Trace element proxies for oceanic and atmospheric chemical evolution: Deep-time, data-driven discovery

R.M.HAZEN[1]*, X-M.LIU[1], D.A.SVERJENSKY[2], R.T.DOWNS[3], J.GOLDEN[3] AND L.KAH[4]

[1]Geophysical Laboratory, Carnegie Institution, Washington, DC 20015, USA (*correspondence: rhazen@ciw.edu)
[2]Department of Earth & Planetary Sciences, Johns Hopkins University, Baltimore, MD 21218, USA
[3]Department of Geosciences, University of Arizona, Tucson, AZ 85721, USA
[4]Department of Earth & Planetary Sciences, University of Tennessee, Knoxville, TN 37996, USA

Discovery in geochemistry relies principally on induction and deduction—approaches to reasoning that focus on observation, modelling, and predictive explanations of known natural patterns and phenomena. However, these powerful methods are inherently inefficient at discovering new complex patterns that require multivariate analysis of large datasets or synthesis of diverse types of data. Recognition of such gradual global processes, such as oxidation of the oceans and atmosphere, radiation of subsurface microbial ecosystems, and co-evolution of the terrestrial geosphere and biosphere, may require decades of integrated geo- and bioscience data. Accordingly, we are developing a deep-time data infrastructure that links trace element and isotope data for rocks and minerals to paleoenvironment, paleobiology, proteomics, and thermochemical data resources. The potential now exists for an alternative "abductive" approach to investigate Earth's co-evolving geo- and biosphere [1-4].

Our initial efforts have focused on accumulating data on trace element distributions in sulphide and carbonate minerals across deep time [5]. These data reveal significant temporal changes in Earth's near-surface oxidation state; for example, discontinuous changes in ratios of redox-sensitive trace elements. Growing data resources also point to new opportunities for applying multivariate statistical methods and adapting visualization strategies for deep-time data.

Ultimately, we envision an integrated deep-time data infrastructure—a new kind of open-access "scientific instrument" that may facilitate transformation of current Earth science paradigms.

[1] Fayyad *et al* (1996) *AI Magazine* **Fall 1996**:37-54. [2] Hey *et al* [Eds] (2009) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redland, WA: Microsoft External Research. [3] Hazen *et al* (2011) *Am. Mineral.* **96**:953-963. [4] Keller & Schoene (2012) *Nature* **485**:490-493. [5] Golden et al. (2013) *Earth Planet. Sci. Lett.* **366**:1-5.