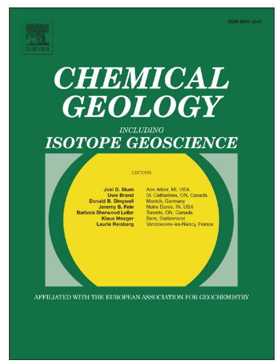
Journal Pre-proof

The Sedimentary geochemistry and paleoenvironments project phase 2 data release: An open data resource for the study of Earth's environmental history

Úna C. Farrell, Hunter C. Olson, Maya O. Thompson, Michelle L. Abshire, Oyeleye O. Adeboye, Anne-Sofie C. Ahm, Lewis J. Alcott, Thomas J. Algeo, Ross P. Anderson, Arif H. Ansari, Lucas Pinto Heckert Bastos, Kohen W. Bauer, Brian Beaty, Justin E. Birdwell, Fred T. Bowyer, Jochen J. Brocks, Tessa Brunoir, James F. Busch, Donald E. Canfield, Fabrício A. Caxito, Chao Chang, Meng Cheng, Jean N.R. Clemente, David R. Cordie, Peter W. Crockford, Huan Cui, Celeste M. Cunningham, Tais W. Dahl, Janaina Rodrigues de Paula, Carol M. Dehler, Lucas Del Mouro, Keith Dewing, Dermeval Aparecido do Carmo, Stephen Q. Dornbos, Nadja Drabon, Julie A. Dumoulin, Omabehere Innocent Ejeh, Emily Ellefson, Maya Elrick, Joseph F. Emmings, Bokanda Ekoko Eric, Hao Fang, Gabriella Fazio, Henrique A. Fernandes, Katherine L. French, Robert R. Gaines, Richard M. Gaschnig, Timothy M. Gibson, Geoffrey J. Gilleaudeau, Karin Goldberg, Zheng Gong, Amy P.I. Hagen, Galen P. Halverson, Kalev Hantsoo, Emma R. Haxen, Miles A. Henderson, João P.T.M. Hippertt, Malcolm S.W. Hodgskiss, Paul F. Hoffman, Edward C. Huang, Benjamin W. Johnson, Pavel B. Kabanov, Junyao Kang, C. Brenhin Keller, Brian Kendall, Julien Kimmig, Sara R. Kimmig, Michael A. Kipp, Andrew H. Knoll, Timmu Kreitsmann, Anurag A. Kulkarni, Alexandra Kunert, Marcus Kunzmann, Jiankang Lai, Richard O. Lease, Chao Li, Sen Li, Alex G. Lipp, Yang Liu, David K. Loydell, Xinze Lu, Katie M. Maloney, Kaarel Mänd, Alexie E.G. Millikin, N. Tanner Mills, Kento Motomura, Chiza N. Mwinde, Lyle L. Nelson, Nora M. Nieminski, Brennan O'Connell, Edel O'Sullivan, Juliana Okubo, Jaden K. Olah, Frantz Ossa Ossa, Chadlin M. Ostrander, Kärt Paiste, Camille A. Partin, Egberto Pereira, Shanan E. Peters, Tiffany Playter, Susannah M. Porter, Simon W. Poulton, Sara B. Pruss, Zhen Qiu, Daven P. Quinn, Mariano Remírez, Sebastian Richiano, Sylvain Richoz, Kathryn I. Rico, Samantha R. Ritzer, Zachary Roney, Alan D. Rooney, William C. Rose, Elias J. Rugen, Swapan K. Sahoo, Shane D. Schoepfer, Judith A. Sclafani, Nathan D. Sheldon, Yanan Shen, Graham A. Shields, Pulkit Singh, Arvind Kumar Singh, Sarah P. Slotznick, Emily F. Smith, Haijun Song, Sam C. Spinks, Richard G. Stockey, Justin V. Strauss, Eva E. Stüeken, Zongyuan Sun, Dongjie Tang, Lidya G. Tarhan, Danielle



Thomson, Nicholas J. Tosca, Rosalie Tostevin, Chenyi Tu, Maoli N. Vizcaíno, Yuxuan Wang, Changle Wang, Xiaomei Wang, Lucas Veríssimo Warren, Lucy C. Webb, Philip R. Wilby, Christina R. Woltz, Rachel Wood, Yuyang Wu, Xiuqing Yang, Inessa A. Yurchenko, Junpeng Zhang, Jessica H. Whiteside, Benjamin C. Gill, Akshay K. Mehra, Kimberly V. Lau, Noah Planavsky, David T. Johnston, Erik A. Sperling

PII: S0009-2541(26)00080-X

DOI: https://doi.org/10.1016/j.chemgeo.2026.123311

Reference: CHEMGE 123311

To appear in: *Chemical Geology*

Received date: 11 February 2026

Accepted date: 18 February 2026

Please cite this article as: Ú.C. Farrell, H.C. Olson, M.O. Thompson, et al., The Sedimentary geochemistry and paleoenvironments project phase 2 data release: An open data resource for the study of Earth's environmental history, *Chemical Geology* (2024), https://doi.org/10.1016/j.chemgeo.2026.123311

The Sedimentary Geochemistry and Paleoenvironments Project Phase 2 Data Release: An open data resource for the study of Earth's environmental history

Úna C. Farrell[1,2]*, Hunter C. Olson[1], Maya O. Thompson[1], Michelle L. Abshire[3], Oyeleye O. Adeboye[4], Anne-Sofie C. Ahm[5], Lewis J. Alcott[6], Thomas J. Algeo[7,8], Ross P. Anderson[9], Arif H. Ansari[10], Lucas Pinto Heckert Bastos[11], Kohen W. Bauer[12], Brian Beaty[13], Justin E. Birdwell[14], Fred T. Bowyer[15,16], Jochen J. Brocks[17], Tessa Brunoir[18], James F. Busch[19], Donald E. Canfield[20,21], Fabrício A. Caxito[22], Chao Chang[23], Meng Cheng[24], Jean N.R. Clemente[25,26], David R. Cordie[27], Peter W. Crockford[28], Huan Cui[29], Celeste M. Cunningham[30], Tais W. Dahl[31], Janaina Rodrigues de Paula[22], Carol M. Dehler[32], Lucas Del Mouro[33], Keith Dewing[34], Dermeval Aparecido do Carmo[35], Stephen Q. Dornbos[36], Nadja Drabon[37], Julie A. Dumoulin[38], Omabehere Innocent Ejeh[39], Emily Ellefson[1], Maya Elrick[40], Joseph F. Emmings[41], Bokanda Ekoko Eric[42], Hao Fang[24], Gabriella Fazio[43], Henrique A. Fernandes[44], Katherine L. French[45], Robert R. Gaines[46], Richard M. Gaschnig[47], Timothy M. Gibson[13], Geoffrey J. Gilleaudeau[48], Karin Goldberg[29], Zheng Gong[49], Amy P.I. Hagen[50], Galen P. Halverson[51], Kalev Hantsoo[52], Emma R. Haxen[31,53], Miles A. Henderson[54], João P.T.M. Hippertt[55], Malcolm S.W. Hodgskiss[56], Paul F. Hoffman[5,37], Edward C. Huang[57], Benjamin W. Johnson[58], Pavel B. Kabanov[34,59], Junyao Kang[50,60], C. Brenhin Keller[19], Brian Kendall[61], Julien Kimmig[62,63], Sara R. Kimmig[64], Michael A. Kipp[65], Andrew H. Knoll[66], Timmu Kreitsmann[67], Anurag A. Kulkarni[1,68], Alexandra Kunert[61], Marcus Kunzmann[69], Jiankang Lai[8], Richard O. Lease[38], Chao Li[24,70], Sen Li[8], Alex G. Lipp[71], Yang Liu[72], David K. Loydell[73], Xinze Lu[61], Katie M. Maloney[74], Kaarel Mänd[67], Alexie E.G. Millikin[75], N. Tanner Mills[76], Kento Motomura[77], Chiza N. Mwinde[78], Lyle L. Nelson[79], Nora M. Nieminski[80], Brennan O'Connell[56], Edel O'Sullivan[81], Juliana Okubo[82], Jaden K. Olah[1], Frantz Ossa Ossa[83,84,85], Chadlin M. Ostrander[86], Kärt Paiste[67], Camille A. Partin[87], Egberto Pereira[11,88], Shanan E. Peters[89], Tiffany Playter[90], Susannah M. Porter[91], Simon W. Poulton[15], Sara B. Pruss[92], Zhen Qiu[94], Daven P. Quinn[89], Mariano Remírez[31], Sebastian Richiano[94], Sylvain Richoz[95], Kathryn I. Rico[96], Samantha R. Ritzer[1], Zachary Roney[97], Alan D. Rooney[13], William C. Rose[1], Elias J. Rugen[71], Swapan K Sahoo[98], Shane D. Schoepfer[99], Judith A. Sclafani[100], Nathan D. Sheldon[96], Yanan Shen[101], Graham A. Shields[71], Pulkit Singh[1], Arvind Kumar Singh[10,102], Sarah P. Slotznick[19], Emily F. Smith[97], Haijun Song[8], Sam C. Spinks[103,104], Richard G. Stockey[105], Justin V. Strauss[19], Eva E. Stüeken[106], Zongyuan Sun[24], Dongjie Tang[107], Lidya G. Tarhan[13], Danielle Thomson[108], Nicholas J. Tosca[56], Rosalie Tostevin[109], Chenyi Tu[65], Maoli N. Vizcaíno[13], Yuxuan Wang[15,110], Changle Wang[111], Xiaomei Wang[93], Lucas Veríssimo Warren[82], Lucy C. Webb[1], Philip R. Wilby[112], Christina R. Woltz[1,113], Rachel Wood[16], Yuyang Wu[8,114], Xiuqing Yang[115], Inessa A. Yurchenko[116], Junpeng Zhang[117], Jessica H. Whiteside[118], Benjamin C. Gill[50], Akshay K. Mehra[119], Kimberly V. Lau[52,120], Noah Planavsky[13], David T. Johnston[37], Erik A. Sperling[1]**

[1] Stanford University, Department of Earth and Planetary Sciences, Stanford, CA 94305, USA
[2] Trinity College Dublin, Department of Geology, Dublin, Ireland
[3] Minnesota State University, Moorhead, Anthropology and Earth Science Department, Moorhead, MN 56563, USA
[4] Oklahoma State University, Boone Pickens School of Geology, Stillwater, OK 74078, USA
[5] University of Victoria, School of Earth and Ocean Sciences, Victoria, BC V8N 1B3, Canada

[6] School of Earth Sciences, University of Bristol, Bristol, BS8 1RJ, United Kingdom
[7] University of Cincinnati, Department of Geosciences, Cincinnati, OH 45221, USA
[8] State Key Laboratory of Geomicrobiology and Environmental Changes (GMEC), China University of Geosciences, Wuhan, Hubei 430074, P.R., China
[9] University of Oxford, Museum of Natural History, Oxford, Oxfordshire OX1 3PW, United Kingdom
[10] Birbal Sahni Institute of Palaeosciences, Precambrian Palaeobiology, Lucknow, Uttar Pradesh 226007, India
[11] Universidade do Estado do Rio de Janeiro (Rio de Janeiro State University), Chemostratigraphy and Organic Geochemistry Laboratory (LGQM), Rio de Janeiro, RJ, 20550-900, Brazil
[12] University of Victoria, Ocean Networks Canada, Victoria, BC V8N 1V8, Canada
[13] Yale University, Department of Earth and Planetary Sciences, New Haven, CT, USA
[14] U.S. Geological Survey, Central Energy Resources Science Center, Denver, CO, USA
[15] University of Leeds, School of Earth and Environment, Leeds LS2 9JT, United Kingdom
[16] University of Edinburgh, School of GeoSciences, Edinburgh EH9 3FE, United Kingdom
[17] The Australian National University, Research School of Earth Sciences, Canberra, ACT 2601, Australia
[18] University of California Davis, Department of Earth & Planetary Sciences, Davis, CA 95616, USA
[19] Dartmouth College, Department of Earth and Planetary Sciences, Hanover, NH 03755, USA
[20] University of Southern Denmark, Biology, Campusvej 55, 5230 Odense M, Denmark
[21] Danish Institute for Advanced Studies, University of Southern Denmark, Campusvej 55, 5230 Denmark
[22] Universidade Federal de Minas Gerais, Programa de Pós Graduação em Geologia Geoscience Institute, Department of Geology, 6627 Antonio Carlos Ave., Belo Horizonte, Minas Gerais, Brazil, 31270901
[23] Northwest University, State Key Laboratory of Continental Evolution and Early Life, Shaanxi Key Laboratory of Early Life and Environment, Xi'an, China
[24] Chengdu University of Technology, State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation & Institute of Sedimentary Geology, Chengdu 610059, China
[25] MIT-WHOI Joint Program in Oceanography, Woods Hole, MA 02543, USA
[26] Woods Hole Oceanographic Institution, Department of Geology and Geophysics, Woods Hole, MA 02543, USA
[27] Edgewood University, Department of Biological Sciences, Madison, WI 53711, USA
[28] Carleton University, Earth Sciences, Ottawa, ON K1S 5B6, Canada
[29] Kansas State University, Department of Geology, Manhattan, KS 66506, USA
[30] St. Francis Xavier University, Department of Earth and Environmental Sciences, Antigonish, NS B2G 2W5, Canada
[31] University of Copenhagen, Department of Geosciences and Natural Resource Management, Copenhagen K, Denmark
[32] Utah State University, Geosciences, Logan, UT 84322, USA
[33] Federal University of Santa Catarina, Department of Geology, Florianópolis, Santa Catarina, Brazil
[34] Geological Survey of Canada, Natural Resource Canada, Calgary, AB T2L 2A7, Canada
[35] University of Brasília, Institute of Geosciences, Brasília, Brazil

