability to conduct powerful analyses without requiring a deep level of computational knowledge. Phinch (<a href="http://phinch.org/">http://phinch.org/</a>) is an interactive, exploratory visualization framework that can be used to identify biological patterns in high-throughput environmental datasets (microbial ecology OTUs and metagenomes). Leveraging a close collaboration between UC Riverside and Pitch Interactive (a data visualization studio in Oakland, CA), this project takes advantage of standard file formats from computational pipelines in order to bridge the gap between biological software (e.g. QIIME) and existing data visualization capabilities (e.g. visualization-specific programming language such as D3.js). Phinch v2.0 represents a refactored framework released as an Electron desktop app, reducing previous reliance on cloud servers and improving support for larger file size imports. Other improvements include extended support for all BIOM file formats, extended export/sharing features, and a new suite of planned visualization tools and user interactions.</p>
<b>References</b>	<p>Bik HM, Pitch Interactive (2014) Phinch: An interactive, exploratory data visualization framework for –Omic datasets, bioRxiv, <a href="http://dx.doi.org/10.1101/009944">http://dx.doi.org/10.1101/009944</a></p>

<b>Abstract Title</b>	A proposed detrital-zircon based method of reconstructing global subduction flux through Earth history
<b>Presenting Author</b>	Dwight C. Bradley
<b>Institution/Affiliation</b>	Dartmouth College
<b>E-mail</b>	bradleyorchard2@gmail.com
<b>Abstract Text</b>	<p>Judging from the age distribution of oceanic crust, the global plate flux has averaged ~3.4 km<sup>2</sup>/yr since 180 Ma (Rowley, 2002). The detrital zircon record may hold clues to changes in global plate flux through Earth history. Because most zero-age zircons at the Earth's surface are products of subduction and most older zircons are products of collision, the abundance of zircon ages through time is likely to preserve a signal of the global plate flux rate. Data teased from a global compilation of 198,000 detrital zircon (DZ) ages (Voice et al., 2011) illustrate the concepts. DZ from modern sands decline in abundance with age, due to zircon attrition. Superimposed on this long-term trend are order-of-magnitude fluctuations, with maxima at ca. 2700, 1900, 1000, and 450 Ma and minima at ca. 2300, 1600, 900, 700, 400, and 210 Ma. The fluctuations correlate with plate reorganizations that are the inevitable consequence of supercontinent cycles (Bradley, 2011). The age distribution is also shaped by secular changes in the global tectonic regime; by differences in zircon production at rifts, arcs, and collisional orogens; by preservation, exhumation, destruction, and recycling of zircons; and by methodological and regional biases. A distillation of many of these influences is shown by the normalized age distribution of DZ, which shows a quasi-exponential form with a "half-life" of about 540 m.y. This trend allows DZ age distributions to be corrected for attrition. Among several unsolved but tractable problems are how to best calibrate the corrected age distribution from modern sand DZ against the known plate flux, and how to cull those zircons that are not products of plate convergence. A carefully planned, crowd-sourced sample suite of ~6000 modern sands and ancient siliciclastic rocks (analytical costs: \$3,000,000) is needed to establish the plate flux rate since 3 Ga.</p>
<b>References</b>	<p>Bradley, D.C., 2011, Secular trends in the geologic record and the supercontinent cycle: <i>Earth-Science Reviews</i>, v. 108, p. 16-33. doi: 10.1016/j.earscirev.2011.05.003.</p> <p>Rowley, D.B., 2002. Rate of plate creation and destruction: 180 Ma to present. <i>Geological Society of America Bulletin</i> 114, 927–933.</p> <p>Voice, P.R., Kowalewski, M., and Eriksson, K.A., 2011. Quantifying the timing and rate of crustal evolution: global compilation of radiometrically dated detrital zircon grains. <i>Journal of Geology</i> 119, 109–126. DOI: 10.1086/658295</p>

<b>Abstract Title</b>	Planetary Data Protection - The CheMin Database
<b>Presenting Author</b>	Thomas Bristow
<b>Institution/Affiliation</b>	NASA Ames Research Center
<b>E-mail</b>	thomas.f.bristow@nasa.gov
<b>Additional Authors</b>	Lafuente B, Stone N, Blake DF, Downs RT, Pires A, Sarrazin P, Keller RM, Dateo CE, Fonda M
<b>Abstract Text</b>	<p>The CheMin X-ray diffractometer (XRD), onboard the Mars Science Laboratory (MSL) rover Curiosity, is the first crystallographic instrument deployed on another planet. CheMin data has proven key in reconstructing ancient habitable surficial environments on Mars, preserved in ~3.7 Ga fluvial-lacustrine strata at Gale crater. The mineralogy and compositions of primary detrital minerals in these sediments provide novel insights into the nature of martian igneous rocks and styles of volcanism. Evidence of periods of aqueous alteration influencing these rocks billions of years after lithification is recorded by secondary minerals. The mineral abundances and compositions determined from CheMin comprise the only full mineralogical data set for Mars surface materials that is currently available. The CheMin database, published using the Open Data Repository (<a href="http://odr.io/CheMin">http://odr.io/CheMin</a>), is a living repository of CheMin and related MSL data integrated with tools and procedures for visualization and analysis. We aim to give universal access to the information and data analysis tools that are required to understand and re-analyze the original raw data, replicate experiments, or even perform entirely new studies with different starting hypotheses. Each data record includes: 1) sample description; 2) interactive XRD and XRF patterns with associated metadata and downloadable files; 3) mineral abundances derived from diffraction data; 4) access to the library of CIF files used in diffraction pattern analysis; 5) links to raw data and results from other MSL instruments (such as elemental composition data from APXS) for each of the samples analyzed by CheMin; 6) library of downloadable open source references associated with each analysis; 7) access to the Experiment Data Records (EDRs) for each sample; 8) a detailed narrative of how the analysis was performed. The database also provides access to QAnalyze (<a href="http://www.qanalyze.com/">http://www.qanalyze.com/</a>), an automated cloud-based application for quantitative analysis of mineral samples using X-ray diffraction (XRD).</p>

<b>Abstract Title</b>	Enzymatic Archaeology: Understanding The RNA World Through Modern Enzyme Substrate Interactions
<b>Presenting Author</b>	Christopher J. Butch
<b>Institution/Affiliation</b>	Earth Life Science Institute
<b>E-mail</b>	chrisbutch@gmail.com
<b>Additional Authors</b>	Donato Giovannelli
<b>Abstract Text</b>	<p>The RNA World theory proposes that prior to modern life, proteins and DNA did not exist, with information and catalytic function instead encoded in RNA or RNA-like polymers (Joyce 2002). Modern biochemistry and LUCA are then theorized to have emerged from the functions initially encoded in these multipurpose RNA molecules. If this trajectory truly took place, proteins preserved since the last universal common ancestor may have evolved initially to handle polymers other than RNA or DNA. We attempt to probe the substrate permissivity of these ancient enzymes by investigating interactions of homologous enzymes sampled broadly from across the bacterial tree of life with ligands which could have been alternatives to RNA in the pre LUCA world.</p>
<b>References</b>	<p>Joyce, Gerald F. "The antiquity of RNA-based evolution." <i>Nature</i> 418.6894 (2002): 214-221.</p> <p>Cleaves, H. James, Markus Meringer, and Jay Goodwin. "227 Views of RNA: Is RNA Unique in Its Chemical Isomer Space?." <i>Astrobiology</i> 15.7 (2015): 538-558.</p>

<b>Abstract Title</b>	Structures in 4D: a 'double tale' of biological accretion
<b>Presenting Author</b>	Caetano-Anollés G
<b>Institution/Affiliation</b>	University of Illinois at Urbana-Champaign
<b>E-mail</b>	gca@illinois.edu
<b>Additional Authors</b>	Caetano-Anollés D, Caetano-Anollés K
<b>Abstract Text</b>	<p>"A double tale I'll tell. At one time one thing grew to be just one from many, at another many grew from one to be apart. Double the birth of mortal things, and double their demise." -On Nature, P. Strasb. Gr. Inv. 1665-6, Empedocles [1]</p> <p>The evolution of structure in biology is driven by accretion and change. Accretion brings together disparate parts to form bigger wholes. Change provides opportunities for growth and innovation. Here we review patterns and processes that are responsible for a 'double tale' of evolutionary accretion at various levels of complexity, from proteins and nucleic acids to building structures in cities. Parts are at first weakly linked and associate variously. As they diversify, they compete with each other and are selected for performance. The emerging interactions constrain their structure and associations. This causes parts to self-organize into modules with tight linkage. In a second phase, variants of the modules evolve and become new parts for a new generative cycle of higher-level organization. Evolutionary genomics and network biology support the 'double tale' of structural module creation [2] and validate an evolutionary principle of maximum abundance that drives the gain and loss of modules. Examples of structures in 4D include the origin of the ribosome, the proteome and metabolic networks [3-7] and the emergence of tall buildings in the Island of Manhattan [8].</p>
<b>References</b>	<ol style="list-style-type: none"> <li>1. Janko R (2004) Zeitschrift papyrologie epigraphik 150:1-26</li> <li>2. Mittenthal JM, Caetano-Anollés D, Caetano-Anollés G (2012) Front Genet 3:147</li> <li>3. Caetano-Anollés K, Caetano-Anollés G (2013) PloS One 8:e59300</li> <li>4. Nasir A, Kim KM, Caetano-Anollés G (2014) PLoS Comput Biol 10:e1003452</li> <li>5. Caetano-Anollés D, Caetano-Anollés G (2015) Comput Struct Biotech J 13:427-447.</li> <li>6. Aziz MF, Caetano-Anollés K, Caetano-Anollés G (2016) Sci Rep 6:25058</li> <li>7. Nasir A, Kim KM, Caetano-Anollés G (2017) Front Microbiol 8:1178</li> <li>8. Schlomo A, Lamson-Hall P (2014) Marron Inst Urban Mgmt (New York Univ) 14:1-48</li> </ol>

<b>Abstract Title</b>	Does the Earth have a Pulse? Constraints from the Geological Record and Implications for Tectonic Processes
<b>Presenting Author</b>	Peter A. Cawood
<b>Institution/Affiliation</b>	School of Earth, Atmosphere and Environment
<b>E-mail</b>	peter.cawood@monash.edu
<b>Additional Authors</b>	Chris Hawkesworth, Bruno Dhuime, Aubrey Zerkle
<b>Abstract Text</b>	<p>Earth is a dynamic, evolving system in which the surficial and solid components interact through a series of cycles and at a variety of scales in response to energy from internal heat and external sources within the solar system. The grand challenge for the Earth Sciences is to unravel the feedbacks between the deep and surficial Earth, the record of which is preserved in the continental crust. The continental record indicates that the distribution of rock units and events is heterogeneous with distinctive peaks and troughs in the ages of igneous crystallization, metamorphism, continental margins and mineralization, and progressive changes in atmosphere and ocean compositions.</p> <p>Based on the geological record, we recognize 6 stages of Earth evolution: 1) Initial accretion and differentiation of the core/mantle system within the first few 10's of m.y., on an anoxic prebiotic Earth; 2) Generation of crust prior to 3.2 Ga in a pre-plate tectonic regime associated with evolution of early life and low oxygen atmosphere; 3) Transition to plate tectonic regime from 3.2 Ga to 2.5 Ga involving development of rigid lithosphere with change from mafic to more felsic composition of continental crust, an increase in crustal thickness and recycling, and the initial emergence of continents with impacts on ocean and atmospheric chemistry; 4) Early sustained plate tectonics involving hot subduction with shallow slab breakoff over the period from 2.5-1.7 Ga, associated with further increases in the proportion of thick felsic continental crust, and changes in surficial Earth including the initial rise in atmospheric oxygen and global glaciations; 5) Earth's middle age from 1.8-0.8 Ga, characterized by lithospheric, environmental, and evolutionary stability, and the evolution of early eukaryotes; 6) Initiation of modern cold subduction at ~0.8 Ga, associated with a second rise in atmospheric oxygen, extensive global glaciations, and the radiation of animal life.</p>

