

# The Sedimentary Geochemistry and Paleoenvironments Project (SGP): a collaborative approach to statistical analyses of deep-time data

UNA C. FARRELL<sup>1</sup>, DAVID T. JOHNSTON<sup>2</sup>, NOAH J. PLANAVSKY<sup>3</sup>, KIMBERLY V. LAU<sup>4</sup>, SAVITHA ANJANAPPA<sup>5</sup>, ERIK A. SPERLING<sup>6</sup> AND SGP COLLABORATIVE TEAM<sup>6</sup>

<sup>1</sup>Trinity College Dublin

<sup>2</sup>Harvard University

<sup>3</sup>Yale University

<sup>4</sup>The Pennsylvania State University

<sup>5</sup>Aionis

<sup>6</sup>Stanford University

Presenting Author: FARRELU@tcd.ie

The Sedimentary Geochemistry and Paleoenvironments Project (SGP) is a community-focused, database-driven research consortium which aims to address questions of deep-time environmental change[1]. Relevant data and associated geological, geographical, and methodological details are organized in a relational database, whose structure is inspired by existing systems in the wider geological community and museum collection databases. The data are presented through a search portal which is tailored to the sedimentary geochemical community. The scientific and community approach of SGP is modeled on paleobiological studies, where researchers have worked hard to understand the impact of sampling intensity and geological bias on their results, and efforts in fields like medical genetics, where multiple research groups have combined their data for more statistical power.

Here we describe the SGP data model and data collection methods, and discuss the ways in which SGP aims to facilitate collaborative research projects and to foster a culture of data-curation and data-sharing in sedimentary geochemistry. In particular, collaborators are given priority access to the compiled data and the opportunity to work on group publications before the data are made publicly available through the SGP search website (<https://sgp-search.io>). We outline the Phase 1 data product, now openly available, which is focused on shale data from the Neoproterozoic-Paleozoic with a global distribution. Initial working-group outcomes include publications on weathering and composition of continental crust through time, changing Cambrian redox conditions in relation to evolutionary dynamics, and approaches to processing and analyzing large datasets. We specifically demonstrate the value of applying powerful statistical methods to large context-rich datasets by summarizing a study which examines redox changes in Neoproterozoic-Paleozoic oceans using machine learning techniques. Finally, we introduce Phase 2 goals, which include addition of more carbonate data and an expansion to all time periods. We emphasize that SGP is open to all; we welcome new collaborators to help build and analyze our sedimentary