[36] University of Wisconsin-Milwaukee, Department of Geosciences, Milwaukee, WI 53211, USA
[37] Harvard University, Department of Earth and Planetary Sciences, Cambridge, MA 02138, USA
[38] U.S. Geological Survey, Alaska Science Center, Anchorage, AK 99508, USA
[39] Delta State University, Department of Geology, Abraka, Delta, Nigeria
[40] University of New Mexico, Earth and Planetary Sciences, Albuquerque, NM 87131 USA
[41] Geotek Ltd., Daventry, Drayton Fields, NN11 8PB, United Kingdom
[42] University of Buea, Cameroon, Department of Geology, Chemical and Petroleum Engineering, Buea, Cameroon
[43] Observatório Nacional (ON/MCTI), Geophysics Department, Rio de Janeiro, Brazil
[44] Universidade de São Paulo, Instituto de Geociências, São Paulo, Brazil
[45] U.S. Geological Survey, Central Energy Resources Science Center, Denver, CO, USA
[46] Pomona College, Geology Department, Claremont, CA 91711, USA
[47] University of Massachusetts Lowell, Department of Environmental, Earth and Atmospheric Sciences, Lowell, MA 01854, USA
[48] George Mason University, Department of Atmospheric, Oceanic, and Earth Sciences, Fairfax, VA 22030, USA
[49] Nanjing University, School of Earth Sciences and Engineering, Nanjing, Jiangsu, China
[50] Virginia Polytechnic Institute and State University, Department of Geosciences, Blacksburg, VA 24061, USA
[51] McGill University, Earth and Planetary Sciences/Geotop, Montreal, QC H3A 0E8, Canada
[52] Penn State University, Department of Geosciences, University Park, PA 16802, USA
[53] Lund University, Department of Experimental Medical Science, Lund, Sweden
[54] The University of Texas Permian Basin, Department Geosciences, Odessa, TX 79705, USA
[55] Federal University of Ouro Preto (UFOP), Department of Geology (DEGEO), Ouro Preto, Brazil
[56] University of Cambridge, Department of Earth Sciences, Cambridge, United Kingdom
[57] Stanford University, Department of Computer Science, Stanford, CA 94305, USA
[58] Iowa State University, Department of Earth, Atmosphere, and Climate, Ames, IA 50010, USA
[59] University of Calgary, Department Of Earth, Energy and Environment, Calgary, AB T2N 1N4, Canada
[60] InPlanet GmbH, Heinrich-Geißler-Straße 20, 80939 München, Germany
[61] University of Waterloo, Department of Earth and Environmental Sciences, Waterloo, ON N2L 3G1, Canada
[62] State Museum of Natural History Karlsruhe, Geosciences, Erbprinzenstr. 13, 76133 Karlsruhe, Germany
[63] University of North Dakota, Harold Hamm School of Geology and Geological Engineering, Grand Forks, ND 58202, USA
[64] Karlsruhe Institute of Technology (KIT), Institute of Applied Geosciences, Chair of Geochemistry & Economic Geology, Karlsruhe, Germany
[65] Duke University, Nicholas School of the Environment, Division of Earth and Climate Sciences, Durham, NC 27708, USA
[66] Harvard University, Department of Organismic and Evolutionary Biology, Cambridge MA 02138, USA
[67] University of Tartu, Department of Geology, Tartu, Tartumaa, Estonia
[68] University of Texas at Austin, Department of Earth and Planetary Sciences, Jackson School of Geosciences, Austin, TX 78712, USA
[69] Rio Tinto Exploration, 37 Belmont Ave., Belmont, Western Australia, 6104, Australia

[70] Chengdu University of Technology, International Center for Sedimentary Geochemistry and Biogeochemistry Research, Chengdu 610059, China
[71] University College London, Earth Sciences, London WC1E 6BS, United Kingdom
[72] Dalian Maritime University, College of Environmental Science and Engineering, Dalian, China
[73] University of Portsmouth, School of the Environment and Life Sciences, Portsmouth, United Kingdom
[74] Michigan State University, Earth and Environmental Sciences, East Lansing, MI 48824, USA
[75] University of Colorado Boulder, Department of Geological Sciences, Boulder, CO 80309, USA
[76] University of Texas at Austin, Institute for Geophysics, Jackson School of Geosciences, Austin, TX 78758, USA
[77] Kyushu University, Department of Earth and Planetary Sciences, Fukuoka, Japan
[78] University of Chicago, Geophysical Sciences, Chicago, IL 60637, USA
[79] Massachusetts Institute of Technology, Department of Earth, Atmospheric & Planetary Sciences, Cambridge, MA 02139, USA
[80] Alaska Division of Geological & Geophysical Surveys, Anchorage, AK 99508, USA
[81]Marine Institute, Rinville, Oranmore, Co. Galway, H91 R673, Ireland
[82] São Paulo State University, Department of Geology, Rio Claro, SP, Brazil
[83] Khalifa University of Science and Technology, Earth Science, Abu Dhabi, United Arab Emirates
[84] Khalifa University of Science and Technology, Polar Research Center, Abu Dhabi, United Arab Emirates
[85] University of Johannesburg, Department of Geology, Johannesburg 2092, South Africa
[86] University of Utah, Department of Geology & Geophysics, Salt Lake City, UT 84105, USA
[87] University of Saskatchewan, Department of Geological Sciences, Saskatoon, SK S7N 5E2, Canada
[88] Universidade do Estado do Rio de Janeiro (Rio de Janeiro State University), Department of Stratigraphy and Paleontology (DEPA), Rio de Janeiro, RJ, 20550-900, Brazil
[89] University of Wisconsin-Madison, Department of Geoscience, Madison, WI 53706, USA
[90] University of Alberta, Department of Earth and Atmospheric Sciences, Edmonton, AB T6G 2E3, Canada
[91] University of California at Santa Barbara, Earth Science, Santa Barbara, CA, USA
[92] Smith College, Department of Geosciences, Northampton MA 01063, USA
[93] China National Petroleum Corporation, Research Institute of Petroleum Exploration & Development, Beijing 100083, China
[94] National University of La Patagonia San Juan Bosco (UNPSJB), Patagonian Institute of Geology and Paleontology (IPGP-CONICET) and Geology (UNPSJB), Puerto Madryn, Chubut, Argentina.
[95] Lund University, Department of Geology, Lund, Sweden
[96] University of Michigan, Department of Earth and Environmental Sciences, Ann Arbor, MI 48109, USA
[97] Johns Hopkins University, Department of Earth & Planetary Sciences, Baltimore, MD 21210, USA
[98] SHRI (Shaligram Research Institute), Fulshear, TX, USA
[99] Western Carolina University, Geoscience and Natural Resources Department, Cullowhee, NC 28723, USA

[100] University of the Pacific, Department of Geological and Environmental Sciences, Stockton, CA 95211, USA
[101] University of Science and Technology of China, School of Earth and Space Sciences, Hefei 230026, China
[102] Academy of Scientific and Innovative Research, Ghaziabad, Uttar Pradesh, India
[103] Teck Australia Pty Ltd., Exploration, Perth, Western Australia, Australia
[104] CSIRO Mineral Resources, ARRC, Dick Perry Avenue, Kensington, WA 6151, Australia
[105] School of Ocean and Earth Science, University of Southampton, Southampton, Hampshire SO14 3ZH, United Kingdom
[106] University of St Andrews, School of Earth & Environmental Sciences, St Andrews, Fife, KY16 9TS, United Kingdom
[107] State Key Laboratory of Geomicrobiology and Environmental Changes, Frontiers Science Center for Deep-time Digital Earth, China University of Geosciences (Beijing), Beijing 100083, China
[108] Canadian Natural Resources Limited, Calgary, AB, Canada
[109] University of Cape Town, Department of Geological Sciences, Rondebosch, Cape Town, 7001, South Africa.
[110] GFZ Helmholtz Centre for Geosciences, Potsdam, Germany
[111] Institute of Geology and Geophysics, Chinese Academy of Sciences, State Key Laboratory of Lithospheric and Environmental Coevolution, Beijing, China
[112] British Geological Survey, Keyworth, Nottingham NG12 5GG, UK
[113] Imperial College London, Department of Earth Science and Engineering, London SW7 2AZ United Kingdom
[114] College of Marine Science and Technology, China University of Geosciences, Wuhan 430074, China
[115] Chang'an University, School of Earth Science and Resources, Xi'an, Shaanxi Province, China
[116] The University of Texas at Dallas, Sustainable Earth Systems Sciences, Dallas, TX 75080, USA
[117] Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing, Jiangsu Province, China
[118] San Diego State University, Department of Earth and Environmental Sciences, San Diego, CA 92182, USA
[119] University of Washington, Department of Earth and Space Sciences, Seattle, WA 98195, USA
[120] Earth and Environmental Systems Institute, Penn State University, University Park, PA 16802, USA

Authors for correspondence: *una.farrell@gmail.com, **esper@stanford.edu

Abstract
Geochemical data from sedimentary rocks are the primary source of information regarding Earth's surface evolution through time, including its air and water envelopes and interactions with life and deep Earth processes. The Sedimentary Geochemistry and Paleoenvironments Project (SGP) is a scientific consortium centered around open data and community-driven development of cyberinfrastructure tools and resources for sedimentary geochemistry and Earth history. Here we describe the SGP Phase 2 data release, which focused on incorporating Paleoproterozoic and

Mesoproterozoic (2500 – 1000 million years ago) data and better accommodating carbonate data. This data release was built through the involvement of >200 researchers worldwide in academia, government, and industry, and provides the largest available public data resource for our user community in the academic fields of geochemistry, sedimentology, tectonics, paleontology, Earth history, and paleoclimate, as well as the petroleum and minerals industries. The dataset now encompasses 126,006 samples and 4,132,371 geochemical analyses. In addition to direct entry by SGP Team Members, we have ingested and incorporated datasets from the Geoscience Australia OZCHEM database, the Alberta Geological Survey, and the Deep-Time Marine Sedimentary Element Database (DM-SED) compilation. This paper details sampling in the Phase 2 dataset with respect to age, geography, lithology, and other geological characteristics, documents access via our search website and API, discusses possible issues and/or biases in the dataset that could impact analyses, describes plans for governance and stewardship of data from Indigenous lands, and serves as the citable reference paper for the data release.

1. Introduction

How has Earth's surface, including its atmosphere, oceans, and depositional systems changed over time? How have these changes affected life on Earth? Were these environmental changes driven by tectonic changes and solid Earth processes, or driven by life itself? Answers to these first-order questions about Earth and its history are found in the geochemistry of Earth's sedimentary rocks. For decades, geologists, geochemists, paleontologists, and sedimentologists have generated relevant data. As results accumulated, researchers increasingly gravitated towards larger data compilations and global syntheses, a prerequisite for studies across long timescales (Alcott et al., 2025). Compilation efforts initially took the form of flat-file (e.g., Excel) data tables. In 2016, the Sedimentary Geochemistry and Paleoenvironments Project (SGP), an open research consortium styled after efforts in the biomedical community, spurred the development of a sedimentary geochemical database, now accessible via a search website and Application Programming Interface (API). More details on the history of SGP and the philosophy and reasoning behind its inception can be found in Farrell et al. (2021). Our Phase 1 project (Farrell et al., 2021) primarily focused on Neoproterozoic–Paleozoic (1000-250 million years ago, Ma) shale records, and has been fundamental to diverse studies of environmental change, weathering history, carbon and nutrient cycling, fossil preservation, paleoclimate, and ore and petroleum deposits (for a non-exhaustive list, refer to Bian et al., 2025; Bishop and Robbins, 2024; Bubphamanee et al., 2025; Canfield et al., 2025; Cui et al., 2023; Emmings et al., 2022; Ernst et al., 2023; Hantsoo et al., 2024; Kimmig and Pratt, 2022; Lipp et al., 2021; Mehra et al., 2021; Murray and Jagoutz, 2024; Olson et al., 2025; Roest-Ellis et al., 2023; Stockey et al., 2024; Tang et al., 2024; Venugopal et al., 2025, 2023; Walton and Shorttle, 2024; Wang et al., 2023; Wei et al., 2024; Ye et al., 2024; Zhang et al., 2022; Zhao et al., 2024). This paper describes our Phase 2 data release, which involved data ingestion, cleaning, and database updates carried out between late 2020 and early 2025.