<b>Abstract Title</b>	Computational Predictions Amphiphile Aggregation for Early Compartmentalization
<b>Presenting Author</b>	Jim Cleaves
<b>Institution/Affiliation</b>	Earth-Life Science Institute/Institute for Advanced Study
<b>E-mail</b>	henderson.cleaves@gmail.com
<b>Additional Authors</b>	Gillams R ,Meringer M
<b>Abstract Text</b>	<p>Extant biology uses a vast array of lipids to perform a range of tasks, and compartmentalization is critical for Life's existence by providing, a separation of chemical environments, enhanced local concentration of molecules, interfaces with reduced dimensionality, and individuality, leading to competition and evolution. We wished to explore and predict which kind of molecules are able to aggregate to form compartments that can host and/or encourage complex and perhaps even simple life-like chemistry that can be assayed easily in vitro. There may be a very large number of such molecule types, and the use of high-resolution models is computationally prohibitive. We thus set out to develop an efficient way to predict aggregation and screen large in silico-generated compound libraries.</p> <p>There are a range of methods available for producing or accessing libraries of molecules. Through the recent explosion of lipidomics, there are a number of tools developed for mass spectrometry that include large compound libraries (e.g. LipidBlast, LipidHome, etc.). These give access to biologically relevant lipids, but do not facilitate the identification of novel molecules.</p> <p>We have identified computationally cheap methods for the generation of exhaustive lipid libraries and the evaluation of their propensity to self-assemble into either micelles or vesicles. Depending on user-defined parameters such libraries can easily contain well past trillions of molecules. We used MolGen (<a href="http://www.molgen.de/">http://www.molgen.de/</a>) for exhaustive generation of sub-libraries of lipid tails and heads. MolGen allows for disallowed molecular motifs and ranges of molecule parameters to be defined for the output.</p> <p>Once generated, solubility properties are assessed using QSPR models, and geometric properties computed. These are then combinatorially reacted using ChemAxon's Reactor software (<a href="https://chemaxon.com/">https://chemaxon.com/</a>) to give a final library. We finally evaluate them using chemoinformatics approaches to identify molecules that possess properties commensurate with an ability to form micelles, and more discriminatively, vesicles.</p>

<b>Abstract Title</b>	Scaling the opportunities for survival of subsurface life in Earth across space and time
<b>Presenting Author</b>	Frederick Colwell
<b>Institution/Affiliation</b>	Oregon State University
<b>E-mail</b>	rcolwell@coas.oregonstate.edu
<b>Additional Authors</b>	Andrew Thurber
<b>Abstract Text</b>	<p>Scientists who study subsurface microbial communities have familiar environmental rules as starting points for framing where and how deep ecosystems exist. Factors such as temperature, redox gradient, overlying ocean productivity, sediment age, and microbial power requirements provide a scaffold for indisputable physical and chemical constraints that dictate where life exists and where it is prohibited on Earth. But can we catalog all of the geological phenomena that enforce these governing rules and their spatial and temporal spectra? What life-sustaining Earth processes are we currently missing because biologists overlook them? We describe the use of Stommel diagrams - devised originally to guide oceanographic sampling - to depict an array of Earth phenomena that overlap on spatial and temporal scales familiar to geologists and geomicrobiologists. By examining in detail a particular Earth process (e.g., seismicity or subsurface fluid flux) in its various forms we hope to resolve distinct conditions that sustain subsurface life and guide our search for new places to find deep life. Spatially, the geological settings that influence the activity and distribution of microbes range from the molecular to the planetary-scale (up to 10<sup>5</sup> km<sup>3</sup>). Temporally, life in the subsurface may respond in time units of seconds (of course, familiar to humans) to multi-millennial. We hypothesize that microbiologists who study the subsurface are still gathering geological and planetary processes that dictate life in the subsurface. A more complete understanding of deep life should emerge as we scale habitats spatially to continents and ocean basins and temporally to long, dynamic Earth cycles.</p>

<b>Abstract Title</b>	Evaluation and Opportunities for Implementation of Machine Learning in the U.S.G.S. Mineral Resources Program, a Scoping Project
<b>Presenting Author</b>	Coyan JA
<b>Institution/Affiliation</b>	Department of the Interior, U.S. Geological Survey, Spokane, Washington 99201
<b>E-mail</b>	jcoyan@usgs.gov
<b>Additional Authors</b>	Robinson GR
<b>Abstract Text</b>	<p>Implementation of machine learning and big data science has been successful in commercial applications and, more recently, has demonstrated potential for solving geoscience problems. Geoscience data, is vastly more complicated than commercial products or social media datasets, as it incorporates understanding of processes from numerous domains, contains information with ambiguous boundaries and latent variables, has a spatio-temporal component, is highly variable, exists in multiple resolutions, may lack stationarity, and may be incomplete or involve missing data (Karpatne et al., 2017).</p> <p>The USGS has approved a one-year scoping project to investigate methods of machine learning and its current and potential applications to the geological sciences, and more specifically, the Mineral Resources Program (MRP). The scoping project broadly covers 4 topics: 1) evaluation of current and emerging uses for machine-learning technology as it applies to mineral and energy resources, 2) determination of what geologically-oriented neural-network-training datasets are available and what role the USGS may have in creating or contributing to publicly available training datasets (e.g., MINDAT, USGS's Mineral Resources Online Data Catalog System, Mineral Resources Online Interactive Map, USGS Geologic Map Database, etc.), 3) identification and engagement of potential collaborators (e.g., computer programmers, computational data scientists, research institutions, statisticians), and 4) submission of a more comprehensive proposal for a multi-year funded project. The goals of the long-term project include further fleshing-out the machine learning/artificial intelligence/neural network needs of the USGS and MRP, and collaboration with partners to produce publicly available consumables in this domain (e.g., analytical methods, training datasets, ontology of ML algorithms and best practices, enhanced visualizations of mineral systems deposits, etc.). This project provides a unique opportunity to explore application of machine learning tools to a government entity, geological sciences, and, more specifically, mineral resources.</p>
<b>References</b>	Karpatne A. et al. (2017). Theory-guided Data Science: A New Paradigm for Scientific Discovery. IEEE Transactions on Knowledge and Data Engineering ( Volume: 29, Issue: 10, Oct. 1 2017 )

<b>Abstract Title</b>	Constraining the crustal velocity structure of the conterminous United States using the autocorrelation of earthquake-generated body-waves
<b>Presenting Author</b>	Jonathan R. Delph
<b>Institution/Affiliation</b>	Rice University
<b>E-mail</b>	jrdelph@rice.edu
<b>Additional Authors</b>	Levander A, Niu F
<b>Abstract Text</b>	<p>To determine the discontinuity structure Earth's crust, passive seismic methods have primarily focused on differences in vertically and radially-polarized energy in the coda of earthquake-generated body waves (e.g., receiver functions). This approach effectively separates primary body waves from their conversions to other phases at discontinuities (P-to-S-wave or vice-versa); however, it fundamentally relies on a linearly dependent combination of parameters to transfer the timing of conversions to the depth domain, which is more useful to geoscientists. In order to do this, one must make a priori assumptions about the velocity structure in their region of interest, which can bias resulting discontinuity models. As these models are commonly used as a starting point for more complex studies, the often unquantified uncertainties related to these assumptions can propagate much further than the initial study.</p> <p>In our study, we focus not only on differences in body-waves and their coda, but also on their similarities through an autocorrelation technique. This allows us to obtain a system of linearly independent equations that can constrain both P and S-wave velocity structure and depths to discontinuities, rather than only two of the three. We quantify uncertainties in our results using a Bayesian approach, and apply this to &gt;100 seismic stations that are spatially distributed around the United States to obtain a model unbiased by weakly-constrained assumptions.</p> <p>This dataset is part of the larger Earthscope Transportable Array, comprising ~2000 seismic stations that operated for an average of ~2 years each. Over this time, each seismic station recorded ~500 earthquakes appropriate for this method, making manual quality control time-intensive and infeasible. However, using this as training dataset for a machine-learning algorithm can make quality control consistent and feasible, enabling the computation of a high resolution model of crustal structure in the United States when applied to the entire Earthscope dataset.</p>

<b>Abstract Title</b>	Demonstrating the Astrobiology Habitable Environment Database (AHED)
<b>Presenting Author</b>	Angela M. Detweiler
<b>Institution/Affiliation</b>	NASA Ames Research Center/ Bay Area Environmental Research Institute
<b>E-mail</b>	angela.m.detweiler@nasa.gov
<b>Additional Authors</b>	Keller RM, Lafuente Valverde B, Blake DF, Bristow TF, Cooper GW, Dateo CE, Des Marais DJ, Everroad RC, Jahnke LL, Kubo MD, Parenteau MN, Prufert-Bebout LE, Stone N
<b>Abstract Text</b>	<p>Astrobiology is an interdisciplinary field that draws from disciplines including biology, chemistry, geology, physics, astronomy, computer, and planetary science to answer complex questions about the origin, evolution and distribution of life on Earth, and the search for life beyond Earth. The Astrobiology Habitable Environment Database (AHED) group at NASA Ames Research Center brings together researchers from these diverse fields to work towards the creation of a repository of databases that can accommodate the needs of different members in the astrobiology community. Based on software by the Open Data Repository (<a href="http://www.opendatarepository.org">http://www.opendatarepository.org</a>), AHED aims to be a dynamic data system intended to allow researchers to discover, analyze, visualize, and use the data more readily than in a typical repository by providing metadata standards [1], and leveraging a host of outside resources that can be integrated with the AHED data environment. These resources include open-source plugins, visualization libraries (e.g. D3.js, plotly.js), and an API interface for more in-depth data analysis using programming languages such as Python and R.</p> <p>Some of these capabilities are illustrated by databases currently in development by the AHED working group, including a Hypersaline Systems database that catalogs the isotopic composition of methane and associated metadata from hypersaline environments, considered to be analogs for some Mars environments. Additionally, the Ames Exobiology culture collection database of astrobiologically relevant organisms involves development of a specific metadata standard to facilitate linking with other culture collections and laboratory studies. A universal and open-source Gas Chromatography-Mass Spectrometer (GC-MS) interface is also in development to facilitate sharing and interpretation of biomarker and organic molecular data. The ultimate goal is to create a platform where data is readily available in a format that will facilitate active, onsite data cataloging, collaboration, discovery, and analysis for the growing field of astrobiology.</p>
<b>References</b>	[1] Keller RK, Detweiler AM, Lafuente Valverde B, Blake DF, Bristow TF, Cooper GW, Dateo CE, Des Marais DJ, Jahnke LL, Kubo MD, Parenteau MN, Prufert-Bebout LE, Stone N. Steps Toward Improved Integration, Search, and Analysis of Heterogeneous Data in the Astrobiology Habitable Environments Database, 4D Workshop, June 4-6, 2018 (submitted)

<b>Abstract Title</b>	Application of Machine Learning to Mineralogical Identification
<b>Presenting Author</b>	Joshua F Einsle
<b>Institution/Affiliation</b>	Dept of Earth Sciences and Department of Materials Science and Metallurgy, University of Cambridge
<b>E-mail</b>	jfe26@cam.ac.uk
<b>Additional Authors</b>	Martineau B, Buisman I, Vukmanovic Z, Johnstone D, Eggeman A, Midgley PA, Harrison RJ
<b>Abstract Text</b>	<p>Modern electron microscopes offer a variety of techniques for mapping chemical and crystallographic information in samples from micron to nanometer scales in both two and three dimensions. A typical data set comprises hundreds of thousands of spectra acquired pixel by pixel as the electron beam scans over the specimen. Routine analysis procedures include background subtraction, peak identification, and quantification. However, while this approach may work on point spectra and on systems where the nanostructures are not buried, it is a physically impossible task when working with an elemental map of millions of spectra. Application of data decomposition techniques (similar to Principal Component Analysis (PCA)) to multidimensional data sets reveals key and often overlooked features. The use of machine learning in a wide range of systems has been demonstrated to allow for the unmixing of buried chemical structures and identification of phases[1]–[3]. This paper will demonstrate how the use of machine learning on Energy Dispersive Spectroscopic (EDS) datasets allows for the unique identification of distinct chemical phases and present insights into the limitations of this approach. SEM-EDS (Scanning Electron Microscopy) maps of a symplectite texture found in a sample from the Bushveld Large Igneous Province demonstrates both the power of machine learning to pull out material phases, but also the limitations of EDS as a technique for phase identification classification. Due to the compositional similarity of many natural materials (in this case, all four phases are Si and O rich), EDS characterization does not present a fully linearly independent description of the material system studied. Since spectrographic data does not include any crystallographic information, the boundary conditions of a fully described dataset are not met. I will present a method for addressing this through the use of a statistical spectral model of the system, combined with statistically derived phase masks.</p>
<b>References</b>	<p>[1] Z. Saghi, et al, Nano Lettets, 11 (2011), p. 4666.  [2] A. S. Eggeman, R. Krakow, and P. A. Midgley, Nature Communications, 6 (2015). p. 1.  [3] J. F. Einsle, et al, In Review, (2018).</p>

