Minerals and the Origins of Life: Insights from Big Data Mineralogy. Robert M. Hazen[1], Ahmed Eleish[2], Chao Liu[1], Shaunna M. Morrison[1], and the Keck Deep-Time Data Collaboration[1], [1]Geophysical Laboratory, Carnegie Institution for Science, 5251 Broad Branch Road NW, Washington DC 20015; [2]Rensselaer Polytechnic Institute, Troy NY 12180. Email: rhazen@ciw.edu

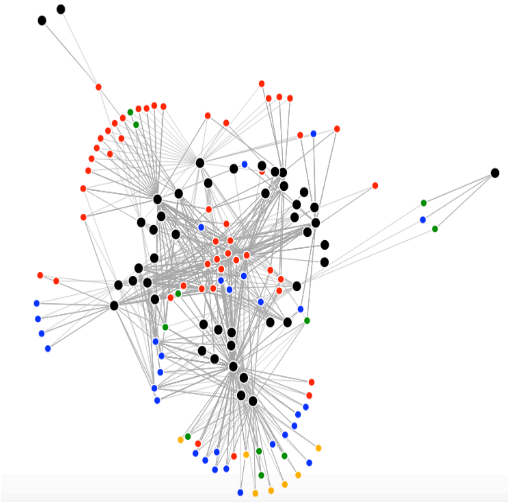
Large and growing data resources catalog every known mineral species and their physical, chemical, and structural properties (rruff.info/ima), as well as mineral localities and associations for samples from Earth, the Moon, Mars, and meteorites (mindat.org). These resources are now being linked to deep-time data to facilitate statistical exploration and visual representation of large-scale patterns of mineral diversity and distribution through deep time. Evolution studies of Earth's changing mineralogy through deep time exploit the "Mineral Evolution Database" with more than 120,000 species/locality/age data. These data reveal a dynamic history of mineralization influenced by tectonic and biological processes, including the Wilson cycle, global glaciation, redox changes of the near-surface environment, microbial evolution, biomineralization, and the rise of the terrestrial biosphere.

Origins-of-life researchers have long recognized the plausible roles of minerals in prebiotic chemistry, including molecular catalysis, selection, protection, and assembly. Deep-time data point to several features of Earth's evolving mineralogy that are relevant to these investigations. For example, mineralogical evidence suggests the presence of a relatively limited number of mineral species on Earth at the time of life's origins compared to the more than 5000 species documented today. In addition, data on major and trace element distributions and species diversification confirm other lines of evidence suggesting significant temporal changes in Earth's near-surface oxidation state and, consequently, the distribution of near-surface minerals.

Trends in mineralogical data through deep time are revealed and clarified by use of a variety of visualization methods. Of special note are recent applications of network analysis to investigate complex patterns of mineral coexistence (see Figures).

Figure Captions: Bipartite mineral networks illustrate associations of copper minerals (colored nodes) with major localities (black nodes). Node colors correspond to mineral compositions based on the presence (+) or absence (-) of sulfur and oxygen: red [+S,-O]; orange [+S,+O], green [-S,-O], blue [-S,+O]. Networks for Archean (A) and Cenozoic (B) Cu minerals document a sharp temporal increase in mineral diversity. Distinctive network geometries reveal that relatively few phases are common (i.e., linked to many localities and thus located near the centers of the networks), whereas many more minerals are rare (and thus positioned around the periphery of the black locality nodes).

(A) Approximately 20 common sulfide minerals dominate the central region of the bipartite network of Archean copper minerals (below). The periphery has a greater number of rarer, predominantly sulfide and oxides phases



(B) The bipartite network of Cenozoic copper minerals (below), by contrast, has a much greater diversity of species than observed in Archean rocks. Oxide and sulfate minerals are more prevalent in the Cenozoic compared to the Archean, though a few dozen common sulfide minerals dominate the central region of the network.