Following the model established by biomedical research consortia, we publicly release data in scheduled 'data freezes.' First, we build a proprietary data product with novel scientific value. This step allows our scientific Working Groups to focus solely on analysis, leading to high-impact scientific studies that would otherwise not be possible. After a set period of proprietary access, the data are publicly released via our search website and API. Concurrently, we publish a group paper as a citable reference describing the data product (e.g., this paper). This approach is similar to the long-successful embargo period for International Ocean Drilling Program/International Ocean Discovery Program (IODP) (ocean drilling) data and samples, which gives participating scientists time to publish their manuscripts while also establishing a repository for future research. This database and its access points are designed in accordance with and align to FAIR principles (Findability, Accessibility, Interoperability, Reusability) for scientific data management and reusability (Wilkinson et al., 2016).

2. Database

SGP uses a PostgreSQL relational database that remains relatively unchanged from the Phase 1 database described in Farrell et al. (2021; additional changes described below). It is based around a modified version of the British Geological Survey Geochemistry database. SGP is a sample-centric database — rock samples are the core entity at the center of the schema, with all other data types linked to them. Samples are linked to sample-level data (e.g., lithology, age) as well as to sites (e.g., drillcore or stratigraphic section) that record higher-level geographic features. Geochemical methodologies are tracked using three controlled vocabularies for a) preparation method (how the sample was ground to a powder; e.g., in a tungsten carbide vessel), b) experimental method (processing steps; e.g., a weak nitric acid digestion), and c) analytical method (how the measurement was made; e.g., ICP-MS). Details on measurement accuracy, precision, reference materials, and upper/lower measurement limits are recorded as reported in ingested data tables. Tables for geological, geographical, and sample details are based on collection management databases such as Specify6 and Arctos, as well as the Observations Data Model 2 (an information model for Earth observations). However, additional tables and modifications have been introduced to address the specific geological and geographical context data required for the SGP project (and sedimentary geochemistry more broadly). Our full schema is described in https://ufarrell.github.io/sgp_phase2/. While the database is tailored to the specific research needs of the sedimentary geochemistry community, in keeping with FAIR principles (Wilkinson et al., 2016) it is also designed with larger communities in mind. We have used common vocabularies and standard community-approved terms wherever possible, and the database structure shares features (and identifiers) with existing databases such as EarthChem, Macrostrat, and the Paleobiology Database (PBDB). This facilitates the transfer of data to, or interaction with, other public community databases.

Recently we have implemented a direct database-level integration with Macrostrat (Peters et al., 2018), a geoinformatics project that compiles data on geological maps, stratigraphic columns, and age models—the geological framework for our geochemical data. Specifically, we have 1) harmonized and aligned definitions for rock attributes, 2) linked SGP samples to Macrostrat units and columns, and are 3) building new web interfaces for stratigraphic column visualization and entry, and 4) engaging the SGP community in capturing column data that contextualize geochemical data. Our samples will ultimately be linked to Macrostrat's continuous-time and automatically revised age model (in addition to our declarative age model) and details on depositional environment, lithology, and paleogeography—a critical step for maintaining the future utility of sedimentary geochemical data archived in SGP. This also represents steps towards the establishment of unified chronostratigraphic frameworks for geochemical and paleontological data that have been identified as a key goal by the deep-time research community (Alcott et al., 2025).

2.1 Carbonate and metal isotope data

In Phase 2, modifications were made to the database to support correct storage of carbonate geochemical data. Because carbonate data are more commonly from a specific allochem (for instance a brachiopod fossil) or a specific matrix (for instance micrite versus cement) compared to shale geochemical data (which are generally from a "bulk" analysis) we have introduced "sample type" as a sample-level feature. We have also incorporated 'parent/child' relationships for sub-samples (e.g., analyzing multiple different brachiopod fossils within a carbonate hand sample). Although most analyses in the database remain "bulk," this allows for future flexibility in incorporating essentially any type of sub-sample analysis (e.g., LA-ICP-MS data). Following community input from carbonate workers, we have also incorporated a more detailed break-down of depositional environments ("depositional environment detail"), allowing for discrimination between multiple inner shelf environments such as peritidal, carbonate shoal, reef, etc., in addition to the coarser "depositional environment bin" of fluvial/lacustrine/inner shelf/outer shelf/basinal.

Adding carbonate data has also required a re-structuring of analyte naming conventions. Correct interpretation of sedimentary geochemical data requires knowing what mineral or phase was analyzed: for instance, if a $\delta^{13}C$ analysis is from organic carbon or carbonate, or if a $\delta^{238}U$ analysis is from carbonate or shale. In these examples, the two paired phases have substantial isotopic differences related to carbon isotope fractionation during photosynthesis (Park and Epstein, 1960) and nuclear volume effect changes during uranium reduction and incorporation into a shale (Andersen et al., 2017), respectively. These are apples-and-oranges measurements that cannot be mixed in a data table. Naming analytes according to phase existed to an extent in SGP Phase 1, for instance recording $\delta^{13}C$-organic versus $\delta^{13}C$-carbonate, or $\delta^{34}S$-pyrite versus $\delta^{34}S$-gypsum, rather than the isotopic measurement itself (i.e., simply $\delta^{13}C$ or $\delta^{34}S$). In Phase 2 we have more broadly applied this logic to elemental and metal isotopic data, using a tripartite system to track analyses targeting specific phases:

Auth: authigenic, from a weak/dilute acid leach on shales, targeting metals associated with organic matter or pyrite. This usually involves a single-acid leach; note that the strength and type of the acid can vary.

Carb: carbonate, from a weak/dilute acid leach on carbonates, targeting metals in the carbonate crystal lattice. This usually involves a single-acid leach; note that the strength and type of the acid can vary. Analyses may include relatively strong acids (e.g., a 6N HCl digest) that may leach elements from clay or detrital minerals compared to a weaker leach, but that are philosophically targeting the carbonate fraction.

Total: Elemental abundance or metal isotopes in all phases including detrital, resulting from total digestion of the rock. Note that aqua regia digests are considered total, although such analyses underestimate the total as they do not digest silicate and oxide minerals.

This tripartite division, and exactly what digestion types are included in each category, will remain static in the Phase 2 website but this is an area where we aim to continue improving and evolving in consultation with the community. Aqua regia (concentrated nitric + hydrochloric acid) digests are a good example; these are run by many commercial geochemical labs (e.g., Actlabs, Bureau Veritas) and since many of the chalcophile elements like Mo or Ni in shale are hosted by organic matter or pyrite that are dissolved by aqua regia, it is a 'nearly total' extraction for these elements.

But it is not as ‘total’ as a four-acid digestion involving hydrofluoric acid, or a fusion methodology, and could equally be considered authigenic (or even be poorly quantitative, depending on the element and host phase(s) (refer to Xu et al., 2012)). This tripartite division also diverges from other databases such as EarthChem, which simply reports the analyte irrespective of phase or extraction type. We believe that such data tabulations (for instance a table of $\delta^{13}C$ data without discrimination whether it comes from organic matter or carbonate) have low utility for sedimentary geochemistry and Earth history studies. While capturing information on phase is critical, this does introduce problems in future data harmonization efforts that are an important goal of the geoinformatics community (Chamberlain et al., 2021; note however that our more granular analyte names can easily be collapsed—for instance $\delta^{13}C$-organic or $\delta^{13}C$-carbonate to simply $\delta^{13}C$—whereas the reverse is not true).

In the current website, “total” element abundances are exposed under the Show → Elements tab (or filtered in the Analyte Filters → Elements tab in a Detailed Search). “Carbonate” elemental abundances (along with common carbonate geochemical proxies for diagenesis such as Mn/Sr) are exposed in the Show → Carbonate Proxies tab. All metal isotopes, including auth, carb, and total, are exposed in the Show → Metal Isotopes tab. We have relatively few “authigenic” elemental abundances (primarily thallium and vanadium) and these are exposed with identifying subscripts under the Show → Elements tab. More details on data presentation and access are included below.

3. Data Collection and Phase 2 Updates

Data collection, collation, and entry into the SGP database occurs through SGP Collaborative Team direct entry and by ingestion of larger published data products. These “data sources” can be selected/filtered through the API or in the first tab under Sample Filters on the Detailed Search or Analysis search pages. In SGP direct entry, Collaborative Team Members fill in standardized context sheets with relevant geological meta-data, which are then directly entered into the database. More information on direct entry is available in Farrell et al. (2021) and the SGP wiki. This process continued during Phase 2, and some data entered during Phase 1 were updated based on newly published geological information or geologic/informatic errors identified by database users (specifically, the geological context data of thousands of samples were enhanced). SGP direct entry now comprises 48,154 samples.

In Phase 1 we ingested data from the United States Geological Survey (USGS) Critical Metals in Black Shales project (Granitto et al., 2017) (USGS-CMIBS) and the USGS National Geochemical Database: Rock (USGS-NGDB).

In Phase 2 we ingested data from three new data sources: the Geoscience Australia OZCHEM database, the Alberta Geological Survey, and the Deep-Time Marine Sedimentary Element Database, or DM-SED (Lai et al., 2025). We also excluded a number of samples (primarily from USGS-NGDB and USGS-CMIBS) in our efforts to better incorporate Indigenous data stewardship principles (refer to “Indigenous Data Governance” below). Some Phase 2 samples do not have associated data, and therefore total sample counts are slightly higher than the sum of samples by data source (samples are linked to data source through batches of data). Extended details on sample counts, data ingestion, and illustrations of sampling by age, geography, and lithology for each data source can be found on the SGP wiki (https://github.com/ufarrell/sgp_phase2/wiki).

3.1 USGS-NGDB and CMIBS updates

During Phase 2 we updated all interpreted ages for these projects that were solely derived from the Macrostrat age model with the newest available ages. We also replaced some samples with a more precisely coded version from the original authors as part of SGP Collaborative Team direct entry. Reference works were added and linked to CMIBS samples, based on the CMIBS "publ_id." We also note a growing recognition that the CMIBS project may include some mineralized samples and that the NGDB database includes a large proportion of terrestrial samples. Internet searches of Phase 1 NGDB and CMIBS formation names that had three or more samples suggested that 129 formations (29%) were terrestrial, 264 formations were marine (59%), and 52 formations (12%) were mixed terrestrial/marine (Canfield et al., 2025). The majority of terrestrial and mixed formations (86%) were from NGDB, largely due to a robust historical USGS sampling program of western U.S. sandstones. Because NGDB samples are not coded with respect to depositional environment ("environmental bin") and terrestrial samples cannot easily be excluded, caution is warranted when considering whether to include these data in studies aiming to reconstruct marine geochemical trends. The environmental interpretations of Canfield et al. (2025), based on Google searches of formation names in NGDB and CMIBS, are now available under "environmental notes" and could be used for manual sampling culling. With respect to CMIBS, although the papers included during Phase 1 were initially screened to exclude clearly mineralized samples/ore systems, the focus of that literature compilation was identifying anomalous enrichments in critical metals. These studies commonly analyzed samples 'distal' to ore systems or from complex volcanic-sedimentary terranes. Such geological settings may have experienced some mineralization, and these samples would not have been removed by our initial manuscript-level screening. Note that both USGS projects remain uncoded for "sample type" because such details are not available, even though both are largely comprised of data from "bulk" samples. Publicly available USGS-NGDB and USGS-CMIBS data sources now contribute 42,585 and 12,276 samples, respectively.

3.2 Geoscience Australia OZCHEM

Inorganic geochemistry data were added from the Geoscience Australia OZCHEM National Whole Rock Geochemistry Dataset, accessed through the Exploration for the Future Portal (https://portal.ga.gov.au/persona/eftf) in 2021. 5,813 samples from 3,317 sites broadly distributed across Australia were entered, with 262,417 analytical results. Samples were screened to remove igneous and metamorphic/altered samples based primarily on lithology, lithology descriptions/qualifiers, and stratigraphic unit names. The samples included are predominantly fine-grained (~30%) and coarser-grained (~39%) siliciclastics. Carbonates comprise about 16% of the samples. Most samples are Proterozoic–Paleozoic in age. Interpreted ages were entered primarily based on the Australian Stratigraphic Units Database (ASUD). A large number of samples (38%) do not have interpreted ages, and manual or third-party age coding in SGP Phase 3 would increase the amount of data available for Earth history studies. The data consist primarily of whole-rock major, minor, and trace elements measured by ICP-MS, XRF, and AAS. As with the USGS-CMIBS samples, given the working relationship between Geoscience Australia and the active mineral exploration industry in Australia, there is a likelihood that some samples have experienced a degree of mineralization.