<b>Abstract Title</b>	The Analytics Pipeline: Data Visualization and Exploratory Analysis
<b>Presenting Author</b>	Ahmed Eleish
<b>Institution/Affiliation</b>	Rensselaer Polytechnic Institute
<b>E-mail</b>	eleisa@rpi.edu
<b>Abstract Text</b>	<p>The human visual system is an amazing analytical tool, we use it all the time to detect patterns and anomalies on the fly without much prior knowledge. We rely on it to navigate the physical world and as perhaps our primary channel of informational input. As such, our eyes are uniquely capable of identifying interesting aspects of visualizations and diagrams produced from data or from some aggregation or summarization of data. This goes hand in hand with a proliferation of software tools and packages across multiple platforms and technology stacks that facilitate the consumption of large datasets and the development of analyses and visualizations with a high degree of customizability and a focus on presentation and dissemination.</p> <p>Visualizations are especially useful in exploratory data analysis where batches of diagrams could be created with relative ease to demonstrate and enable the comparative examination of various characteristics of a dataset. These diagrams could be further customized with annotations to make interesting observations more apparent and communicable.</p> <p>This poster presentation will introduce the audience to the latest and greatest in data analysis and visualization software tools and packages with a focus on the features for visualization, as demonstrated through use cases from the earth sciences. The topics to be presented would include producing diagrams that represent various aspects of a dataset including the summarization and aggregation, frequency distribution, dimensionality reduction, and projection of data.</p>

<b>Abstract Title</b>	Computational Systems Chemistry
<b>Presenting Author</b>	Christoph Flamm
<b>Institution/Affiliation</b>	University of Vienna
<b>E-mail</b>	christoph.flamm@univie.ac.at
<b>Additional Authors</b>	Andersen JL, Merkle D, Stadler PF
<b>Abstract Text</b>	<p>The design and analysis of complex chemical systems is a difficult task, which for interesting systems requires assistance from software. We present formal methods based on core computer science for modelling of chemical compounds and reactions. The methods are aimed at exploration and generation of chemical networks based on a high-level description of the chemistry and for analysing the networks to find both generic pathways and specialised motifs such as autocatalysis. The molecule model provides a solid foundation for supporting isotope labelling experiments with computational analysis</p>
<b>References</b>	<ol style="list-style-type: none"> <li>1. Andersen JL, Flamm C, Merkle D, Stadler PF (2018) Rule Composition in Graph Transformation Models of Chemical Reactions. <i>MATCH</i> 80(3):661-704.</li> <li>2. Fagerberg R, Flamm C, Kianian R, Merkle D, Stadler PF (2018) Finding the K Best Synthesis Plans. <i>J Cheminfo</i> 10:19 doi:10.1186/s13321-018-0273-z</li> <li>3. Andersen JL, Flamm C, Daniel Merkle D, Stadler PF (2017) An Intermediate Level of Abstraction for Computational Systems Chemistry. <i>Royal Soc Phil Trans A</i> 375(2109):20160354 doi:10.1098/rsta.2016.0354</li> <li>4. Andersen JL, Flamm C, Merkle D, Stadler PF (2016) A Software Package for Chemically Inspired Graph Transformation. <i>LNCS</i> 9761:73-88 doi:10.1007/978-3-319-40530-8-5</li> <li>5. Andersen JL, Flamm C, Merkle D, Stadler PF (2014) Generic Strategies for Chemical Space Exploration. <i>Int J Comp Biol Drug Des</i> 7(2/3):225-258 doi:10.1504/IJCBDD.2014.061649</li> <li>6. Mann M, Nahar F, Schnorr N, Backofen R, Stadler PF, Flamm C (2014) Atom mapping with constraint programming. <i>Algo Mol Biol</i> 9:23 doi:10.1186/s13015-014-0023-3</li> </ol>

<b>Abstract Title</b>	Dating the Origins of Microbial Sulfate Reduction
<b>Presenting Author</b>	Fournier G
<b>Institution/Affiliation</b>	MIT
<b>E-mail</b>	g4nier@mit.edu
<b>Additional Authors</b>	Momper L, Magnabosco C
<b>Abstract Text</b>	<p>Microbial reduction of sulfate is a major component of modern biogeochemical cycling. The abundance of sulfate in these microbial systems has varied over Earth's history, and is closely tied to the oxidative weathering of continental crust that has operated since the initial rise of atmospheric oxygen ~2.3 Ga. When did these metabolic pathways first evolve, and when did major microbial lineages involved in sulfate cycling become established? Using time calibrated phylogenies of conserved proteins within microbial groups, as well as phylogenies of proteins specifically involved in sulfite reduction, the last step of dissimilatory sulfate reduction (DsrA/DsrB), we reconstruct the early history of sulfate reduction. These age estimates are consistent with a history of the sulfur cycle in which the genes for dissimilatory sulfate reduction arose early in Earth's history, and horizontally transferred to many microbial clades in the Late Archean, possibly due to the increased availability of oxidized sulfur following the evolution of oxygenic photosynthesis. By the time of the GOE, the major extant groups of sulfate reducing microbes had been established, with an evolutionary pattern predominated by vertical inheritance rather than gene transfer. However, the diversification of the most abundant groups of sulfate reducers, including Desulfovibrionales within Deltaproteobacteria, occurred during the Paleoproterozoic.</p> <p>We suggest this was likely in response to the increased availability of weathering-derived sulfate within sediments as a consequence of the persistent presence of atmospheric oxygen. Molecular clocks informed by relative age constraints from horizontal gene transfers can greatly improve our understanding of the co-evolution of microbial and biogeochemical systems in Deep Time.</p>

<b>Abstract Title</b>	Data-mined ion substitutions in crystals: Reassessment of Goldschmidt's rules of ion substitution
<b>Presenting Author</b>	Olivier C. Gagné
<b>Institution/Affiliation</b>	Geophysical Laboratory, Carnegie Institution for Science (visiting investigator)
<b>E-mail</b>	olivier.c.gagne@gmail.com
<b>Additional Authors</b>	Hazen, RM
<b>Abstract Text</b>	<p>The basic rules of crystal chemistry were outlined over 90 years ago by Goldschmidt and Pauling. While some of Pauling's rules have since been the subject of further study and developments (e.g., the quantification of Pauling's electrostatic rule, leading to the development of the bond-valence model), Goldschmidt's seminal rules of ion substitution in crystals have received far less scrutiny aside from a minor amendment by Ringwood in the 1950s regarding differences in electronegativity. With today's unprecedented access to vast amounts of crystal-structure data, it is relatively easy to not only verify such results, but to increase their precision and enhance the scope of their application. In such an effort, we have analyzed over 13,000 instances of ion substitution in inorganic crystals (6,000+ instances for unique combinations of crystal structure and crystallographic site) using the Inorganic Crystal Structure Database (ICSD). We use these data to quantify the factors that affect ion substitution in oxide and oxysalt crystals, beyond those proposed by Goldschmidt and Ringwood, by cross-referencing (1) frequency of occurrence, and (2) site-occupancy ratio of the substitution ions to a variety of crystal chemical data (e.g., differences in ionic radius, bond strength, distortion value, etc.). Stepwise multiple-regression analysis shows that the parameters proposed by Goldschmidt and Ringwood (i.e., differences in charge, size, and electronegativity of the substituting ions) are not sufficient to accurately model the frequency for which ions are observed to substitute for one another.</p>

<b>Abstract Title</b>	The Mineral Evolution Database and data model derived visualizations
<b>Presenting Author</b>	Joshua J. Golden
<b>Institution/Affiliation</b>	University of Arizona
<b>E-mail</b>	jgolden@email.arizona.edu
<b>Additional Authors</b>	Pires, A.J., Rolf, J., Hystad, G., Morrison, S.M., Ma, X., Hummer, D., Liu, C., Eleish, A., Prabhu, A., Runyon, S., Downs, R.T., Hazen, R.M.
<b>Abstract Text</b>	<p>The Mineral Evolution Database (MED) was designed to discover the oldest recorded age for each mineral species using data found in pre-existing literature and databases. This is accomplished using a data model to establish connections between mindat's localities (<a href="https://www.mindat.org/">https://www.mindat.org/</a>), the International Mineralogical Association's list of approved species (<a href="http://rruff.info/ima/">http://rruff.info/ima/</a>), and the ages with additional contextual data found in the literature. The connections identified and recorded in the MED lead to new opportunities for data visualization for both mineralogy and geology as a whole.</p> <p>The current model contains 285,669 localities from mindat and 5,327 approved mineral species. Currently, age data sources include primarily geologic publications and USGS Mineral Resource data (<a href="https://mrdata.usgs.gov/">https://mrdata.usgs.gov/</a>), and are considered by the model if they match a mindat locality and if mindat lists the minerals at the locality. From 7,727 age records that satisfy those conditions, contextual information is used in the model to define age associations between localities, minerals, and elements; generating an age for nearly 22% of the 776,509 mineral occurrences.</p> <p>Age associations derived from the data model are stored in a MySQL database, allowing data to be filtered by any combination of the following criteria depending on desired purpose: locality, mineral chemistry, one or more minerals, or age. Data from the model is displayed as both web pages and is available as a series of csv files specific to minerals, localities, and ages.</p> <p>Derived from the data model, various visualizations have been created showing the age distribution of minerals and elements through time, element and mineral coexistence, oxidation state through time, and has allowed for the prediction of undiscovered mineral species. Age data that is fed to the model can be accessed on the Open Data Repository (<a href="https://odr.io/MEDAges">https://odr.io/MEDAges</a>).</p>

<b>Abstract Title</b>	Random Forests Classification of Pyrite: A Potential New Biosignature
<b>Presenting Author</b>	Daniel D Gregory
<b>Institution/Affiliation</b>	University of California, Riverside
<b>E-mail</b>	dgregory@ucr.edu
<b>Additional Authors</b>	Cracknell MJ, Figueroa MC, Large RR, Lyons TW
<b>Abstract Text</b>	<p>The search for biosignatures is fraught with many potential complications that can frustrate efforts to identify signs of past life. These complications include, but are not limited to, contamination during sampling, metamorphic overprints altering the biosignatures originally present in the rock, and difficulty in conclusively distinguishing between biotic structures (fossils) and abiotic structures that look similar. To help overcome these problems, we propose utilizing the trace element content of the mineral pyrite coupled with the classification algorithm Random Forests to create a new, novel biosignature. In sedimentary settings, pyrite forms readily when sulfate reducing bacteria produce sulfide that can bind with iron to form pyrite (FeS<sub>2</sub>). In doing so, the pyrite is stamped with a distinctive trace element fingerprint diagnostic of pyrite formed in this environment. Pyrite trace element chemistry has several advantages as a biosignature. It is readily available, and occurs in many different rock types. Its trace element chemistry is preserved up to mid-greenschist facies, past the P-T conditions that would reset many organic chemistry based biosignatures. Similarly, because of pyrite's chemical stability, contamination during sample collection is extremely unlikely to result in a false positive in the identification of a biosignature.</p> <p>That said, pyrite geochemistry is complicated and the range of trace element abundance in pyrite from different sources often overlap. To thoroughly characterize the geochemical fingerprint of pyrite, we need to consider a large suite trace elements simultaneously. This is done by using Random Forests, a supervised classification algorithm with proven effectiveness in earth sciences. In this study we selected nearly 4000 pyrite analyses from sedimentary (biogenic) pyrite and 6 different hydrothermal sources. The initial testing was promising, with the sedimentary pyrite being correctly identified 98% of the time and the different varieties of hydrothermal pyrite being correctly identified 80-95% of the time.</p>

<b>Abstract Title</b>	Data mining for critical minerals
<b>Presenting Author</b>	Jane M. Hammarstrom
<b>Institution/Affiliation</b>	US Geological Survey
<b>E-mail</b>	jhammars@usgs.gov
<b>Additional Authors</b>	Gilpin R. Robinson
<b>Abstract Text</b>	<p>Different types of mineral deposit have characteristic mineral associations that can be used to classify mineral localities by their likely mineral deposit types and to identify regions that may contain new resources. Mineral databases provide information for assessing mineral associations, mineral deposit types, and localities for evaluating known and potential sources of future supplies of rare or scarce mineral commodities. Databases such as MINDAT, the Mineral Resources Data System (MRDS), and compilations of data for individual types of mineral deposits can be used in combination as tools to explore for potential new sources of mineral resources. MINDAT, the largest mineral database available, contains worldwide data on minerals, mineral localities, references, and other mineralogical information. The site contains nearly 300,000 localities; many localities are mine sites of known deposit type. The elements germanium (Ge), gallium (Ga), and indium (In) are critical for solar cells, fiber optics, and other applications and are principally recovered as byproducts of mining other commodities. Very few (&lt;100) natural minerals contain any of these elements as primary constituents and many of those minerals are known from only 1 or 2 localities worldwide. These elements substitute in the mineral in sphalerite (ZnS), which is the primary zinc ore mineral. We compiled a mineralogy database for 1,400 zinc-bearing mineral deposits classified by deposit type to look for associations of minerals that contain these elements in 10 different types of zinc deposits. Preliminary results show that while only 7% of the 1,400 sites report Ge, about half of the epithermal-type deposits have Ge in the form of the silver-bearing Ge mineral argyrodite. We hope that concepts and methods developed in big data studies may help us explore mineral assemblages more efficiently so that we can identify sites that might represent new sources for these and other critical mineral commodities.</p>
<b>References</b>	<p>MINDAT, Mindat.org run by the Hudson Institute of Mineralogy, <a href="https://www.mindat.org">https://www.mindat.org</a></p> <p>Mineral Resources Data System (MRDS), Mineral Resources Online Spatial Data, U.S. Geological Survey, <a href="https://mrdata.usgs.gov/">https://mrdata.usgs.gov/</a></p>

<b>Abstract Title</b>	Late Archean weathering of phosphorus: implications for Earth's early P cycle
<b>Presenting Author</b>	Jihua Hao
<b>Institution/Affiliation</b>	Université Claude Bernard Lyon1 & Ens-Lyon
<b>E-mail</b>	haojihua@gmail.com
<b>Additional Authors</b>	Dimitri A. Sverjensky, Robert M. Hazen, and Isabelle Daniel
<b>Abstract Text</b>	<p>Phosphorus is an essential element for life and its bio-availability might limit biological activities on the early Earth. Geological records such as Banded Iron Formations and black shales have been used to reconstruct the levels of phosphorus in ancient seawater. However, interpretation of these records is complicated by various environmental factors affecting P-deposition, such as silica and cations, and the proxies are not ideal to represent the coastal environment where biological activities flourish.</p> <p>This study focuses on the continental weathering and riverine transport of phosphorus, which is the largest source in the modern P cycle. We firstly compiled the P concentrations in paleosols and calculated their enrichment factor (<math>\epsilon^{\text{P}} \text{P/Al}</math> or <math>\epsilon^{\text{P}} \text{P/Ti}</math>) in comparison with parent rocks. We found that P was depleted up to about 100% in the Archean paleosols (average depletion as 50%). This result is consistent with previous modelling studies, which showed solubility of P in the Archean river water 1-2 orders higher than modern (Hao et al., 2017). Moreover, the Archean surface waters were more acidic than modern (Hao et al., 2017), implying a much quicker weathering of P-hosted minerals like apatite. Last, iron is mobile as Fe(II) during Archean weathering due to the moderately reducing atmosphere, therefore allowing transport of substantial P as free phosphate ions, which are reactive and biologically available unlike Fe-bound P in the modern river water. Considering that continental crust grew steadily and reached more than 50% modern volume during the Archean (Taylor &amp; McLennan, 2009), flux of riverine P in the late Archean might be comparable or more pronounced than the modern Earth. High riverine transport of P favors biological activities, including oxygenic photosynthesis in the late Archean and possibly contribute to the irreversible oxygenation of atmosphere.</p>
<b>References</b>	<p>Hao, Jihua, Dimitri A. Sverjensky, and Robert M. Hazen. "A model for late Archean chemical weathering and world average river water." <i>Earth and Planetary Science Letters</i> 457 (2017): 191-203.</p> <p>Taylor, S. Ross, and Scott McLennan. <i>Planetary crusts: their composition, origin and evolution</i>. Vol. 10. Cambridge University Press, 2009.</p>

<b>Abstract Title</b>	Sequencing data discovery with MetaSeek: enabling data-driven discovery linking the biosphere and geosphere
<b>Presenting Author</b>	Adrienne Hoarfrost
<b>Institution/Affiliation</b>	University of North Carolina at Chapel Hill
<b>E-mail</b>	adrienne.l.hoarfrost@unc.edu
<b>Additional Authors</b>	Brown N, Brown CT, Arnosti C
<b>Abstract Text</b>	<p>Microbial communities and the geosphere have coevolved to shape the geochemical conditions on Earth in the modern day and throughout geological history. The relationships between the biosphere and geosphere are complex and nonlinear, and the greatest opportunities for understanding and modeling the interactions between the Earth system and microbial ecosystem function in Earth's past, present, and future rely on data-intensive investigations integrating large amounts of biological sequences. However, current tools to discover and curate sequencing data into large-scale integrated datasets are challenging and time-consuming, with sequencing data scattered across multiple repositories and metadata that is not easily or comprehensively searchable. MetaSeek (<a href="http://www.metaseek.cloud">www.metaseek.cloud</a>) is a sequencing data discovery tool that integrates sequencing metadata from the major data repositories, allowing the user to flexibly search and filter on datasets in a lightweight application with an intuitive, easy-to-use web-based interface. Users can save and share curated datasets, while other users can browse these data integrations or use them as a jumping off point for their own curation. Missing and/or erroneous metadata are inferred automatically where possible. Once an integrated dataset has been curated, users can follow simple instructions to download their raw data and quickly begin their investigations. The MetaSeek database is also easily queryable via an API, further enabling users and facilitating integrations of MetaSeek with other data curation and analysis tools. MetaSeek is currently being used to integrate large collections of metagenomic data in order to model biogeochemical processes driven by microbial communities using deep learning. MetaSeek lowers the barriers to curation and integration of environmental sequencing data, clearing the path forward to illuminating the ecosystem-scale interactions between biological and abiotic processes.</p>

<b>Abstract Title</b>	The Analytics Pipeline: Data Processing and Preparation
<b>Presenting Author</b>	Fang Huang
<b>Institution/Affiliation</b>	Tetherless World Constellation, Rensselaer Polytechnic Institute
<b>E-mail</b>	huangf5@rpi.edu
<b>Additional Authors</b>	Zhong H., Eleish A., Prabhu A., Pan F., Ma X., Hao J., Fox P.A., and the Keck DTDI Team
<b>Abstract Text</b>	<p>Due to the development of experimental and analytic equipment, large amounts of scientific data are being produced all over the world. The rapidly increasing volume and variety of geoscience-related data give researchers opportunities to answer scientific questions that are hard to solve using traditional methods. However, most of the time, raw data are not directly usable because of their hierarchical structures, missing values, inaccuracy, duplication, and so on. Also the data format needs to be adapted to different analytic techniques. It is often said that 80% time of data analytics is spent on data cleaning and preparing. Therefore, data processing is a crucial part of the data analytics pipeline, and iterative processing is needed when new problems emerge or secondary data are generated, throughout the whole analytics procedure.</p> <p>This poster will mainly focus on data processing and preparation. We start by introducing the principles for cleaning datasets, followed by some quick data exploration methods. Then we demonstrate some ways of dealing with various problems, such as missing values or duplication, in the datasets. The last part of the presentation show applications in real geoscientific datasets. Following good data processing practices can save time by avoiding repetition, and get more accurate results in analyses to help solve scientific problems.</p>

<b>Abstract Title</b>	Ecology and Evolution of Manganese Minerals: Implications for the Redox History of Earth and Life
<b>Presenting Author</b>	Daniel Hummer
<b>Institution/Affiliation</b>	Southern Illinois University
<b>E-mail</b>	daniel.hummer@siu.edu
<b>Additional Authors</b>	Hazen RM, Eleish A, Liu C, Morrison SM, Downs RT, Golden JJ, Pires A, Hystad G, Meyer MB
<b>Abstract Text</b>	<p>Manganese is a widespread redox sensitive transition element with three naturally occurring oxidation states and 560 mineral species, making it a useful proxy for the redox state of Earth's crust. Network analysis of Mn minerals using large mineral databases (mindat.org), which graphically displays the coexistence of mineral species, reveals three distinct clusters: (1) a central cluster composed of primary Mn<sup>2+</sup> silicates, oxides, and carbonates formed in igneous intrusions, hydrothermal ores, and skarns, together with secondary Mn<sup>3+</sup> and Mn<sup>4+</sup> oxidative weathering products, (2) a side cluster of predominantly Mn<sup>2+</sup> phosphates and oxides arising in granitic pegmatites, and (3) a side cluster of predominantly Mn<sup>2+</sup> silicates arising in agpaitic intrusives. Clusters 2 and 3 display some co-occurrence with minerals in cluster 1, but not with each other.</p> <p>Ages of first appearance reveal that common, primary igneous and metamorphic minerals of clusters 1 and 2 appear early in the geologic record, at or before 2.0 Ga. Oxidative weathering products in cluster 1 predominantly appear after photosystem-II initiated the Great Oxidation Event at ~ 2.3 Ga. Minerals of cluster 3 are geologically recent, appearing in the last 1.2 Ga as a result of further reworking and differentiation of the crust. The increasing prevalence of oxidized Mn minerals during the last 1 billion years raised the average oxidation state of crustal Mn in a pattern mirroring reconstructions of atmospheric oxygen, documenting the oxidation of Earth's crust in response to photosynthetic oxygen.</p>

<b>Abstract Title</b>	A Bayesian approach to estimating Earth's undiscovered, mineralogical diversity
<b>Presenting Author</b>	Grethe Hystad
<b>Institution/Affiliation</b>	Purdue University Northwest
<b>E-mail</b>	ghystad@pnw.edu
<b>Additional Authors</b>	Eleish A, Hazen RM, Morrison SM, Downs RT
<b>Abstract Text</b>	A Bayesian approach is introduced to estimate the total number of mineral species in Earth's crust. Markov chain Monte Carlo (MCMC) simulations are used to generate samples from a posterior distribution of the model parameters such that estimates and inference are directly obtained. Species accumulation curves are constructed and employed to estimate the population size as a function of sampling size. The Poisson lognormal distribution is found to provide the best fit to the mineral species frequency spectrum. Finally, the population size estimates obtained by Bayesian methods are compared to the empirical Bayes estimates.
<b>References</b>	Hystad G, Eleish A, Hazen RM, Morrison SM, Downs RT (2018) Bayesian Estimation of Earth's Undiscovered Mineralogical Diversity using Noninformative Priors, Math Geosci, under review

<b>Abstract Title</b>	Steps Toward Improved Integration, Search, and Analysis of Heterogeneous Data in the Astrobiology Habitable Environments Database
<b>Presenting Author</b>	Richard M. Keller
<b>Institution/Affiliation</b>	NASA Ames Research Center
<b>E-mail</b>	rich.keller@nasa.gov
<b>Additional Authors</b>	Detweiler AM, Lafuente Valverde B, Blake DF, Bristow TF, Cooper GW, Dateo CE, Des Marais DJ, Jahnke LL, Kubo MD, Parenteau N, Prufert-Bebout LE, and Stone N
<b>Abstract Text</b>	<p>The Astrobiology Habitable Environments Database (AHED) is a new data system being developed as a long-term, open-access repository for astrobiology data*. AHED is intended to store user-contributed results from NASA or externally-funded research in astrobiology, and to encourage sharing and synergy within the astrobiology community. However, the interdisciplinary nature of astrobiology presents some challenges to data management, integration, and analysis within AHED. In some disciplines (e.g., genomics), open databases thrive because the contributed products are fairly uniform and standardized (e.g., sequence data). In astrobiology, each investigation produces a unique set of data products; this makes it difficult to search across datasets to find similar data, or to combine results from separate investigations.</p> <p>With AHED, we are taking steps to ensure there is adequate metadata – both at the dataset and record levels – to facilitate search, integration, and analysis. At the dataset level, we are developing a new metadata standard for describing astrobiology datasets, with detailed information about content, funding source, and scientific relevance, along with a set of topical keywords for characterizing datasets. At the record level, we are encouraging users to provide more structured content and finer-grained metadata. In many user-contributed science data repositories, few restrictions are placed on the uploaded data format, and minimal or no record-level metadata is required; thus users are unburdened with respect to data preparation. The tradeoff is that deep integration and search across datasets is almost impossible without standardized structures and metadata. Although AHED users are free to upload minimally-described datasets, they will be encouraged to use database authoring tools (supplied by the underlying platform – Data Publisher**) plus a set of customizable astrobiology-specific templates to help structure their data and provide standardized metadata. In reward for their extra effort, AHED will be able to deliver enhanced search, discovery, and analysis capabilities.</p>
<b>References</b>	<p>*Lafuente B, Bristow T, Stone N, Pires A, Keller RM, Downs RT, Blake D, Fonda M (2017) Data Sharing in Astrobiology: The Astrobiology Habitable Environments Database (AHED), 48th Lunar and Planetary Science Conference, March 2017, <a href="http://www.lpi.usra.edu/meetings/lpsc2017/pdf/2943.pdf">http://www.lpi.usra.edu/meetings/lpsc2017/pdf/2943.pdf</a></p> <p>**Stone N, Lafuente B, Bristow T, Keller RM, Downs RT, Blake D, Fonda M, Pires (2016) Data publication and interoperability for long tail researchers via the Open Data Repository's (ODR) Data Publisher, AGU Fall Meeting, December 2016, <a href="https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/176555_28">https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/176555_28</a></p>