3.3 Alberta Geological Survey

Six open-access datasets were added from the Alberta Geological Survey in 2022, chosen based on their alignment with SGP goals, including inorganic and organic geochemistry and mineralogical data. Most of the data came with detailed methodological information, in some cases including the labs where analyses were made and the experimental methods (e.g., acid digestion procedures). 4,192 samples were entered, with 332,226 analytical results, from 550 sites in Alberta, Canada. The samples are mostly fine-grained siliciclastic rocks (72%) with some sandstones (6%) and carbonates (13%). The samples are mostly Paleozoic and Mesozoic in age. Sampling is biased towards organic-rich fine-grained units (i.e., potential source rocks) that may generate petroleum.

3.4 DM-SED

The Deep-Time Marine Sedimentary Element Database (DM-SED) is a compilation study published in 2025 (Lai et al., 2025), with a similar focus to SGP. The project built on SGP Phase 1, with new data added from studies from a broad range of ages (Proterozoic and Phanerozoic), with global coverage and from studies dating back to the 1960s. Samples in DM-SED were divided into two projects: “New Compilation” and “SGP.” The “New Compilation” data were assembled by a team of researchers who collated data, extracted geological meta-data (including digitizing stratigraphic heights from older publications or tracking down latitude/longitude information from other studies) and developed sample-level age models. This diligent compilation work is nonetheless commonly less precise for meta-data like sample lithology than when coded directly by the original authors. Specifically in the case of DM-SED, samples are assigned a broad “LithType” category of “siliciclastic” or “carbonate”, with additional detail provided in the “LithName” field. However, for more than half of the ingested samples (56%), no “LithName” is available, and the lithology is therefore recorded only at the higher level. We used the DM-SED 0.0.01.csv download (V3). New compilation data that overlapped with SGP direct entry were first removed. We also excluded data from studies younger than 2017 or from Ocean Drilling Project/Deep Sea Drilling Project (ODP/DSDP) studies. We focused on older papers because 1) we hope to engage authors of more recently published studies in directly coding their data into the SGP database and becoming involved in the project, and 2) geochemical data and geological meta-data from older papers is less commonly digitized and harder to extract, and our intention is not to duplicate the efforts of the DM-SED project. We did not incorporate ODP/DSDP data because ongoing efforts to digitize all such data (for instance, Sessa et al., 2023) should allow for future bulk ingestion with consistent age models. In total, data ingestion from DM-SED incorporated 8,029 samples with 212,506 analytical results from 456 sites across the globe; samples have been associated with their source publications (123 publications total). Note the DM-SED dataset only incorporated a subset of published geochemical analytes (generally elemental abundances and light stable isotopes; refer to details in Lai et al. (2025)). Thus some samples/papers may be included in the SGP reference list and searchable as “projects” but some data, for instance metal isotopes or iron or phosphorus speciation data, are not present.

3.5 Consideration of data sources

All data in the SGP database have passed a minimum bar that we consider appropriate for studying Earth’s sedimentary carapace and its history. Nonetheless, all data sources have biases; for instance the direct entry from predominantly academic studies tend to be biased towards stratigraphic or biological events in the Phanerozoic, whereas the geological surveys are commonly tilted towards targets of economic interest (such as petroleum for the Alberta Geological Survey or ore deposits for OZCHEM). Careful consideration of data provenance, together with the use of available geological context and geochemical methodology data to filter samples or correct for sampling biases (Figure 1), will help improve the accuracy of scientific interpretations based on data from SGP. Note that samples can be included/excluded at the "Data Source" level in Detailed Search → Sample Filters → Data Source/Project on the search website. Ultimately, we believe that the best approach will be similar to those adopted by paleontological studies (Dunhill et al., 2012; Kidwell, 2005; Peters, 2005; Peters and Foote, 2001; Raja et al., 2022; Smith and Benson, 2013): all datasets are biased by geographical, sociological, or methodological collection approaches, and our goal is to figure out when biases do and do not matter, and then to statistically correct for them—as much as possible—in order to infer how the Earth surface has evolved through time (e.g., Mehra et al., 2021).

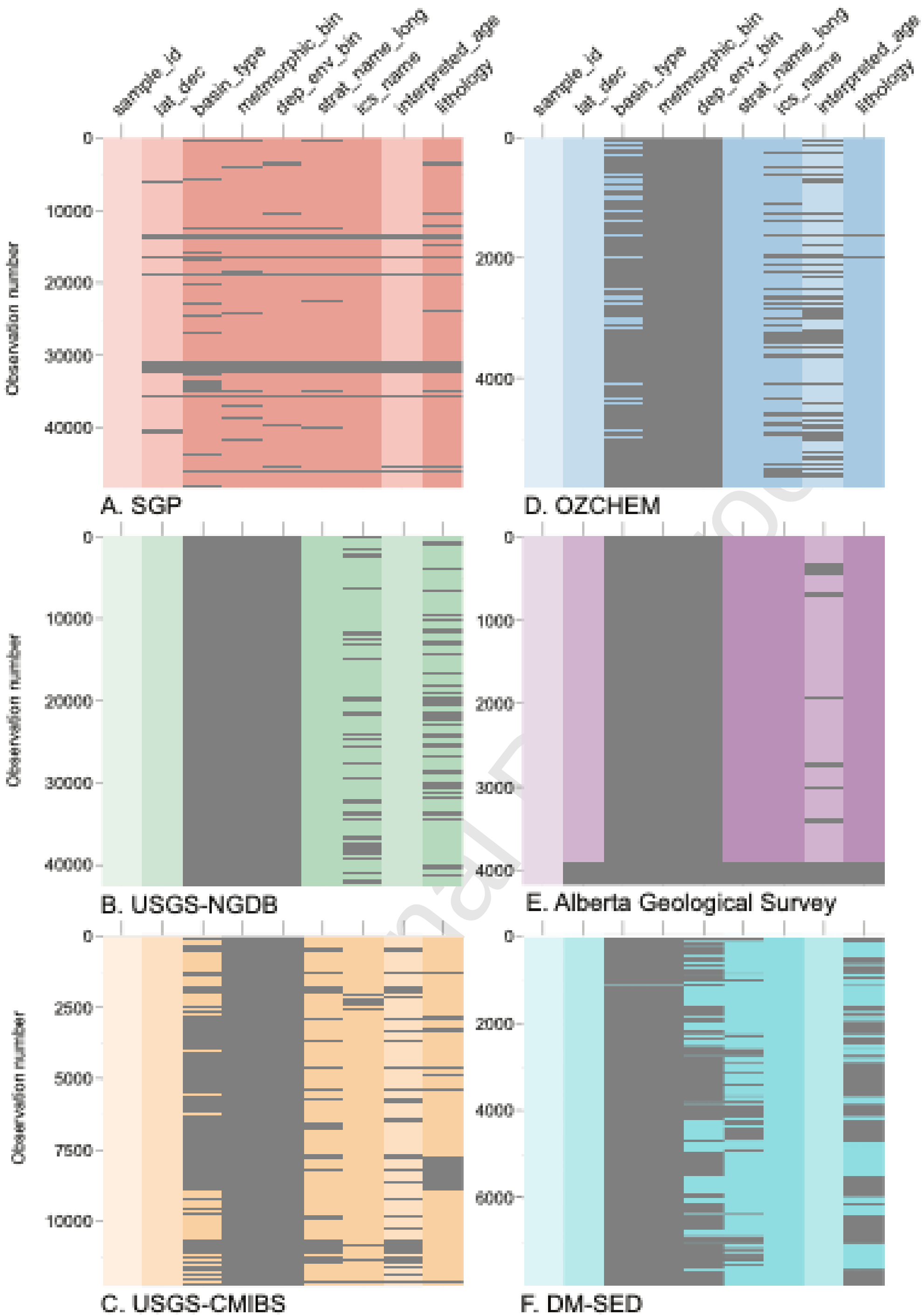


Figure 1: Completeness of geological context data for the six data sources compiled in the SGP Phase 2 data release including SGP (panel A, red, n = 48,154), USGS-NGDB (panel B, green, n = 42,585), USGS-CMIBS (panel C, orange, n = 12,276), OZCHEM (panel D, blue, n = 5,813), Alberta Geological Survey (AGS) (panel E, purple, n = 4,192), and DM-SED (panel F, aqua, n = 8,029) (acronyms in Table 1). From left to right, columns indicate the data field as written in the database, including sample ID, latitude in decimal degrees (lat_dec; all samples with latitude also

have longitude), metamorphic bin, depositional environment bin (dep_env_bin), lithostratigraphic unit (strat_name_long; i.e., formation), geologic period according to the International Commission on Stratigraphy time scale (ics_age), numerical age in millions of years (interpreted_age), and lithology. Each horizontal line (observations) corresponds to one individual sample from that data source. Colored bars indicate that context data are present; dark grey indicates that data are absent.

4. Phase 2 Data Description

The published and public Phase 2 database includes 126,006 samples. Of these, 121,047 samples have geochemical data (some samples are entered into the database but not yet linked to data), and 120,499 samples have geochemical data (4,132,371 analytical results) available through the search website (some unusual/less useful analytes have data stored in the database but not presented on the website). Note that the number of available samples/analyses accessible on the Phase 2 website will change slightly as data from land under U.S. Tribal jurisdiction becomes available (additional details below).

4.1 Site type

49% of samples are from outcrop, 43% are from core, 7% are unknown, and 1% are modern/other.

4.2 Lithology

As in Phase 1, the majority of samples in the SGP database are fine-grained siliciclastics (lithology names coded as: argillite, clay, claystone, marl, marlstone, meta-argillite, metapelite, metasiltstone, mud, mudstone, oil shale, pelite, silt, siltite, siltstone and shale; available as "Shales only" under the Simple Search website option) (Figure 2; 59% of samples with a lithology code). Coarse-grained siliciclastics comprise 10.7% for sandstones and 0.8% for conglomerates. Carbonate lithologies now comprise 19% of coded samples (including: lime mudstone, wackestone, packstone, grainstone, boundstone, crystalline limestone, dolomudstone, dolowackestone, dolopackstone, dolograinstone, crystalline dolomite, ooze, carbonate, limestone, dolomite, dolomicrite, limestone/dolomite; available as "Carbonates only" under the Simple Search website option). While the overall percentage of carbonate samples has only grown a little since Phase 1 (largely because the Alberta Geological Survey, OZCHEM, and DM-SED data sources brought in a large number of siliciclastic samples), we have substantially grown the absolute number of coded carbonate samples from 8,514 to 19,427. Additional lithologies with minor representation in the database include iron formation (0.29%), chert (1.35%), and phosphorite (1.7%). Additional details on lithology textures, modifiers, and composition are also available for many samples.

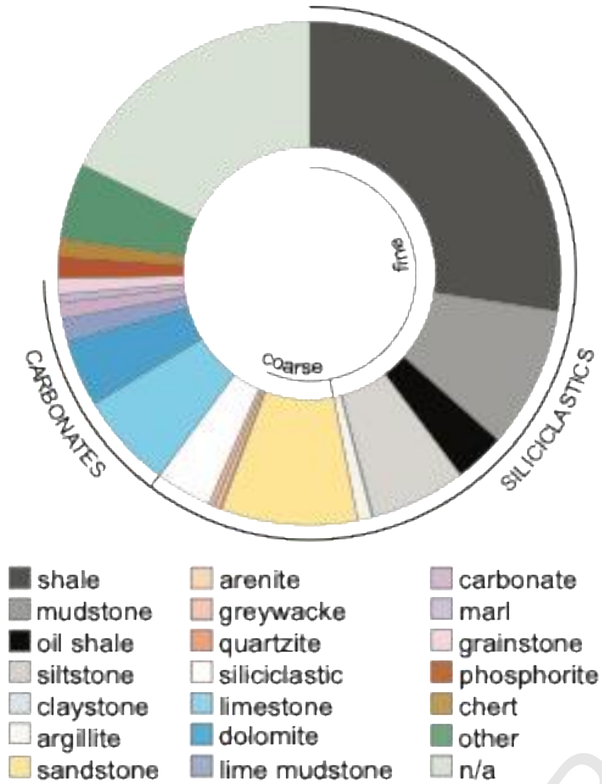


Figure 2: Distribution of lithologies in the SGP Phase 2 data release.

4.3 Geologic Age

Data collection in Phase 1 focused on the Neoproterozoic and Paleozoic, whereas Phase 2 focused on adding data primarily in the Paleoproterozoic, Mesoproterozoic, and lower Neoproterozoic. We have added substantial data holdings in deeper time, increasing the number of pre-Neoproterozoic samples from 5,023 to 18,082 (Figure 3). Most of these samples were through SGP direct entry, with accompanying rich geological context data, but the OZCHEM database also added a substantial number of Paleoproterozoic samples. We have also generally doubled the number of samples available in each Neoproterozoic time bin. Note that while the Phanerozoic is far better sampled than the Proterozoic, this is at least partly a function of the available rock record (e.g., Husson and Peters, 2017 and Fig. 3D) and in a qualitative sense we believe we have captured much of the published Proterozoic geochemical samples (at least for fine-grained siliciclastics). The Phase 2 data release has 2,811 Archean samples, 8,050 Paleoproterozoic samples, 7,221 Mesoproterozoic samples, 14,389 Neoproterozoic samples, 46,488 Paleozoic samples, 23,011 Mesozoic samples, and 17,137 Cenozoic samples.