<b>Abstract Title</b>	Center for Life Detection Science
<b>Presenting Author</b>	Barbara Lafuente
<b>Institution/Affiliation</b>	NASA Ames Research Center
<b>E-mail</b>	barbara.lafuentevalverde@nasa.gov
<b>Additional Authors</b>	Hays L, Hoehler T, Pohorille A, CLD team
<b>Abstract Text</b>	<p>NASA programs and missions have begun to include the search for life as a prominent objective. To provide the scientific underpinning and context that would allow this objective to be pursued in a rigorous and informed fashion, the Center for Life Detection Science (CLDS) is developing a web-based platform in which to compile, organize, and curate a “living” repository of information and community dialog relating to life detection science. The tool is intended to support and provide connectivity between the program planning, instrument development, and life detection science communities.</p> <p>The CLDS web tool broadly encompasses two content areas: one devoted to exploring potential biosignatures ("features") independently of the instruments that may be used to seek them; and, separately, an ‘Instrument Forum’ that explores heritage and emerging approaches that target specific features. Adopting the basic approach represented in the “Ladder of Life Detection”[1], features are considered through reference to a standard set of attributes ("criteria") that collectively address the potential utility of seeking that feature as part of a life detection investigation. The feature/criteria organizing scheme is used to convey information at three levels: (i) the Comparison Matrix - a dynamic, filterable table used to compare multiple features at a highly distilled level; (ii) the Feature Summary – a narrative document (one per feature) that conveys background relevant to a given feature and discusses individual criteria at an expanded level of detail; (iii) the Knowledge Base – a tool, based on the preexisting “Hypothesis Browser” software, that organizes information (published papers) and ongoing community dialog according to its bearing on a specific criterion for a specific feature.</p> <p>The CLDS tool is envisioned to be an evolving informational resource in which content is continually updated through active management that identifies, vets, and incorporates relevant new results and dialog relating to individual features.</p>
<b>References</b>	[1] Neveu M, Hays LE, Voytek MA, New MH, Schulte MD. The Ladder of Life Detection. Astrobiology, accepted.

<b>Abstract Title</b>	Evidence from Ore Deposits indicates High Levels of Oxygen in the Proterozoic Ocean-Atmosphere System.
<b>Presenting Author</b>	Ross Large
<b>Institution/Affiliation</b>	Centre for Ore Deposits and Earth Sciences, University of Tasmania
<b>E-mail</b>	ross.large@utas.edu.au
<b>Additional Authors</b>	Indrani Mukherjee, Daniel D Gregory, Jeffrey Steadman and Sebastien Meffre
<b>Abstract Text</b>	<p>The Proterozoic is the major period in Earth history when metals were concentrated in sedimentary basins to form large deposits of copper, uranium and zinc. Examples include the Central African Copper Belt, Athabasca uranium province in Canada, and the Proterozoic Zinc-Belt of northern Australia. The transport and concentration of metals in sedimentary basins is affected by a number of factors including redox state, pH and temperature of the ore fluids. Of these, redox state is the overriding factor that sets these deposit groups apart. Due to the very low levels of atmospheric oxygen in the Archean, basinal fluids were commonly reduced (<math>H_2S \gg SO_4^{2-}</math>) and capable of transporting large quantities of Fe and Au, leading to the great reserves of banded iron formations, such as in the Hamersley Basin, and gold in the Witwatersrand Basin; the two greatest global accumulations of iron and gold respectively. However following the GOE, and especially after sulfate became abundant in the oceans (post 1800 Ma), basin fluids were far more oxidized (<math>SO_4^{2-} \rightarrow H_2S</math>) and capable of transporting significant amounts of copper, uranium, zinc, lead and silver, but lesser amounts of iron and gold. This resulted in the massive sedimentary copper, sedimentary uranium and sedimentary zinc-lead-silver deposits (SEDEX) accumulated in Proterozoic sedimentary basins. Although certain geochemical proxies, such as Cr isotopes, suggests that oxygen in the ocean-atmosphere system was very low (<math>pO_2 &lt; 10^{-3}</math> PAL) during the Proterozoic, the evidence from the ore deposits is that sedimentary basin fluids (principally derived from seawater and meteoric fluids) were generally oxidized, and of similar redox character to Phanerozoic basin fluids. We therefore suggest that rather than a sharp decline in atmosphere oxygen to very low levels immediately following the GOE, atmosphere oxygen remained elevated, and may have declined slowly through the Proterozoic, possibly over the range 0.7 to 0.1 PAL.</p>

<b>Abstract Title</b>	Preservational Bias of Mineral and Ore Deposits during Assembly of Rodinia
<b>Presenting Author</b>	Chao Liu
<b>Institution/Affiliation</b>	Geophysical Lab, Carnegie Institution for Science
<b>E-mail</b>	cliu@carnegiescience.edu
<b>Additional Authors</b>	Runyon S., Knoll A., Hazen R.
<b>Abstract Text</b>	<p>Rodinian assembly (1.3 Ga - 0.9 Ga) apparently lacks significant development of many types of ore deposits and minerals that are frequently formed during assembly of other supercontinents, including orogenic gold, porphyry copper, and volcanic hosted massive sulfides, and many minerals (e.g., minerals bearing Au, Co, Se, S) associated with these deposits. One possible explanation for these depletions is that convergent margins formed during Rodinian assembly may have experienced enhanced erosion. To test this hypothesis, we compile and compare the formation conditions of ore deposits and minerals formed during assembly of Rodinia with that of Nuna. Preliminary results indicate that many of the preserved records of Rodnian ore deposits and minerals formed along extensional settings, including intraplate rifts and distensional regions during hiatus of periodic collisions otherwise formed within high-grade metamorphic terranes. In contrast, many of the ore deposits associated with Nuna are hosted in low-grade metamorphic terranes; such observations may suggest that enhanced erosion of Rodinian convergent margins has resulted in an apparent lack of mineralogical diversity.</p>

<b>Abstract Title</b>	Microbial effects on volatile fluxes across the Costa Rica convergent margin
<b>Presenting Author</b>	Karen G. Lloyd
<b>Institution/Affiliation</b>	University of Tennessee
<b>E-mail</b>	klloyd@utk.edu
<b>Additional Authors</b>	Barry P, Battaglia A, Beaudry P, Bini G, Fullerton K, de Moor M, Giovannelli D, González G, Hummer D, Iacovino K, Lopez T, Martinez M, Matamoros M, Miller H, Pratt K, Ramírez C, Schrenk M, Segura Y, Turner S
<b>Abstract Text</b>	<p>Current understanding of volatile fluxes in subduction arcs focuses on the main volcanic emissions, and largely ignores any chemical transformations driven by the subsurface biosphere in low temperature regions. However, few studies have integrated measurements of gas fluxes across a transect of a convergent margin with geochemical and microbiological measurements to elucidate the complex subsurface processes that may affect them. Our multidisciplinary team assayed concentrations and isotopic ratios of gasses as well as dissolved organic and inorganic carbon compounds in the outer forearc, forearc, arc, and backarc across the Costa Rica convergent margin. We also sequenced the 16S rRNA gene, which is a marker for microbial identities, and produced large metagenomic datasets of all the functional genes in these environments. Our preliminary results suggest that a substantial gas flux of carbon dioxide in the Costa Rican outer forearc and forearc may be masked by a complex interplay of calcite deposition and microbiological factors. Additionally, all samples contained clear evidence of subsurface microbiological input, suggesting that subsurface processing, both biotic and abiotic, influences carbon recycling and volatile fluxes across the Costa Rica convergent margin.</p>

<b>Abstract Title</b>	A Justification on the Need to Build a Machine-Readable Knowledge Base of Deep Time
<b>Presenting Author</b>	Xiaogang (Marshall) Ma
<b>Institution/Affiliation</b>	Department of Computer Science, University of Idaho
<b>E-mail</b>	max@uidaho.edu
<b>Additional Authors</b>	Fox P, Hazen R, Hummer D, Golden J, Downs R, Hystad G, Muscente AD
<b>Abstract Text</b>	<p>The co-evolution between the geosphere and biosphere is one of the fundamental questions in Earth science [1]. Exploring the positive and negative feedbacks between the living and non-living components in the Earth's history needs various subjects of datasets to be integrated [2,3]. In the past decades, many geoscience data facilities have been built and made open on the Web. However, the shortage of efficient methods for accessing and synthesizing multi-source datasets hamper the data-intensive co-evolution research. Geologic time is an essential topic in the co-evolving geosphere and biosphere, and can be used as a common reference to connect various parameters among the data silos [4]. There is a critical need to build a Web-based and machine-readable knowledge base of deep time to automate geoscience data synthesis and to support executable workflows for data-driven scientific discovery. This presentation will summarize the previous work of computer models for geologic time scale, analyze the heterogeneity among local, regional and global geologic time standards, design the workflow for building a smart knowledge base of deep time, and propose the services of the knowledge base that can address the needs of data synthesis in co-evolution study. In the United States, the big data and data science communities are now promoting a national effort of open knowledge network to enable meaningful and smart data services in the cyberinfrastructure [5,6]. The work in this research will help leverage community of practice to improve the interoperability among the established geoscience data facilities. The best practices of multi-source data synthesis will provide strong evidence to demonstrate why an open knowledge network is useful, as well as how to build the network in a cyberinfrastructure ecosystem.</p>

## References

1. National Research Council (2008). *Origin and Evolution of Earth: Research Questions for A Changing Planet*. National Academies Press, Washington, DC, 150 pp.
2. Hazen, RM (2014). Data-driven abductive discovery in mineralogy. *American Mineralogist*, 99 (11-12): 2165-2170.
3. Hazen, RM (2017). Chance, necessity and the origins of life: a physical sciences perspective. *Philosophical Transactions of the Royal Society A* 375 (2109): 20160353. DOI: 10.1098/rsta.2016.0353.
4. Wang, C, Ma, X, Chen, J (2018). Ontology-driven data integration and visualization for exploring regional geologic time and paleontological information, *Computers & Geosciences* 115: 12-19.
5. Baru, C (2018). Perspectives on Open Knowledge Networks, The First U.S. Semantic Technologies Symposium (US2TS), Dayton, OH, USA. Oral Presentation.
6. Guha, RV, Moore, A (2016). *The OKN White Paper: Open Knowledge Network: Creating the Semantic Information Infrastructure for the Future*. 7 pp. <http://ichs.ucsf.edu/wp-content/uploads/2017/08/OKN-White-Paper..docx> (accessed on April 12, 2018)

<b>Abstract Title</b>	The size and extend of the subsurface biosphere
<b>Presenting Author</b>	Cara Magnabosco
<b>Institution/Affiliation</b>	Flatiron Institute
<b>E-mail</b>	cmagnabosco@flatironinstitute.org
<b>Abstract Text</b>	<p>Since the 1920s, scientists have explored the deep biosphere through drill cores, wells, and deep mines. To date, the deepest biological samples were retrieved from 4.8 km below land surface at the Donghai Drill Site in China. By compiling a global database of subsurface cell counts, we are able to provide more precise estimates of the size and extent of the subsurface biosphere. Using Shiny, we are able provide a web interface for users to explore this dataset and identify trends.</p> <p><a href="https://caramagnabosco.shinyapps.io/SubsurfaceBiologicalStudies/">https://caramagnabosco.shinyapps.io/SubsurfaceBiologicalStudies/</a></p>

<b>Abstract Title</b>	The influence of plate tectonics and mantle convection on the deep carbon cycle
<b>Presenting Author</b>	Claire Mallard
<b>Institution/Affiliation</b>	The University of Sydney
<b>E-mail</b>	sabin.zahirovic@sydney.edu.au
<b>Additional Authors</b>	Zahirovic S, Merdith A, East M, Johansson L, Pall J, Müller RD
<b>Abstract Text</b>	<p>The exchange of carbon between deep and surface reservoirs on Earth is dominated by plate tectonic and mantle convective processes on geological timescales. Perturbations to atmospheric CO<sub>2</sub> is captured by a number of geological proxy records, but these include a combination of physical Earth processes and interactions with the biosphere and hydrosphere. One avenue to help untangle the relative contributions of these processes is to first better quantify the 'tectonic forcing' parameters that are contributing to the planetary deep carbon cycle. We present emerging technologies that couple plate reconstructions that extend to one billion years of Earth history with aspects of the deep carbon cycle in the open-source and cross-platform GPlates software (<a href="http://www.gplates.org">www.gplates.org</a>). We use pyGPlates Python libraries to model slab flux and subducting plate area through time, which has implications for the volume of sediment injected at subduction zones and the resulting volumes of CO<sub>2</sub> generated by arc volcanism. In addition, we track the eruption of continental Large Igneous Provinces (LIPs) that emit significant volumes of CO<sub>2</sub> into the atmosphere over short geological timeframes of usually less than several million years. However, an additional consideration is the transit of LIPs through the near-equatorial humid belt where silicate weathering causes drawdown of atmospheric CO<sub>2</sub> over longer timescales. Our digital community plate reconstructions and workflows have also been used to track the interaction between subduction zones and major carbonate platforms, which is likely to trigger episodes of significant crustal decarbonation. The resulting collection of time-series will be crucial inputs to emerging DCO community carbon box models that will help us better quantify the role of plate tectonics in driving the exchange of carbon between deep and shallow reservoirs and the resulting modulation of long-term global climate.</p>

<b>Abstract Title</b>	Fully-dynamic global models to explore the processes that control global tectonics
<b>Presenting Author</b>	Claire Mallard
<b>Institution/Affiliation</b>	The University of Sydney
<b>E-mail</b>	claire.mallard@sydney.edu.au
<b>Additional Authors</b>	Arnould M, Coltice N, Zahirovic S
<b>Abstract Text</b>	<p>Plate tectonics theory describes first order surface motions at the surface of the Earth. Since the 1980s, significant progress has been made to reconstruct global tectonics (Lithgow-Bertelloni and Richards, 1998; Seton et al., 2012; Muller et al 2016), but increasing uncertainties in the geological record back in time makes it difficult to constrain plate motions before Pangea breakup. Although it is agreed upon that convection in the mantle drives the plates, the relationships between deep dynamics and surface tectonics are still largely unknown.</p> <p>Recent fully-dynamic global models (Tackley, 2008) are now able to generate large-scale ascending and descending mantle currents, as well as plate-like surface tectonics with narrow regions of localized deformation. These models self-consistently generate an expansion of the oceanic floor similar to that of the last 200 million years on Earth, and continental drift similar to what can be reconstructed with palaeomagnetism and provides access to a range of time-evolving parameters (temperature, velocities, viscosity).</p> <p>The development of new analytical tools (e.g. Cramer 2018; Mallard et al. 2017) and the constant evolution of the quality of mantle convection models allow us to improve our understanding of the link between mantle dynamics and surface tectonics, but also to target necessary improvements in existing convection models and tectonic reconstructions.</p> <p>The goal is to be able to qualitatively and quantitatively compare the results of convection computations, their plate boundaries and reorganisations, dynamic topography and mantle plume motions with surface motions, as reconstructed using the rules of plate tectonics and geological observations (Muller et al 2016). For instance, the definition and number of plates and past subduction zones can be biased, as well as the reconstruction of past dynamic topography and our understanding of mantle plume stability. Using the underlying forces within the lithosphere and mantle to further refine them is now possible.</p>

## References

- Lithgow-Bertelloni, C., Richards, M.A., (1998) The dynamics of Cenozoic and Mesozoic plate motions. *Rev. Geophys.* 36, 27–78.
- Seton, M., Müller, R., Zahirovic, S., Gaina, C., Torsvik, T., Shephard, G., Talsma, A., Gurnis, M., Turner, M., Maus, S., et al. (2012) Global continental and ocean basin reconstructions since 200 Ma. *Earth-Sci. Rev.* 113, 212–270.
- Müller, R. D., Seton, M., Zahirovic, S., Williams, S. E., Matthews, K. J., Wright, N. M., Shephard, G. E., Maloney, K., Barnett-Moore, N., Hosseinpour, M., et al. (2016) Ocean basin evolution and global-scale plate reorganization events since pangea breakup, *Annual Review of Earth and Planetary Sciences*, 44(1), 107–138.
- Tackley, P.J. (2008) Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the Yin-Yang grid. *Phys. Earth Planet. Inter.* 171, 7–18.
- Crameri, F. (2018) Geodynamic diagnostics, scientific visualisation and StagLab 3.0, *Geosci. Model Dev. Discuss* (In press)
- Mallard, C., Jacquet, B., & Coltice, N. (2017). ADOPT: A tool for automatic detection of tectonic plates at the surface of convection models. *Geochemistry, Geophysics, Geosystems*, 18(8), 3197-3208.

<b>Abstract Title</b>	Biological utilization of cobalt from the geosphere, and the expanding network of microbial metabolisms in the Archean eon
<b>Presenting Author</b>	Eli K. Moore
<b>Institution/Affiliation</b>	1. Environmental Biophysics and Molecular Ecology Program, Department of Earth, Ocean and Atmospheric Sciences, Rutgers University, New Brunswick, NJ, USA 2. National Agricultural Library, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD, USA 3. Department of Environmental Science, School of Earth & Environment, Rowan University, Glassboro, NJ, USA
<b>E-mail</b>	eli.can.moore@gmail.com
<b>Additional Authors</b>	J. Hao, A. Prabhu, H. Zhong, B. I. Jelen, M. Meyer, R. M. Hazen, P. G. Falkowski
<b>Abstract Text</b>	<p>The availability of different metals and substrates has changed over the course of Earth's history, which impacted the evolution of new metabolic pathways. Geochemical evidence can be used to reconstruct the metabolic network of microbial metabolism in the Archean eon, including the associated metallocofactors. Most metabolic pathways do not leave geochemical evidence to assign approximate dates of origin, resulting in uncertainty of when many metallocofactors were initially exploited by biology. Cobalt has been speculated to be a remnant of ancient metabolism, and is a unique metal in biology due to its importance to a wide range of organisms as the metal center of vitamin B12 (a.k.a. cobalamin, Cbl), despite its low abundance in the environment. We examined important chemical and geological factors that may have influenced the utilization of Co early in the evolution of life. Our mineral weathering model indicates dissolved Co was potentially more bioavailable in the Archean ocean under low S conditions than it is today. Mineral weathering, redox chemistry, Co complexation with nitrogen-containing organics, and hydrothermal environments were crucial in the incorporation of Co in primitive metabolic pathways. These chemical and geological characteristics of Co support its utilization in the Archean, and this approach can be used to inform the biological utilization of other trace metals in early forms of life.</p>

<b>Abstract Title</b>	Electron transfer pathway networks to probe deep-time protein evolution
<b>Presenting Author</b>	Vikas Nanda
<b>Institution/Affiliation</b>	Rutgers University
<b>E-mail</b>	nanda@cabm.rutgers.edu
<b>Additional Authors</b>	Raanan H, Pike DH, Moore EK, Falkowski PG
<b>Abstract Text</b>	<p>There are no direct fossil records of the original proteins at the beginning of life. Phylogenetic approaches that infer ancestry from sequences and structures of extant proteins are of limited use over billions-year evolutionary time scales. Leveraging the structures of proteins containing transition-metal cofactors, we identify structural modules that comprise the diverse family of oxidoreductases critical for electron transfer reactions, and patterns of connectivity between minimal structural elements. The result is a bipartite network that is potentially rich in information about evolutionary and functional relationships. We are exploring the role of such networks for assessing evolutionary homology versus analogy of distantly related proteins.</p>
<b>References</b>	Raanan H, Pike DH, Moore EK, Falkowski PG, Nanda V (2018) Modular origins of biological electron transfer. PNAS 115:1280

<b>Abstract Title</b>	Predicting multi-component mineral compositions with label distribution learning
<b>Presenting Author</b>	Feifei Pan
<b>Institution/Affiliation</b>	Rensselaer Polytechnic Institute
<b>E-mail</b>	panf2@rpi.edu
<b>Additional Authors</b>	Shaunna M. Morrison, Anirudh Prabhu, Ahmed Eleish, Olivier C. Gagne, Peter Fox, Robert M. Downs, Robert M. Hazen and the Keck Deep-Time Data Infrastructure Team ( <a href="http://dtdi.carnegiescience.edu">http://dtdi.carnegiescience.edu</a> )
<b>Abstract Text</b>	<p>In many previous studies, refined unit-cell parameters are used to estimate mineral composition by studying the relationships between ionic radii and changes in the unit-cell dimensions or diffraction peak positions. However, due to the complexity of the multi-component system, existing work usually limits the number of chemical components to 3 or fewer.</p> <p>We propose to solve the problem by adapting machine learning techniques such as Label Distribution Learning (LDL). LDL is a novel framework for classification problems. Compared to the traditional single or multi-label classification, LDL can predict not only the label for each instance, but also a degree for each label which demonstrates by how much each label can represent the instance. Algorithms under the LDL framework have been widely applied to the computer vision research domain to solve problems such as age estimation, head pose estimation, and natural scene component identification. We have also successfully applied this framework to additive manufacturing research to predict failure types in the manufacturing procedure.</p> <p>In this poster, we present our work on extending the application of LDL to predict multi-component mineral compositions. We define our task as: (1) Adapt the LDL framework to small mineralogy datasets (30-200 data points in general); (2) Use crystallographic parameters of each mineral to predict its chemical components; (2) Estimate the amount for each chemical component. We modify multiple algorithms under the LDL framework to accommodate our problem setting and achieve significant results. We evaluate the performance using distance and similarity between label distributions as well as mean square error and we also compare the results to the performance of traditional machine learning methods. Experiments results demonstrate LDL has recognizable advantages.</p>
<b>References</b>	<p>X. Geng. Label Distribution Learning. IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE), 2016</p> <p>X. Geng, C. Yin, and Z.-H. Zhou. Facial Age Estimation by Learning from Label Distributions. IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI), 2013</p>

<b>Abstract Title</b>	Macrostrat: A Platform for Geological Data Integration and 4D Earth Crust Research
<b>Presenting Author</b>	Shanan E. Peters
<b>Institution/Affiliation</b>	University of Wisconsin-Madison
<b>E-mail</b>	peters@geology.wisc.edu
<b>Additional Authors</b>	Husson JM, Czaplewski J
<b>Abstract Text</b>	<p>Characterizing the lithology, age, and physical-chemical properties of rocks and sediments in the Earth's upper crust is necessary to fully assess energy, water, and mineral resources and to address many fundamental questions about the evolving Earth-life system. Here we describe Macrostrat (<a href="https://macrostrat.org">https://macrostrat.org</a>), a relational geospatial database and supporting cyberinfrastructure that is designed to enable quantitative spatial and geochronological analyses of the entire assemblage of surface and subsurface sedimentary, igneous, and metamorphic rocks. Macrostrat currently contains general, but also comprehensive summaries of the age and properties of 34,826 lithologically and chronologically defined geological units distributed across 1,518 regions in North and South America, the Caribbean, New Zealand, and the deep sea. Sample-derived data, including fossil occurrences in the Paleobiology Database, more than 180,000 geochemical and outcrop-derived measurements and 2.3 million geologic map units derived from over 200 map sources, are linked to specific rock/sediment units and/or lithologies within those units. Data within Macrostrat, particularly lithostratigraphic names, are automatically linked to publications and information within those publications via GeoDeepDive (<a href="https://geodeepdive.org">https://geodeepdive.org</a>) infrastructure. Macrostrat is primarily designed to enable quantitative summarization of the mass-age-lithology properties of the upper crust. However, the data and cyberinfrastructure developed as part of this project can be used as a platform in mobile and analytical applications that serve a wide range of end use cases. Expanding geographic coverage of chronostratigraphic columns globally, refining age models and including data to better constrain material and chemical properties of rock units is necessary in order to arrive at a more precise characterization of the upper crust and to test fundamental hypotheses about the deep time evolution of coupled Earth systems.</p>