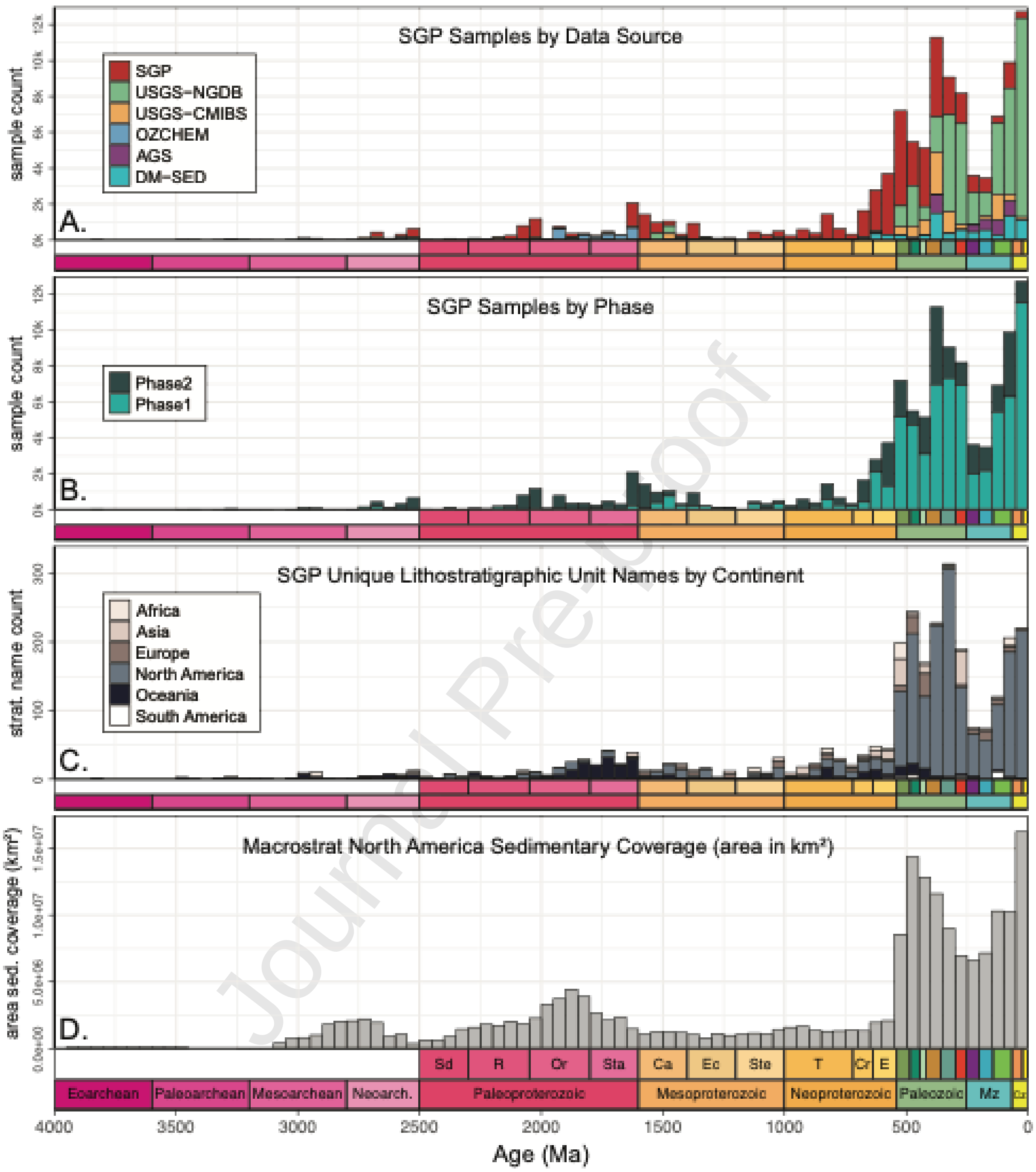


Figure 3: Plots through time of SGP samples, color-coded by data source (A) and phase (B); SGP unique lithostratigraphic unit names by continent (C); and, for comparison, area of North American sedimentary coverage from Macrostrat (D). Plot D created using the rmacrostrat package from Jones et al. (2024). Data are plotted in 50 million year bins. On plots A and B sample counts "k" = thousand, such that 2k = 2000 samples. Timescale abbreviations from left to right, top to bottom: Sd- Siderian, R- Rhyacian, Or- Orosirian, Sta- Statherian, Ca- Calymmian, Ec- Ectasian, Ste- Stenian, T- Tonian, Cr- Cryogenian, E- Ediacaran, Neoarch.-Neoarchean, Mz-Mesozoic, Cz- Cenozoic. Phanerozoic periods not labelled but follow the standard International Commission on Stratigraphy color scheme.

4.4 Geographic sampling

Geographic sampling in the database remains heavily biased towards North America. This is due to the ingestion of data from the USGS-NGDB and Alberta Geological Survey (both of which are regionally focused) and the fact that, like in other fields of Earth science (e.g., Raja et al., 2022) socio-political factors have led to substantial data generation in that region. There is also the sociological bias that SGP organizers are based in North America and are more familiar with North American researchers and data. Australia now represents a sampling hotspot due to data from OZCHEM, USGS-CMIBS, and SGP direct entry. Substantial sampling from China is present through SGP direct entry and DM-SED (note this is primarily focused on Mesoproterozoic, Neoproterozoic, and Cambrian data, with relatively little post-Cambrian data). While these regions (plus northern Europe) represent the most densely sampled geographic regions, global sampling has expanded substantially over Phase 2. This has been driven, at least in part, by an increased geographic diversity of SGP Collaborative Team members from Africa, South America, and India. Despite this growth in global sampling, we recognize that the complete stratigraphic column in many areas worldwide is poorly sampled in the SGP database.

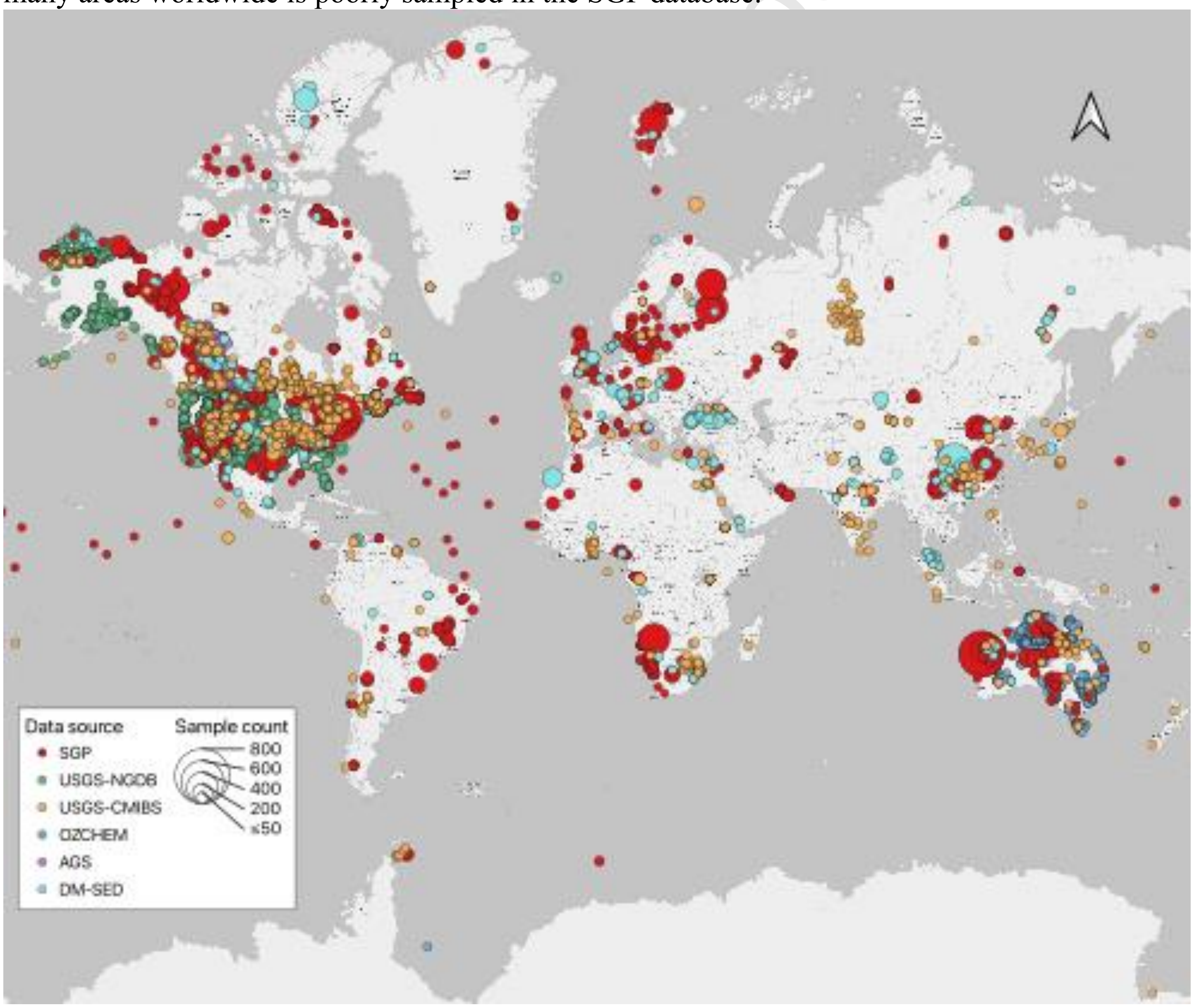


Figure 4: SGP Phase 2 geographic coverage. Samples are from 19,734 sites in 84 countries/oceans. The size of the dot at each site scales to the number of samples and is color-coded to data source.

4.5 Sampling completeness through time

Combining data from Figure 3 (temporal sampling) and Figure 4 (geographic sampling) allows us to analyze global sampling completeness through time in the SGP database. Following Jones and Eichenseer (2021), we evaluated two metrics of sampling completeness: the number of occupied equal-area grid cells (100 kilometer spacings) and the length of the Minimum Spanning Tree (MST; the convex hull encompassing sampling points) (Figure 5). Over broader timescales (last 2500 million years) the global sampling pattern primarily reflects macrostratigraphic trends in the rock record (Husson and Peters, 2017), with substantially greater volumes of Phanerozoic strata more highly sampled by geologists and geochemists. Nonetheless, the Phase 2 data release now includes relatively even sampling across the Proterozoic, allowing for increasingly robust analyses across billion-year timescales. Notable global sampling gaps exist in the Siderian, upper Rhyacian, and Ectasian–Stenian.

Both sampling metrics reflect our focus on the Neoproterozoic–Paleozoic in Phase 1 and Mesoproterozoic-Paleoproterozoic in Phase 2. Consequently, the best-sampled interval remains the lower Paleozoic. In Phanerozoic-scale sampling completeness plots, both metrics show a low around the Silurian–Devonian transition. The number of occupied grid cells then increases dramatically over the upper Devonian and Carboniferous, but without a substantial rise in MST. This is due to substantial sampling in North America and Europe with relatively little global sampling. The Triassic, Jurassic, and Lower Cretaceous are relatively poorly sampled globally for both metrics. The middle and Upper Cretaceous are better sampled (especially compared to Phase 1) due to increased global data from SGP direct entry and the Alberta Geological Survey data. The Paleogene represents an additional interval with relatively poor global sampling.