<b>Abstract Title</b>	The Analytics Pipeline: Data Visualization and Exploratory Analysis
<b>Presenting Author</b>	Anirudh Prabhu
<b>Institution/Affiliation</b>	Rensselaer Polytechnic Institute
<b>E-mail</b>	prabha2@rpi.edu
<b>Additional Authors</b>	Eleish A, Zhong H, Huang F, Pan F, Fox P, Morrison SM, Moore E, Muscente AD, Hazen R and the DTDI Team.
<b>Abstract Text</b>	<p>The last 30 years have seen a revolution in the development and availability of big data resources in the geo- and biosciences , such as EarthChem, NavDat, mindat.org and Paleobiology Database. The increases in the volume, variety and velocity of data make it very important for geoscientists to the know methods and techniques to identify and “learn” patterns and trends hidden in the data.</p> <p>Data and Information analytics extend analyses (descriptive and predictive models to obtain knowledge from data) by using insight to recommend actions or to guide and communicate decision-making. Thus, analytics is not only concerned with individual analyses or analysis steps, but with an entire methodology.</p> <p>This poster uses real-world “geoscience use cases” to introduce the audience to the process of Data Modeling and Analysis. We use both supervised (predictive methods like Decision Trees, Random Forest, Support Vector Machines, Neural Networks etc) and unsupervised (Clustering Methods K-means, Partitioning around Medoids etc) models to better understand complex data from different domains and then visualize and examine the results in order to gain valuable insights and help answer important scientific questions.</p>

<b>Abstract Title</b>	Data-driven discovery targets the “where” of Earth’s future resources
<b>Presenting Author</b>	Robinson GR
<b>Institution/Affiliation</b>	Department of the Interior, U.S. Geological Survey, Reston, Virginia 20192
<b>E-mail</b>	grobins@usgs.gov
<b>Additional Authors</b>	Mars JC, Shapiro JL, Hammarstrom JM, Cohan JA
<b>Abstract Text</b>	<p>Data-driven analytical models coupled with large and growing mineral resource and other geologic databases will help us to better understand and forecast the occurrence and distribution of mineralizing systems and resources over space, time, processes, and discovery. These databases typically include information on known mineral deposits, their contained resources, mineral localities and mineral associations, and systematic classified landscape data on geologic features, chemistry, remote sensing, geophysics, and topography. Targeting “where” is termed mineral prospectivity mapping. The basic approach is to analyze the locations of mineral deposits in relation to ore-controlling features and landscape factors that are assumed to control or influence the spatial distribution of mineral deposits, and can be expressed as maps. The idea is similar to the ecological niche concept in biology. Many data-driven mineral prospectivity mapping methods have been developed, including weights-of-evidence, evidential belief functions, logistic regression, neural network techniques, random forest, and maximum entropy models. To be useful, these methods need to deal with small numbers of occurrence records, as mineral deposits are rare events. A variety of techniques, including Receiver Operating Characteristic (ROC) curves, prediction-area curves, and kappa scores, can be used to evaluate model performance and uncertainty.</p> <p>We will explore the use of mineral prospectivity mapping for intrusion-related hydrothermal mineral deposits using satellite-based multispectral sensor data (ASTER). Specifically, prospectivity mapping focused on hydrothermal alteration mineral assemblages (e.g., Argillic (clays), Phyllic (micas), Propylitic (epidote-chlorite), Carbonate, and Hydrothermal Silica (jasper)) that are discriminated based on characteristic mineral spectral features. The prospectivity modelling integrates the spatial density and association of the alteration mineral groups weighted in relation to known mineral deposits in a well-explored study area. The model results are applied to a broader region in the Southwest United States.</p>

<b>Abstract Title</b>	Global patterns of subsurface microbial diversity through deep time and space
<b>Presenting Author</b>	Joshua Ladau, Cara Magnabosco, Emil Ruff
<b>Institution/Affiliation</b>	University of Calgary
<b>E-mail</b>	emil.ruff@ucalgary.ca
<b>Additional Authors</b>	Colwell R., Gaidos E., Hoarfrost A., Lloyd K., Kiel Reese B., Rogers K., Sheik C., Sherwood Lollar B., Sogin M.
<b>Abstract Text</b>	<p>Marine and terrestrial subsurface sediments and rocks are the largest habitat for life on Earth. This deep biosphere harbors an estimated <math>10^{29}</math> microbial cells. These microorganisms influence global chemical cycles and may represent an essentially untapped archive of diversity and functions. Yet, their community structure, their global diversity, their biogeography and modes of dispersal are largely unknown. Here we compiled and analyzed the largest dataset of subsurface communities available to date, comprising 302 samples from 24 globally distributed marine and terrestrial subsurface ecosystems. Using state-of-the-art bioinformatics we discovered four major trends. 1. Marine subsurface ecosystems are more diverse than terrestrial subsurface ecosystems 2. The rate of microbial dispersal is higher in marine ecosystems. 3. None of the ecosystems was exhaustively sampled, leaving lots of biodiversity to be discovered. 4. The communities living in bedrock appear to correlate with tectonic plate movement over deep time. These findings highlight the longevity of these communities and have major implications for our understanding of microbial dispersal and community assembly over deep time and space.</p>

<b>Abstract Title</b>	U-bearing Mineral Chemistry and its Relation to Uranium Ore Deposit Types
<b>Presenting Author</b>	Simone E. Runyon
<b>Institution/Affiliation</b>	Carnegie Geophysical Lab / University of Wyoming
<b>E-mail</b>	srunyon@carnegiescience.edu
<b>Additional Authors</b>	Hall SM, Perry SN, Eleish A, Prabhu, A, Morrison SM, Liu C, Golden J, Pires A, Smith ML, Wendlandt RF, Zhong H, Fang H, Burns PC, Hazen RM, and the Keck Deep-Time Data Infrastructure Team ( <a href="http://dti.carnegiescience.edu">http://dti.carnegiescience.edu</a> )
<b>Abstract Text</b>	<p>Uranium is an important element in the production of nuclear energy. Uranium geochemistry is strongly governed by oxidation state and is highly mobile during oxidizing conditions as the uranyl ion, <math>UO_2^{2+}</math>. U(VI) can form a variety of minerals by complexing with carbonate, chloride, sulfate, phosphate, fluoride, hydroxyl, and/or silicate (e.g., Cuney, 2009). Uranium deposits form over a broad range of geologic conditions, up to 800°C and 5 -7 kbar (Cuney, 2009). The oldest known uranium deposits are ~3.1 Ga and uranium deposits are widespread after the Great Oxidation Event through the Quaternary (Cuney, 2010). Uranium deposits, therefore, can range drastically in age, host rock, tectonic setting, primary mineral assemblage, Pb content, and overprinting supergene assemblages. As a result of this complexity, there are multiple classification schemes used to describe uranium-bearing deposits, including classification schemes developed by the IAEA (International Atomic Energy Agency: Bruneton, 2014; IAEA, 2017), Cuney (2009), Skirrow et al. (2009), Dahlkamp (1993, 2009), and others. The IAEA classification scheme is largely based on host rocks or deposit morphology, whereas other classification schemes have aimed to propose a classification based on genetic processes (e.g., Cuney, 2009). By combining data from the IAEA (UDEPO database) and Mineral Evolution Database (MED), and adding information from primary literature, we can analyze the occurrence and commonality of U-bearing minerals and their corresponding deposit types. In this preliminary work, we investigate the distribution of U-bearing minerals between different deposit types to better understand genetic implications of these geochemical patterns.</p>
<b>References</b>	<p>Bruneton P (2014) IAEA classification of uranium deposits. IAEA Vienna URAM 2014.</p> <p>Cuney M (2009) The extreme diversity of uranium deposits. <i>Mineralium Deposita</i>, 44.</p> <p>Cuney M (2010) Evolution of uranium fractionation processes through time: Driving the secular variation of uranium deposit types. <i>Economic Geology</i>, 105.</p> <p>Dahlkamp FJ (1992) <i>Uranium ore deposits</i>. New York, Springer-Verlag.</p> <p>Dahlkamp FJ (2009) <i>Uranium deposits of the world USA and Latin America</i>. Berlin, Springer-Verlag.</p> <p>IAEA (2017) <i>Uranium 2016: resources, production and demand</i>. A joint report by the OECD Nuclear Energy Agency and the International Atomic Energy Agency, OECD, Paris.</p> <p>Skirrow RG, Jaireth S, Huston DL, Bastrakov EN, Schofield A, van der Weilen SE, Barnicoat AC (2009) <i>Uranium Mineral Systems: Processes, exploration criteria and a new deposit framework</i>. Australian Government Geoscience Australia Record 2009/20 GeoCat # 69124.</p>

<b>Abstract Title</b>	Universal scaling across biochemical networks on Earth
<b>Presenting Author</b>	Harrison B. Smith
<b>Institution/Affiliation</b>	Arizona State University
<b>E-mail</b>	hbs@asu.edu
<b>Additional Authors</b>	Kim H, Mathis C, Raymond J, Walker SI
<b>Abstract Text</b>	<p>The application of network science to biology has advanced our understanding of the metabolism of individual organisms and the organization of ecosystems but has scarcely been applied to life at a planetary scale [1-4]. To characterize planetary-scale biochemistry, we constructed biochemical networks using a global database of 28,146 annotated genomes and metagenomes, and 8,658 cataloged biochemical reactions. We uncover scaling laws governing biochemical diversity and network structure shared across levels of organization from individuals to ecosystems, to the biosphere as a whole. Comparing real biochemical networks to random chemical networks reveals the observed biological scaling is not solely a product of the biochemistry shared across life on Earth. Instead, it emerges due to how the global inventory of biochemical reactions is partitioned into individuals. We show the three domains of life are topologically distinguishable, with &gt; 80% accuracy in predicting evolutionary domain based on biochemical network size and average topology. Taken together our results point to a deeper level of organization in biochemical networks than what has been understood so far.</p>
<b>References</b>	<ol style="list-style-type: none"> <li>1. Barabási AL. &amp; Oltvai ZN (2004) Network biology: understanding the cell's functional organization. <i>Nat. Rev. Genet.</i> 5:101–113.</li> <li>2. Kitano H. (2002) Computational systems biology. <i>Nature</i> 420:206–210.</li> <li>3. Klitgord N &amp; Segrè D (2011) Ecosystems biology of microbial metabolism. <i>Curr. Opin. Biotechnol.</i> 22:541–546.</li> <li>4. Proulx SR, Promislow DEL, Phillips PC (2005) Network thinking in ecology and evolution. <i>Trends Ecol. Evol.</i> 20:345–353.</li> </ol>

<b>Abstract Title</b>	Earth and Space Sciences Data Are a World Heritage: Community Partnership to Develop Best Practices Across the Data Lifecycle to Advance Open and FAIR Data
<b>Presenting Author</b>	Shelley Stall
<b>Institution/Affiliation</b>	American Geophysical Union
<b>E-mail</b>	sstall@agu.org
<b>Additional Authors</b>	Yarmey L, Lehnert K, Robinson E, Hanson B, Parsons M, Wyborn L, Cutcher-Gershenfeld J, Nosek B
<b>Abstract Text</b>	<p>Open, accessible, and high-quality data and related products such as software are critical to the integrity of published research. They ensure transparency and support reproducibility and are necessary for accelerating the advancement of science.</p> <p>Scientific results are increasingly dependent on large complex data sets and models that transform these data, or data sets with difficult-to-acquire observations. Increasingly these are stored or made available separately from the actual publication.</p> <p>Even when data are saved, their curation is uneven and discovery and linking of data sets that should be allied is difficult or impossible. Data are often stored with publishers as PDF or other supplements without any metadata or in general repositories without any quality control or curation. Researchers are often not able to understand the data sets without contacting the author, if that is even possible.</p> <p>To address this critical need, the Laura and John Arnold Foundation has awarded a grant to a coalition of groups representing the international Earth and space science community, convened by the American Geophysical Union (AGU), to develop standards that will connect researchers, publishers, and data repositories in the Earth and space sciences to enable FAIR (findable, accessible, interoperable, and reusable) data on a large scale.</p> <p>This effort will build on the work of The Coalition on Publishing Data in the Earth and Space Sciences (COPDESS.org), ESIP, RDA, the scientific journals, and domain repositories to ensure that well documented data, preserved in a repository with community agreed-upon metadata, and supporting persistent identifiers becomes part of the expected research products submitted in support of each publication. It is expected that the broader community will play a key role in the recommended guidelines and approach.</p>
<b>References</b>	Stall, S., E. Robinson, L. Wyborn, L. R. Yarmey, M. A. Parsons, K. Lehnert, J. Cutcher-Gershenfeld, B. Nosek, and B. Hanson (2017), Enabling FAIR data across the Earth and space sciences, <i>Eos</i> , 98, <a href="https://doi.org/10.1029/2017EO088425">https://doi.org/10.1029/2017EO088425</a> . Published on 08 December 2017.