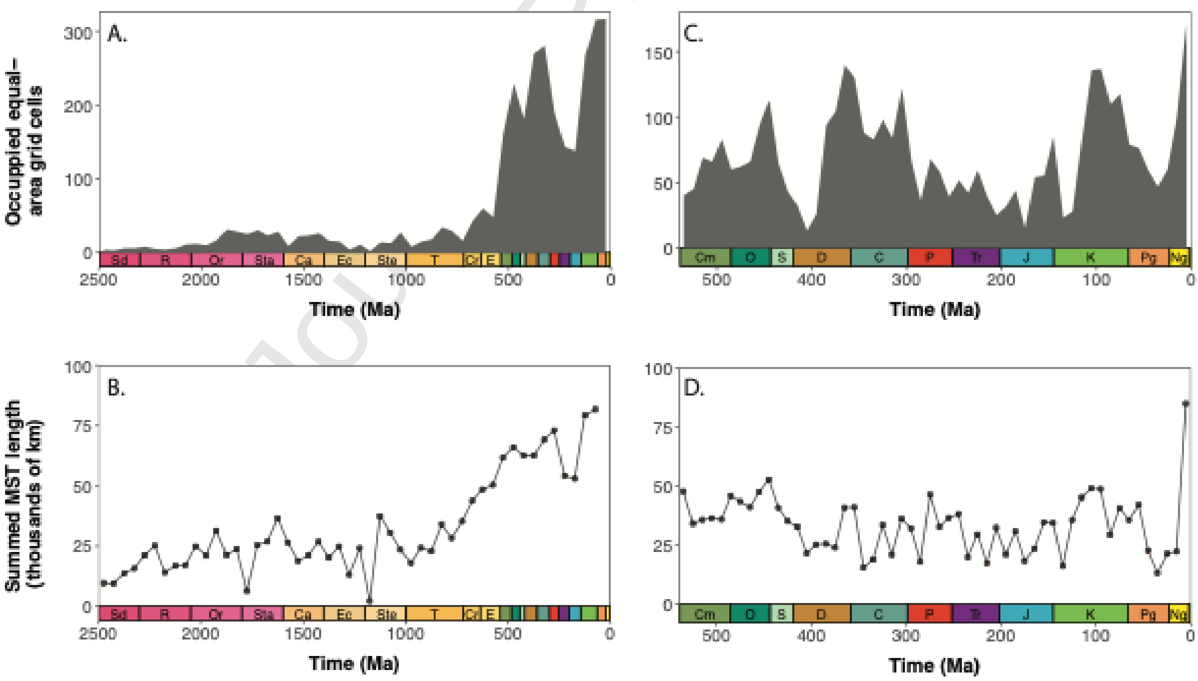


Figure 5: Plots of geographic sampling in the SGP Phase 2 data release through time for the last 2500 million years in 50 million year bins (A and B) and Phanerozoic in 10 million year bins (C and D). Number of occupied equal area grid cells (A and C) shows the number of cells sampled but may be geographically restricted (e.g., North America). The Summed Minimum Spanning Tree (MST, B and D) is the length of the convex hull encompassing all sampling points. Proterozoic

time scale abbreviations as in Fig. 4. For Phanerozoic abbreviations: Cm- Cambrian, O- Ordovician, S- Silurian, D- Devonian, C- Carboniferous, P- Permian, Tr- Triassic, J- Jurassic, K- Cretaceous, Pg- Paleogene, Ng- Neogene.

5. Data Presentation and Access

The SGP search website (https://sgp-search.io) is hosted on Amazon Web Services (AWS) (front-end: React.js, back-end: Node.js). SGP data products (back-end) are exposed to front-end users via a REST-ful API and using industry-standard JSON format for query and response. Our website was substantially revised for the Phase 2 data release in direct response to user feedback at our SGP Collaborative Team meetings. Specifically, we have made basic searches more straightforward and enhanced custom searches to access full methodological details. There are now three available search options:

5.1 Simple Search

This search is designed to provide quick access to data, with pre-set filters on sample and geological meta-data. Samples are grouped based on lithology (refer to Phase 2 Data Description above), allowing users to choose between "All Samples," "Shales only," or "Carbonates only." Users can then narrow their search based on age range or geographic location (the two most common search criteria) and choose geochemical analytes of interest. The query is returned with pre-set geological metadata. In this search type the units are standardized, replicate results are averaged, and oxides are converted to elements.

5.2 Detailed Search

This option gives users complete control over their search and is more similar to the Phase 1 website interface. Users can filter based on:

- 22 parameters related to geological and geographic context
- Ranges of values for a given analyte
- Specific methodologies (for instance, only including trace metals analyzed by ICP-MS).
- Data source or Project (samples grouped together by publication or data ingestion source, such as SGP direct entry or OZCHEM).

Users can then select geochemical analytes of interest and chose exactly what geological meta-data to include in the data export. Note that the "No HHXRF" search type, a variant of the Detailed Search in Phase 1 with data from handheld XRF excluded, no longer exists as this functionality can be achieved with the new methodology search features. As with Simple Search, units are standardized, replicate results are averaged, and oxides are converted to elements.

5.3 Analyses Search

This option allows users to access individual analytical results exactly as published or provided (i.e. units are not standardized, replicate results are not averaged, and oxides are not converted to elements). Analytes and units appear as values in rows, rather than as column headers.

A major upgrade during Phase 2 is the ability for users to filter and export based on analytical methods (available in Detailed and Analyses searches; refer to https://github.com/ufarrell/sgp_phase2/wiki/C.-Analyses) for a summary of Phase 2 data distribution by method). This is critical as specific methodologies can result in false interpretations of geochemical signals (for instance, grinding vessel contamination (Hickson and Juras, 1986) or the strength and type of acid in carbonate dissolution (Cao et al., 2020; Clarkson et al., 2020; Liu et al., 2018; Zhang et al., 2024)). Using standardized vocabularies allows for further filtering, sorting, or automated data correction in programs such as Python or R, or for weighting and/or inclusion as a predictor variable in multivariate or machine learning analyses.

5.4 Indigenous ancestral lands

Many samples in the SGP database have come from Indigenous ancestral lands around the world. While the geographic boundaries of ancestral lands are difficult to determine, SGP has instituted the ability on our website to access information from the Native Land Digital website (https://native-land.ca) as a starting point. While Native Land Digital is considered the most comprehensive worldwide reference for such information, it is neither complete nor a legal reference nor recognized for use in government-to-government consultation efforts. As such, this functionality is provided for the use and education of SGP website users (and is not used in Section 6 on Indigenous data governance and stewardship). The data can be displayed in the data table and included in data downloads via Show → Samples Context and clicking the "traditional territory" button.

5.5 Data visualization and export: After users have completed their search, they have the option to explore their query results in a data table. This table is sortable based on high/low values for a column, has expandable menus with methodology details, and has a geological context info button for each sample. The user can also copy the API call for future re-use or to include in a paper as a reproducibility step. Data downloads are returned as a .csv file. Finally, the user can explore the queried data on a new dynamic map feature that is color-coded to the geological timescale. Users can zoom in to the site level and easily access the geological context of samples from that site.

Note that since the Simple and Detailed searches average replicate results on the same analyte, measurements above (the maximum) or below (the minimum) detection limits will be removed, as they cannot be averaged. This is important for studies of analytes whose abundances are near detection limits for common geochemical methodologies—in the case of low-abundance elements, most analyses (i.e., below detection) will be removed, and only anomalously 'high' analyses will be returned. The following analytes have more than 25% of analyses below detection limits: Ta-carb, Ir, Pd, Sb-carb, Pt, Zr-carb, Os, Au, Re, In, Te, Bi, Ge, Ag, Re-carb, Sn, Cd, Ta, W, As-carb, Tl, Eu, Ho, Sb, Be, As, Hf, Pr, Hg, Hf-carb, Se, Sm, Li, Se-carb, Nb, Th, Er, B, Tb, U, S-org, Mo, Tm, S-SO4, Dy (listed from highest proportion below detection, Ta-carb at 100%, to Dy at 25.2%; more details can be found on the SGP wiki (https://github.com/ufarrell/sgp_phase2/wiki/C.-Analyses#abovebelow-detection). We do not imply that statistical analyses of such analytes is inappropriate; rather that users should be aware of this specific feature of the Simple and Detailed data export.

6. Indigenous Data Governance and Stewardship

SGP acknowledges that many of the lands and waters where we live, where geological samples are collected, and where geochemical measurements are conducted may be on the present-day and ancestral lands of Indigenous People. We recognize the continued significance of these lands and waters for Indigenous Peoples since time immemorial and are cognizant of the need for proper and respectful stewardship of these data.

In some cases, geological data have been collected from Indigenous land worldwide without permission or consultation (Kempf et al., 2023; Monarrez et al., 2022). Due to the long history of geoscience data collection, large quantities of such data exist in legacy datasets that persist in repositories. The absence of Indigenous stewardship of geologic data can directly impact community health (Moore-Nall, 2015), economic stability (Na-Cho Nyäk Dun elders et al., 2019), territorial control (Goodman, 2018), and/or continued perpetuation of past injustices (Eichstaedt, 1994). The Global Indigenous Data Alliance (GIDA) established four principles for Indigenous data governance: Collective benefit, Authority to control, Responsibility, and Ethics (CARE) (Carroll et al., 2021, 2020). Since the initial establishment of the CARE principles, there has been increased interest in the proper application of these principles in different branches of science, such as biology (Jennings et al., 2023), genomics (Carroll et al., 2022), archaeology (Gupta et al., 2023), and Earth Science (Jennings et al., 2025; O'Brien et al., 2024). Applying CARE principles to a compilation dataset (like SGP) first involves understanding which data are from Indigenous lands, then contacting the appropriate Indigenous group (O'Brien et al., 2024). As outlined in Hudson et al. (2023), even if data from Indigenous lands were ingested from a public source where permission was granted, publication and/or presentation of that data in a new context should have new approval.

Our first efforts here are focused on the United States, where SGP is based, and we have given specific attention to lands under the jurisdiction of federally recognized Tribes. Such lands include reservations and trust lands, as well as Oklahoma Tribal Statistical Areas, some of which have been or are becoming recognized as reservations in the wake of *McGirt* v. *Oklahoma* 140 S. Ct. 2452 (2020) and its progeny. Using public TIGER/Line shape files provided by the U.S. Census Bureau, we have identified 3,814 samples (2.94% of the 129,820 samples we had originally compiled), from 743 geographic sites, with 134,165 geochemical analytes, where the recorded SGP latitude and longitude site-matched with lands under Tribal jurisdiction. These 3,814 samples were ingested from four distinct publicly available data sources including the USGS-NGDB database (91%), third-party literature compilations by USGS-CMIBS (6%) and DM-SED (1%), and SGP direct entry from published literature (2%). Note that many of these same samples/data are also held by other non-governmental Earth Science repositories such as EarthChem and Macrostrat. Even though contributions from United States Government agencies, academia, and industry may have been obtained in accordance with any applicable laws and policies in place at the time of collection/reception, these data will now be presented in this new context of the SGP search website and API. Therefore, we have taken the proactive step of removing these data from the Phase 2 release.

6.2 Future SGP goals regarding CARE principles

Moving forward, we have integrated a PostGIS-based spatial check into our data ingestion pipeline to flag new samples that may come from lands under U.S. Tribal jurisdiction. Our next step will be detailed vetting of the samples that site-matched to lands under U.S. Tribal jurisdiction to determine whether this was the likely point of origin (note that SGP samples have varying degrees of geographic precision; refer to Farrell et al. (2021)). With our finalized list of samples, we will notify Tribal Nations of our intent to include data about the samples on the SGP website. If a Tribe wants to restrict (in whole or in part) any data from the SGP website, we will abide by the Tribe's decision. We will also seek the Tribe's input for future data ingestion and presentation. While the products of these consultations will be incorporated, we recognize that individual Native Nations' needs and opinions on data holdings can and will change over time and continued open communication is necessary for proper stewardship.

We are cognizant that the proposed work here represents just one step towards implementation of CARE principles. Many complex issues remain, including treatment/stewardship of data from ancestral lands, Indigenous data from outside the United States, and data handling considerations prior to and during consultation (refer to Carroll et al., 2021; O'Brien et al., 2024). SGP is committed to developing new modes of collaboration, engagement, and partnership with Indigenous peoples for the care and stewardship of past and future heritage collections.

7. Future Goals and Directions

Many of our future goals and directions for SGP Phase 3 flow from Figures 1-5. With respect to lithologies, we now have a database and methodological structure appropriate for carbonate data but only a small fraction of the carbonate geochemical data that have been generated by the geochemical community to date has been entered. Expanding data holdings in carbonate geochemistry is thus a top goal. From a temporal perspective, the Mesozoic–Cenozoic has limited geographic sampling compared to the Paleozoic and is dominated by North American samples. This means that Phanerozoic-scale studies (and geochemical comparisons with the fossil record across this interval) have higher potential for spatial and temporal sampling imbalances. Building the Archean through Rhyacian and Ectasian–Stenian records are secondary temporal goals. With respect to analytical data types, database revisions (Section 2.1) have greatly enhanced our ability to usefully input and export trace metal isotope data, and these analytes will be a focus in Phase 3. Including organic geochemical (biomarker) data is an aspirational goal but would require substantial database updates and extensive community consultation to ensure SGP is appropriately capturing critical data and methodological details. These lithological, analytical, and temporal goals will require proactive outreach to new research communities, specifically carbonate geochemists, organic geochemists, and researchers working in the Archean and Mesozoic–Cenozoic, as researchers at both ends of the geological timescale commonly specialize temporally due to the unique features of those records. We will also continue to seek external data sources for bulk import and integration. In particular, we are interested in engaging and collaborating with national/regional geological surveys worldwide. In addition to our planned proactive approaches, we encourage any researchers in these areas to reach out directly and join SGP. This is an open, community-driven project and we welcome your involvement. Finally, Figure 4 illustrates that we have considerable work ahead to build a globally comprehensive database. We are planning regional outreach events over Phase 3 in Africa, Asia, and South America, and again we welcome and encourage geologists and geochemists in these regions to reach out to us about accessioning data and joining the SGP community.

8. Data license and Citation

The SGP Phase 2 data release is available under a Creative Commons BY 4.0 International license. Users are free to share (copy and redistribute the material in any medium or format for any purpose) or adapt (remix, transform, and build upon the material for any purpose) the data and database if attribution and credit is provided. Such attribution and credit to the geoscientists that generated these data and provided relevant geological meta-data is given through citation of this manuscript.

Acknowledgements

We thank Chrissy Garcia, Lydia Jennings, Chris Hammond, and Laramie Duncan for helpful discussion. We thank Kassie Sharp and John Freshwaters for continued support of SGP geoinformatics infrastructure at Stanford University. SGP Phase 2 was funded by National Science Foundation grants 2143164 to EAS and 2311092 to EAS, SEP, and DPQ. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Table 1. Acronyms used in this manuscript.

Acronym	**Definition/explanation**
OZCHEM	The Geoscience Australia data product. Sedimentary geochemical data from this database were ingested during SGP Phase 2
DM-SED	A third-party compilation of geochemical data (refer to Lai et al., 2025). Older (pre-2017) data from this compilation were ingested during SGP Phase 2
API	Application Programming Interface: a software intermediary that enables two programs to communicate using a set of definitions and protocols. Data in SGP can be accessed directly through our API, or from our search website, which uses an API call to retrieve requested data.
FAIR	Principles for management and stewardship of scientific data: data should be Findable, Accessible, Interoperable, and Reusable (refer to Wilkinson et al., 2016)
ICP-MS	Inductively Coupled Plasma Mass Spectrometry
LA-ICP-MS	Laser Ablation Inductively Coupled Plasma Mass Spectrometry
USGS-CMIBS	The Critical Metals in Black Shale project from the United States Geological Survey (USGS). A third-party compilation of shale geochemical data that was ingested during SGP Phase 1
USGS-NGDB	The USGS National Geochemical Database, storing geochemical data from decades of USGS projects; ingested during SGP Phase 1
XRF	X-ray Fluorescence
AAS	Atomic Absorption Spectroscopy
CARE	Principles for management and stewardship of Indigenous data: data must facilitate Collective benefit for Indigenous groups, Indigenous groups must have Authority to control, there is a Responsibility to nurture respectful relationships, and be collected and managed in an Ethics framework that focuses on Indigenous Peoples' rights and well-being (refer to Carroll et al., 2020).
TIGER/Line	Topologically Integrated Geographic Encoding and Referencing - U.S. Census Bureau format for geospatial data.

References

Alcott, L.J., Bowyer, F.T., Agić, H., 2025. Future directions for understanding the coevolution of life and oxygen. Commun Earth Environ 6, 725. https://doi.org/10.1038/s43247-025-02689-0

Andersen, M.B., Stirling, C.H., Weyer, S., 2017. Uranium Isotope Fractionation. Reviews in Mineralogy and Geochemistry 82, 799–850. https://doi.org/10.2138/rmg.2017.82.19

Bian, L., Zhao, Z., Wang, X., Sanei, H., Chappaz, A., Dong, J., Dong, Z., Xie, L., Schovsbo, N.H., Goodarzi, F., Zhang, S., Zhao, W., 2025. Degradation of organic matter by radioactive radiation in black shales: An overlooked modification of organic molecular structures. International Journal of Coal Geology 309, 104864. https://doi.org/10.1016/j.coal.2025.104864

Bishop, B.A., Robbins, L.J., 2024. Using machine learning to identify indicators of rare earth element enrichment in sedimentary strata with applications for metal prospectivity. Journal of Geochemical Exploration 258, 107388. https://doi.org/10.1016/j.gexplo.2024.107388

Bubphamanee, K., Kipp, M.A., Meixnerová, J., Stüeken, E.E., Ivany, L.C., Bartholomew, A.J., Algeo, T.J., Brocks, J.J., Dahl, T.W., Kinsley, J., Tissot, F.L.H., Buick, R., 2025. Mid-Devonian ocean oxygenation enabled the expansion of animals into deeper-water habitats. Proceedings of the National Academy of Sciences 122, e2501342122. https://doi.org/10.1073/pnas.2501342122

Canfield, D.E., Zhang, S., Mitchell, R.N., Wang, X., Naemi, A., Spencer, C.J., Zhang, P., 2025. Long-term history of continental weathering and particle transport to the sea. Proceedings of the National Academy of Sciences 122, e2507312122. https://doi.org/10.1073/pnas.2507312122

Cao, C., Liu, X.-M., Bataille, C.P., Liu, C., 2020. What do Ce anomalies in marine carbonates really mean? A perspective from leaching experiments. Chemical Geology 532, 119413. https://doi.org/10.1016/j.chemgeo.2019.119413

Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J., Hudson, M., 2020. The CARE Principles for Indigenous Data Governance. Data Science Journal 19, 43. https://doi.org/10.5334/dsj-2020-043

Carroll, S.R., Herczog, E., Hudson, M., Russell, K., Stall, S., 2021. Operationalizing the CARE and FAIR Principles for Indigenous data futures. Sci Data 8, 108. https://doi.org/10.1038/s41597-021-00892-0

Carroll, S.R., Plevel, R., Jennings, L.L., Garba, I., Sterling, R., Cordova-Marks, F.M., Hiratsuka, V., Hudson, M., Garrison, N.A., 2022. Extending the CARE Principles from tribal research policies to benefit sharing in genomic research. Front. Genet. 13, 1052620. https://doi.org/10.3389/fgene.2022.1052620

Chamberlain, K.J., Lehnert, K.A., McIntosh, I.M., Morgan, D.J., Wörner, G., 2021. Time to change the data culture in geochemistry. Nat Rev Earth Environ 2, 737–739. https://doi.org/10.1038/s43017-021-00237-w

Clarkson, M.O., Müsing, K., Andersen, M.B., Vance, D., 2020. Examining pelagic carbonate-rich sediments as an archive for authigenic uranium and molybdenum isotopes using reductive cleaning and leaching experiments. Chemical Geology 539, 119412. https://doi.org/10.1016/j.chemgeo.2019.119412

Cui, M., Luther, G.W., Gomes, M., 2023. Constraining the major pathways of vanadium incorporation into sediments underlying natural sulfidic waters. Geochimica et Cosmochimica Acta 359, 148–164. https://doi.org/10.1016/j.gca.2023.08.008

Dunhill, A.M., Benton, M.J., Twitchett, R.J., Newell, A.J., 2012. Completeness of the fossil record and the validity of sampling proxies at outcrop level. Palaeontology 55, 1155–1175. https://doi.org/10.1111/j.1475-4983.2012.01149.x

Eichstaedt, P., 1994. If You Poison Us: Uranium and Native Americans. Red Crane Books, Santa Fe, N.M.

Emmings, J.F., Poulton, S., Walsh, J., Leeming, K.A., Ross, I., Peters, S.E., 2022. Pyrite mega-analysis reveals modes of anoxia through geological time. Science Advances 8, eabj5687.

Ernst, D.M., Garbe-Schönberg, D., Kraemer, D., Bau, M., 2023. A first look at the gallium-aluminium systematics of Early Earth's seawater: Evidence from Neoarchean banded iron formation. Geochimica et Cosmochimica Acta 355, 48–61. https://doi.org/10.1016/j.gca.2023.06.019

Farrell, Ú.C., Samawi, R., Anjanappa, S., Klykov, R., Adeboye, O.O., Agic, H., Ahm, A.-S.C., Boag, T.H., Bowyer, F., Brocks, J.J., Brunoir, T.N., Canfield, D.E., Chen, X., Cheng, M., Clarkson, M.O., Cole, D.B., Cordie, D.R., Crockford, P.W., Cui, H., Dahl, T.W., Mouro, L.D., Dewing, K., Dornbos, S.Q., Drabon, N., Dumoulin, J.A., Emmings, J.F., Endriga, C.R., Fraser, T.A., Gaines, R.R., Gaschnig, R.M., Gibson, T.M., Gilleaudeau, G.J., Gill, B.C., Goldberg, K., Guilbaud, R., Halverson, G.P., Hammarlund, E.U., Hantsoo, K.G., Henderson, M.A., Hodgskiss, M.S.W., Horner, T.J., Husson, J.M., Johnson, B., Kabanov, P., Keller, C.B., Kimmig, J., Kipp, M.A., Knoll, A.H., Kreitsmann, T., Kunzmann, M., Kurzweil, F., LeRoy, M.A., Li, C., Lipp, A.G., Loydell, D.K., Lu, X., Macdonald, F.A., Magnall, J.M., Mänd, K., Mehra, A., Melchin, M.J., Miller, A.J., Mills, N.T., Mwinde, C.N., O'Connell, B., Och, L.M., Ossa, F.O., Pagès, A., Paiste, K., Partin, C.A., Peters, S.E., Petrov, P., Playter, T.L., Plaza-Torres, S., Porter, S.M., Poulton, S.W., Pruss, S.B., Richoz, S., Ritzer, S.R., Rooney, A.D., Sahoo, S.K., Schoepfer, S.D., Sclafani, J.A., Shen, Y., Shorttle, O., Slotznick, S.P., Smith, E.F., Spinks, S., Stockey, R.G., Strauss, J.V., Stüeken, E.E., Tecklenburg, S., Thomson, D., Tosca, N.J., Uhlein, G.J., Vizcaíno, M.N., Wang, H., White, T., Wilby, P.R., Woltz, C.R., Wood, R.A., Xiang, L., Yurchenko, I.A., Zhang, T., Planavsky, N.J., Lau, K.V., Johnston, D.T., Sperling, E.A., 2021. The Sedimentary Geochemistry and Paleoenvironments Project. Geobiology 19, 545–556. https://doi.org/10.1111/gbi.12462

Goodman, D., 2018. Gold and the Public in the Nineteenth-Century Gold Rushes, in: Mountford, B., Tuffnell, S. (Eds.), A Global History of Gold Rushes. University of California Press, pp. 65–87. https://doi.org/10.1525/9780520967588-007

Granitto, M., Giles, S.A., Kelley, K.D., 2017. Global Geochemical Database for Critical Metals in Black Shales. https://doi.org/10.5066/F71G0K7X

Gupta, N., Martindale, A., Supernant, K., Elvidge, M., 2023. The CARE Principles and the Reuse, Sharing, and Curation of Indigenous Data in Canadian Archaeology. Advances in Archaeological Practice 11, 76–89. https://doi.org/10.1017/aap.2022.33

Hantsoo, K., Gomes, M., Brenner, D., Cornwell, J., Palinkas, C.M., Malkin, S., 2024. Trends in estuarine pyrite formation point to an alternative model for Paleozoic pyrite burial. Geochimica et Cosmochimica Acta 374, 51–71. https://doi.org/10.1016/j.gca.2024.04.018

Hickson, C.J., Juras, S.J., 1986. Sample contamination by grinding. Canadian Mineralogist 24, 585–589.

Husson, J.M., Peters, S.E., 2017. Atmospheric oxygenation driven by unsteady growth of the continental sedimentary reservoir. Earth and Planetary Science Letters 460, 68–75. https://doi.org/10.1016/j.epsl.2016.12.012

Hudson, M., Carroll, S.R., Anderson, J., Blackwater, D., Cordova-Marks, F.M., Cummins, J., David-Chavez, D., Fernandez, A., Garba, I., Hiraldo, D., Jäger, M.B., Jennings, L.L., Martinez, A., Sterling, R., Walker, J.D., Rowe, R.K., 2023. Indigenous Peoples' Rights in Data: a contribution toward Indigenous Research Sovereignty. Front. Res. Metr. Anal. 8, 1173805. https://doi.org/10.3389/frma.2023.1173805

Jennings, L., Anderson, T., Martinez, A., Sterling, R., Chavez, D.D., Garba, I., Hudson, M., Garrison, N.A., Carroll, S.R., 2023. Applying the 'CARE Principles for Indigenous Data Governance' to ecology and biodiversity research. Nat Ecol Evol 7, 1547–1551. https://doi.org/10.1038/s41559-023-02161-2

Jennings, L., Jones, K., Taitingfong, R., Martinez, A., David-Chavez, D., Alegado, R. 'Anolani, Tofighi-Niaki, A., Maldonado, J., Thomas, B., Dye, D., Weber, J., Spellman, K.V., Ketchum, S., Duerr, R., Johnson, N., Balch, J., Carroll, S.R., 2025. Governance of Indigenous data in open earth systems science. Nat Commun 16, 572. https://doi.org/10.1038/s41467-024-53480-2

Jones, L.A., Dean, C.D., Gearty, W., Allen, B.J., 2024. rmacrostrat: An R package for accessing and retrieving data from the Macrostrat geological database. Geosphere 20, 1456–1467. https://doi.org/10.1130/GES02815.1

Jones, L.A., Eichenseer, K., 2021. Uneven spatial sampling distorts reconstructions of Phanerozoic seawater temperature. Geology 50, 238–242. https://doi.org/10.1130/G49132.1

Kempf, H.L., Olson, H.C., Monarrez, P.M., Bradley, L., Keane, C., Carlson, S.J., 2023. History of Native American land and natural resource policy in the United States: impacts on the field of paleontology. Paleobiology 49, 191–203. https://doi.org/10.1017/pab.2022.41

Kidwell, S.M., 2005. Shell composition has no net impact on large-scale evolutionary patterns in mollusks. Science 307, 914–917.