<b>Abstract Title</b>	Understanding Limitations of Information Content in X-ray Absorption and Emission Spectroscopy
<b>Presenting Author</b>	Richard C. Walroth
<b>Institution/Affiliation</b>	NASA Ames Research Center
<b>E-mail</b>	richard.c.walroth@nasa.gov
<b>Abstract Text</b>	<p>Bioinorganic chemists are constantly seeking novel methods to interrogate the electronic and geometric structure of metal centers. However, when developing novel investigative tools, such as novel spectroscopic techniques, it is important to rigorously interrogate the claims made for any given spectroscopic method, and ask if a given signature is unique, distinguishable from alternatives, and interpretable with minimal reliance on standards. Development of two methods in particular, valence to core X-ray emission spectroscopy (V2C-XES), and X-ray absorption spectroscopy (XAS), will be reported. Special attention will be given to the applicability of this work to data driven discovery. For V2C-XES, a library was assembled from spectra obtained for 12 diverse Cr complexes and used to calibrate density functional theory (DFT) calculations of V2C XES band energies. As the calculated spectra could readily reproduce experiment, the Cambridge Crystallography Database was used to obtain a broad set of complexes to further establish the utility of V2C-XES. Experimentally calibrated, DFT calculated V2C XES spectra of 90 Cr compounds were used to produce a quantitative spectrochemical series showing the V2C XES band energy ranges for ligands comprising 18 distinct classes. Substantial overlaps are detected in these ranges, which complicates the use of V2C XES to identify ligands in the coordination spheres of unknown Cr compounds. For XAS, a set of Cu complexes was used to calibrate TDDFT based spectral predictions. Monte-Carlo based fitting methods were developed to limit bias and extract energy/intensity values for features in the spectra. XAS was then used to narrow possible intermediates from computational studies of a Cu catalyzed alcohol oxidation reaction.</p>

<b>Abstract Title</b>	Revelations from marine sedimentary metagenomic signatures
<b>Presenting Author</b>	Lisa Warden
<b>Institution/Affiliation</b>	Carnegie Institute for Science
<b>E-mail</b>	lwarden@carnegiescience.edu
<b>Additional Authors</b>	Giovannelli, D., Bromberg, Y., Liu, C., Morrison, S. and Hazen, R.
<b>Abstract Text</b>	<p>Microorganisms in marine sediments constitute a distinct, pervasive and largely understudied ecosystem that are believed to make up perhaps one-third of the Earth's total biomass (Whitman et al., 1998) and influence large-scale geochemical cycles (D'hondt et al., 2002). The ability of marine microbes to thrive throughout the marine environment, from hydrothermal vents to methane seeps, is due to the diversity and adaptability encoded within their genetic sequences (Kennedy et al., 2010). Since it is widely accepted that less than 1% of microbes can be isolated using traditional culturing methods, metagenomics, the study of the pooled genetic complement of a given environmental sample (Handelsman, 2004), has been proposed to provide the most comprehensive and accurate view of the microbial world (Von Mering et al., 2007). In order to develop a better understanding of one of the largest reservoirs of microbial biomass on Earth and how the distribution of its microbial inhabitants are influenced by their environmental parameters, we collected published metagenomes from marine sediments in varying marine settings that were deposited in an assortment of online databases. After annotating the microbiomes' molecular functions using mi-faser (Zhu et al., 2017), we will examine the results along with any accompanying metadata that was obtained at the time the sample was collected with any mineral, chemical and geological data deposited in the EarthChem database with similar GPS coordinates and from similar environments as the metagenome samples. The application of network theory to determine co-occurrence patterns among microorganisms is a relatively new frontier in microbial ecology (Raes and Bork, 2008), and we hope that through its application we can illustrate relationships and patterns among the microbial community and their varying environments within marine sediments.</p>

## References

- D'hondt, S., Rutherford, S., & Spivack, A. J. (2002). Metabolic activity of subsurface life in deep-sea sediments. *Science*, 295(5562), 2067-2070.
- Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and molecular biology reviews*, 68(4), 669-685.
- Kennedy, J., Flemer, B., Jackson, S. A., Lejon, D. P., Morrissey, J. P., O'gara, F., & Dobson, A. D. (2010). Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism. *Marine drugs*, 8(3), 608-628.
- Raes, J., & Bork, P. (2008). Molecular eco-systems biology: towards an understanding of community function. *Nature Reviews Microbiology*, 6(9), 693.
- Von Mering, C., Hugenholtz, P., Raes, J., Tringe, S. G., Doerks, T., Jensen, L. J., ... & Bork, P. (2007). Quantitative phylogenetic assessment of microbial communities in diverse environments. *science*, 315(5815), 1126-1130.
- Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences*, 95(12), 6578-6583.
- Zhu, C., Miller, M., Marpaka, S., Vaysberg, P., Rühlemann, M.C., Wu, G., Heinsen, F.A., Tempel, M., Zhao, L., Lieb, W. and Franke, A., 2017. Functional sequencing read annotation for high precision microbiome analysis. *Nucleic acids research*.

<b>Abstract Title</b>	Ecological Marine Units as a Framework for Collaborative Data Science and Knowledge Discovery
<b>Presenting Author</b>	Dawn J. Wright
<b>Institution/Affiliation</b>	Environmental Systems Research Institute and Oregon State University
<b>E-mail</b>	dwright@esri.com
<b>Additional Authors</b>	Sayre, R., Breyer, S., Butler K.A., VanGaraafeiland, K., Goodin, K., Kavanaugh, M.T., Costello, M., Cressie, N., Basher, Z., Harris, P., and Guinotte, J.
<b>Abstract Text</b>	<p>We present a data-derived, ecosystem mapping approach for the global ocean as commissioned by the Group on Earth Observations (GEO) and as a contribution to the Marine Biodiversity Observation Network (MBON). These ecological marine units (EMUs) are comprised of a global point mesh framework, created from over 52 million points from NOAA's World Ocean Atlas with a spatial resolution of 1 by 1 degree (~27 x 27 km at the equator) at 44 varying depths and a temporal resolution that is currently decadal. Each point carries attributes of chemical and physical oceanographic structure (temperature, salinity, dissolved oxygen, nitrate, silicate, phosphate) as likely drivers of many marine ecosystem responses. A k-means statistical clustering algorithm revealed physically-distinct, relatively-homogenous, volumetric regions within the water column (the EMUs). Backwards stepwise discriminant analysis determined if all of six variables contributed significantly to the clustering, and a pseudo F-statistic gave us an optimum number of clusters worldwide at 37. A major intent of the EMUs is to support marine biodiversity conservation assessments, economic valuation studies of marine ecosystem goods and services, and studies of ocean acidification and other impacts. As such, they represent a rich geospatial accounting framework for these types of studies, as well as for scientific research on species distributions. To further benefit the community and facilitate collaborate knowledge building, data products are shared openly and interoperably via <a href="http://www.esri.com/ecological-marine-units">www.esri.com/ecological-marine-units</a>. Work is in progress to delineate EMUs at finer spatial and temporal resolutions and to include ocean currents and various biodiversity observations. A major aim is for the ocean science community members to move the research forward with higher-resolution data from their own field studies or areas of interest, with the original EMU project team assisting with GIS implementation (especially via a new online discussion forum), and hosting of additional data products as needed.</p>
<b>References</b>	Sayre R, et al. (2017) A three-dimensional mapping of the ocean based on environmental data. <i>Oceanography</i> 30(1):90-103.

<b>Abstract Title</b>	Phosphorous fluxes over supercontinent cycles: perspectives from the Phosphorous Mineral Evolution Database (P-MED)
<b>Presenting Author</b>	Dr Aubrey Zerkle
<b>Institution/Affiliation</b>	University of St Andrews
<b>E-mail</b>	az29@st-andrews.ac.uk
<b>Additional Authors</b>	Walton CR, Hazen R, Golden, J
<b>Abstract Text</b>	<p>Mineral diversity may have increased on Earth over time, both in response to and driving major biogeochemical change. The mineral content of terrestrial rocks provides a detailed record of Earth system chemistry that might be usefully quantified with a “big data” approach (Hazen et al, 2014). The surficial cycling of phosphorus – an often-limiting nutrient for life – is controlled by the weathering of unreactive phosphate minerals (Pasek et al, 2014). Hence, changes in phosphate mineral diversity and abundance may have played a primary role in catalysing/limiting biogeochemistry. However, phosphorous has only one stable isotope – this has limited progress in mapping out its cycling over deep time (Reinhard et al, 2017). The Phosphorous Mineral Evolution Database (P-MED) uses spatial data on mineral occurrence and temporal data on their formation ages as a new way of approaching the issue. In order to do this correctly, however, it is important to take into account possible issues with preservation bias, sampling bias and paragenesis. There are already indications in the database that P mineralisation follows a cyclic pattern over Earth history that is coincident with supercontinent cycles. This bolsters the idea that tectonic processes have ultimately governed P supply to the Earth’s surface over time. Exospheric feedbacks will also have played a significant role in this regard, however, as (e.g.) weathering fluxes are probably dependant on atmospheric chemistry as well as the types of P-bearing phases present in the crustal mineral inventory at any given point. Feedbacks between tectonic processes and surficial biochemistry are heavily implied when thinking about P cycling in deep time from this perspective, yet the details remain elusive. It will take a co-ordinated effort bringing together mineral evolution with research into preservation proxies and weathering trends to understand the links between P, tectonics, Earth surface chemistry and life.</p>
<b>References</b>	<p>Pasek, M.A., Sampson, J.M. &amp; Atlas, Z., 2014. Redox chemistry in the phosphorus biogeochemical cycle. <i>Proceedings of the National Academy of Sciences</i>, 111(43), p.15468 LP – 15473. Available at: <a href="http://www.pnas.org/content/111/43/15468.abstract">http://www.pnas.org/content/111/43/15468.abstract</a>.</p> <p>Hazen, R. M. et al., 2014. Mineral Evolution: Episodic Metallogenesis, the Supercontinent Cycle, and the Coevolving Geosphere and Biosphere, Society of Economic Geologists, Inc.</p> <p>Reinhard, C.T. et al., 2016. Evolution of the global phosphorus cycle. <i>Nature</i>, 541, p.386. Available at: <a href="http://dx.doi.org/10.1038/nature20772">http://dx.doi.org/10.1038/nature20772</a>.</p>

<b>Abstract Title</b>	The Analytics Pipeline: Data Acquisition in the Information Era
<b>Presenting Author</b>	Hao Zhong
<b>Institution/Affiliation</b>	Rensselaer Polytechnic Institute
<b>E-mail</b>	zhongh3@rpi.edu
<b>Additional Authors</b>	Huang F, Prabhu A, Pan F, Eleish A, Ma X, Fox PA, and the Keck Deep-Time Data Infrastructure Team
<b>Abstract Text</b>	<p>Over the past century, enormous amount of data has been produced, archived and published across the geoscience community. Development of new experimental devices, analytical tools, as well as scientific methods have been the driving forces underneath the accelerated increase in quantity and improvement in quality of geoscience-related data. In recent decades, such exponential growth of data has uncovered new, data-intensive approaches towards research questions that were once before unsolvable in the absence of enough data, and even inspired many new discoveries. As data science becomes instrumental in geoscience research, understanding and adoption of suitable data science practices cannot be emphasized enough in order to maximize the utilization of existing data and unleash the full potential of data-driven discoveries in geoscience.</p> <p>Data acquisition represents a starting phase in the data science pipeline where researchers acquire data from its sources. While traditional modes of data acquisition include observation, measurement, and generation, in the information era, data acquisition also includes data rescue and data access. Data rescue generally refers to digitization and curation of “dark data”, i.e. data that exists, often from long time ago and in analogue forms, but are not readily available, easily accessible, and directly usable. Data scientists at Tetherless World Constellation have been actively engaged in and successfully accomplished data rescue tasks, e.g. thermodynamic data rescue and diamond data rescue, in collaboration with geoscientists. Data access, on the other hand, is directly accessing and taking full advantage of data from already well-curated and high-quality geoscience databases. Members of the Keck Deep-Time Data Infrastructure Team have developed a number of applications to facilitate data access from geoscience data portals such as the RRUFF IMA Database of Mineral Properties and Mineral Evolution Database, and the Paleobiology Database, and in turn provided high-quality and comprehensive datasets for several successful deep-time data-driven discoveries.</p>