Kimmig, J., Pratt, B.R., 2022. Evidence for microbially mediated silver enrichment in a middle Cambrian Burgess Shale-type deposit, Mackenzie Mountains, northwestern Canada. Canadian Journal of Earth Sciences 59, 123–133.

Lai, J., Song, Haijun, Chu, D., Dal Corso, J., Sperling, E.A., Wu, Y., Liu, X., Wei, L., Li, M., Song, Hanchen, Du, Y., Jia, E., Feng, Y., Song, Huyue, Yu, W., Liang, Q., Li, X., Yao, H., 2025. Deep-Time Marine Sedimentary Element Database. Earth System Science Data 17, 1613–1626. https://doi.org/10.5194/essd-17-1613-2025

Lipp, A.G., Shorttle, O., Sperling, E.A., Brocks, J.J., Cole, D.B., Crockford, P.W., Del Mouro, L., Dewing, K., Dornbos, S.Q., Emmings, J.F., Farrell, U.C., Jarrett, A., Johnson, B.W., Kabanov, P., Keller, C.B., Kunzmann, M., Miller, A.J., Mills, N.T., O'Connell, B., Peters, S.E., Planavsky, N.J., Ritzer, S.R., Schoepfer, S.D., Wilby, P.R., Yang, J., 2021. The composition and weathering of the continents over geologic time. Geochem. Persp. Let. 21–26. https://doi.org/10.7185/geochemlet.2109

Liu, C., Wang, Z., Macdonald, F.A., 2018. Sr and Mg isotope geochemistry of the basal Ediacaran cap limestone sequence of Mongolia: Implications for carbonate diagenesis, mixing of glacial meltwaters, and seawater chemistry in the aftermath of Snowball Earth. Chemical Geology 491, 1–13. https://doi.org/10.1016/j.chemgeo.2018.05.008

Mehra, A., Keller, C., Zhang, T., Tosca, N., McLennan, S., Sperling, E., Farrell, U., Brocks, J., Canfield, D., Cole, D., Crockford, P., Cui, H., Dahl, T., Dewing, K., Emmings, J., Gaines, R., Gibson, T., Gilleaudeau, G., Guilbaud, R., Hodgkiss, M., Jarrett, A., Kabanov, P., Kunzmann, M., Li, C., Loydell, D., Lu, X., Miller, A., Mills, N., Mouro, L., O'Connell, B., Peters, S., Poulton, S., Ritzer, S., Smith, E., Wilby, P., Woltz, C., Strauss, J., 2021. Curation and Analysis of Global Sedimentary Geochemical Data to Inform Earth History. GSAT 31, 4–10. https://doi.org/10.1130/GSATG484A.1

Monarrez, P.M., Zimmt, J.B., Clement, A.M., Gearty, W., Iii, J.J.J., Jenkins, K.M., Kusnerik, K.M., Poust, A.W., Robson, S.V., Sclafani, J.A., Stilson, K.T., Tennakoon, S.D., Thompson, C.M., 2022. Our past creates our present: a brief overview of racism and colonialism in Western paleontology. Paleobiology 48, 173–185. https://doi.org/10.1017/pab.2021.28

Moore-Nall, A., 2015. The Legacy of Uranium Development on or Near Indian Reservations and Health Implications Rekindling Public Awareness. Geosciences 5, 15–29. https://doi.org/10.3390/geosciences5010015

Murray, J., Jagoutz, O., 2024. Palaeozoic cooling modulated by ophiolite weathering through organic carbon preservation. Nat. Geosci. 17, 88–93. https://doi.org/10.1038/s41561-023-01342-9

Na-Cho Nyäk Dun elders, Hogan, J., Saxinger, G., Gartler, S., 2019. Dän Húnày. Our peoples' story: First Nation of Nacho Nyäk Dun elders' memories and opinions on mining. First Nation of Nacho Nyäk Dun / ReSDA - Resources and Sustainable Development in the Arctic / Yukon College, Mayo.

O'Brien, M., Duerr, R., Taitingfong, R., Martinez, A., Vera, L., Jennings, L.L., Downs, R.R., Antognoli, E., Brink, T. ten, Halmai, N.B., David-Chavez, D., Carroll, S.R., Hudson, M., Buttigieg, P.L., 2024. Earth Science Data Repositories: Implementing the CARE Principles. Data Science Journal 23, 37. https://doi.org/10.5334/dsj-2024-037

Olson, H.C., Scheirer, A.H., Ritzer, S.R., Sperling, E.A., 2025. Prediction of organic geochemical parameters from inorganic geochemical data in the Cretaceous-Danian Moreno Formation, San Joaquin Basin, California. Chemical Geology 674, 122551. https://doi.org/10.1016/j.chemgeo.2024.122551

Park, R., Epstein, S., 1960. Carbon isotope fractionation during photosynthesis. Geochimica et Cosmochimica Acta 21, 110–126. https://doi.org/10.1016/S0016-7037(60)80006-3

Peters, S.E., 2005. Geologic constraints on the macroevolutionary history of marine animals. Proceedings of the National Academy of Sciences of the United States of America 102, 12326–12331.

Peters, S.E., Foote, M., 2001. Biodiversity in the Phanerozoic: a reinterpretation. Paleobiology 27, 583–601.

Peters, S.E., Husson, J.M., Czaplewski, J., 2018. Macrostrat: A Platform for Geological Data Integration and Deep-Time Earth Crust Research. Geochemistry, Geophysics, Geosystems 19, 1393–1409. https://doi.org/10.1029/2018GC007467

Raja, N.B., Dunne, E.M., Matiwane, A., Khan, T.M., Nätscher, P.S., Ghilardi, A.M., Chattopadhyay, D., 2022. Colonial history and global economics distort our understanding of deep-time biodiversity. Nat Ecol Evol 6, 145–154. https://doi.org/10.1038/s41559-021-01608-8

Roest-Ellis, S., Richardson, J.A., Phillips, B.L., Mehra, A., Webb, S.M., Cohen, P.A., Strauss, J.V., Tosca, N.J., 2023. Tonian Carbonates Record Phosphate-Rich Shallow Seas.

Geochemistry, Geophysics, Geosystems 24, e2023GC010974. https://doi.org/10.1029/2023GC010974

Sessa, J.A., Fraass, A.J., LeVay, L.J., Jamson, K.M., Peters, S.E., 2023. The Extending Ocean Drilling Pursuits (eODP) Project: Synthesizing Scientific Ocean Drilling Data. Geochemistry, Geophysics, Geosystems 24, e2022GC010655. https://doi.org/10.1029/2022GC010655

Smith, A.B., Benson, R.B.J., 2013. Marine diversity in the geological record and its relationship to surviving bedrock area, lithofacies diversity, and original marine shelf area. Geology 41, 171–174. https://doi.org/10.1130/G33773.1

Stockey, R.G., Cole, D.B., Farrell, U.C., Agić, H., Boag, T.H., Brocks, J.J., Canfield, D.E., Cheng, M., Crockford, P.W., Cui, H., Dahl, T.W., Del Mouro, L., Dewing, K., Dornbos, S.Q., Emmings, J.F., Gaines, R.R., Gibson, T.M., Gill, B.C., Gilleaudeau, G.J., Goldberg, K., Guilbaud, R., Halverson, G., Hammarlund, E.U., Hantsoo, K., Henderson, M.A., Henderson, C.M., Hodgskiss, M.S.W., Jarrett, A.J.M., Johnston, D.T., Kabanov, P., Kimmig, J., Knoll, A.H., Kunzmann, M., LeRoy, M.A., Li, C., Loydell, D.K., Macdonald, F.A., Magnall, J.M., Mills, N.T., Och, L.M., O'Connell, B., Pagès, A., Peters, S.E., Porter, S.M., Poulton, S.W., Ritzer, S.R., Rooney, A.D., Schoepfer, S., Smith, E.F., Strauss, J.V., Uhlein, G.J., White, T., Wood, R.A., Woltz, C.R., Yurchenko, I., Planavsky, N.J., Sperling, E.A., 2024. Sustained increases in atmospheric oxygen and marine productivity in the Neoproterozoic and Palaeozoic eras. Nat. Geosci. 17, 667–674. https://doi.org/10.1038/s41561-024-01479-1

Tang, Q., Zheng, W., Zhang, S., Fan, J., Riedman, L.A., Hou, X., Muscente, A.D., Bykova, N., Sadler, P.M., Wang, X., Zhang, F., Yuan, X., Zhou, C., Wan, B., Pang, K., Ouyang, Q., McKenzie, N.R., Zhao, G., Shen, S., Xiao, S., 2024. Quantifying the global biodiversity of Proterozoic eukaryotes. Science 386, eadm9137. https://doi.org/10.1126/science.adm9137

Venugopal, A., Tripathy, G.R., Goswami, V., Ghosh, S.K., Singh, D., 2023. Oceanic Redox State During the Early Cambrian: Insights From Mo-S Isotopes and Geochemistry of Himalayan Shales. Geochemistry, Geophysics, Geosystems 24, e2023GC011182. https://doi.org/10.1029/2023GC011182

Venugopal, A., Tripathy, G.R., Goswami, V., Khan, T., Ackerman, L., 2025. Unravelling the extent of ocean euxinia during the late Paleoproterozoic: Constraints from Re–Os and Mo isotopes. Geochimica et Cosmochimica Acta 407, 158–173. https://doi.org/10.1016/j.gca.2025.09.013

Walton, C.R., Shorttle, O., 2024. Phanerozoic biological reworking of the continental carbonate rock reservoir. Earth and Planetary Science Letters 632, 118640. https://doi.org/10.1016/j.epsl.2024.118640

Wang, X., Algeo, T.J., Li, C., Zhu, M., 2023. Spatial pattern of marine oxygenation set by tectonic and ecological drivers over the Phanerozoic. Nat. Geosci. 16, 1020–1026. https://doi.org/10.1038/s41561-023-01296-y

Wei, G.-Y., Zhao, M., Sperling, E.A., Gaines, R.R., Kalderon-Asael, B., Shen, J., Li, C., Zhang, F., Li, G., Zhou, C., Cai, C., Chen, D., Xiao, K.-Q., Jiang, L., Ling, H.-F., Planavsky, N.J., Tarhan, L.G., 2024. Lithium isotopic constraints on the evolution of continental clay mineral factory and marine oxygenation in the earliest Paleozoic Era. Science Advances 10, eadk2152. https://doi.org/10.1126/sciadv.adk2152

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., ’t Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

Xu, G., Hannah, J.L., Bingen, B., Georgiev, S., Stein, H.J., 2012. Digestion methods for trace element measurements in shales: Paleoredox proxies examined. Chemical Geology, Special Issue Recent Advances in Trace Metal Applications to Paleoceanographic Studies 324–325, 132–147. https://doi.org/10.1016/j.chemgeo.2012.01.029

Ye, Y., Wang, X., Wang, H., Fan, H., Chen, Z., Guo, Q., Wang, Z., Wu, C., Canfield, D.E., Zhang, S., 2024. Hydrological dynamics and manganese mineralization in the wake of the Sturtian glaciation. Geochimica et Cosmochimica Acta 376, 14–24. https://doi.org/10.1016/j.gca.2024.05.021

Zhang, K., Tarbuck, G., Shields, G.A., 2024. Refining the carbonate-associated iodine redox proxy with leaching experiments. Chemical Geology 646, 121896. https://doi.org/10.1016/j.chemgeo.2023.121896

Zhang, Q., Bendif, E.M., Zhou, Y., Nevado, B., Shafiee, R., Rickaby, R.E.M., 2022. Declining metal availability in the Mesozoic seawater reflected in phytoplankton succession. Nat. Geosci. 15, 932–941. https://doi.org/10.1038/s41561-022-01053-7

Zhao, M., Mills, B.J.W., Poulton, S.W., Wan, B., Xiao, K.-Q., Guo, L., Guo, Z., 2024. Drivers of the global phosphorus cycle over geological time. Nat Rev Earth Environ 5, 873–889. https://doi.org/10.1038/s43017-024-00603-4

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Erik Sperling reports financial support was provided by National Science Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Highlights

- Expanded sedimentary geochemical dataset of >120k samples, including new Proterozoic data
- Open source and community-driven, data provided by collaborators
- Modifications to incorporate carbonate and metal isotope data
- Integration with Macrostrat and new bulk-ingested data from external sources
- Publicly accessible via API and updated search website