An interactive, literature-integrated digital crust in support of critical minerals assessment

DARPA CriticalMAAS AI Exploration program, Technical Area 4

Daven P. Quinn[1], Shanan E. Peters[1], Shivaram Venkataraman[2], and Brian Bockelman[3]

[1]Department of Geoscience, University of Wisconsin-Madison
[2]Department of Computer Science, University of Wisconsin-Madison
[3]Morgridge Institute for Research, Madison, WI

Contents

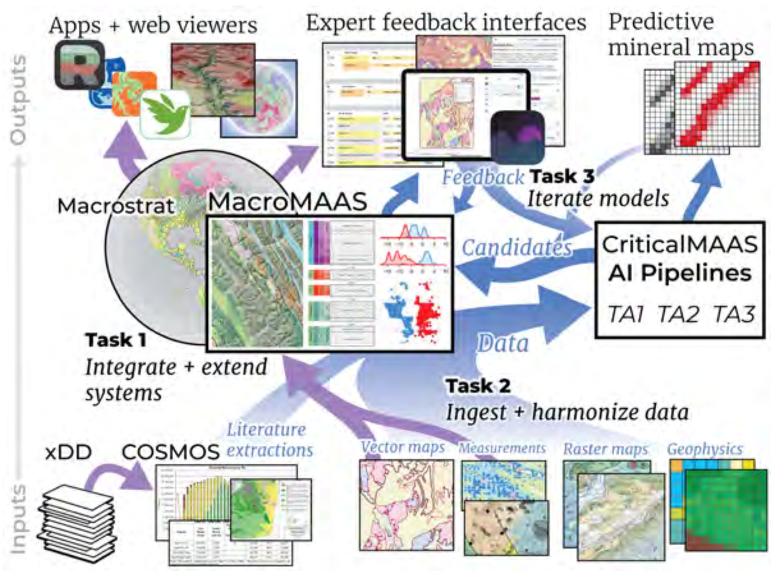

Figure 1: Flow chart showing the organization of project components

Executive summary

Impact

- *Need:* a system to integrate diverse datasets and gather/manage feedback for AI/ML algorithms
- *Problem:* geological data are heterogeneous, complex, and difficult to work with across scales and sources
- *Goal:* Rapid human-in-the-loop workflows that are built on an integrated data platform with automated systems for data discovery and synthesis, harmonized multiscale and multi-modal datasets, and user interfaces optimized for domain experts

Approach

- Macrostrat will store and represent geological information needed by TA1-3 performers using existing data models, extended to handle candidate datasets produced by AI/ML models and feedback on data components
- xDD and COSMOS will provide geologic entity-linked, literature-sourced data to TA1-2
- Existing web-based interfaces will be adapted to support human-in-the-loop feedback, with annotation "widgets" for map feedback and map and stratigraphic editing for rapid correction of AI outputs

Context

We will extend two existing platforms:

- Macrostrat, a comprehensive, widely used database of synthesized geologic information, with: APIs for flexible representation of geological data (e.g., maps, columns, samples) and web-accessible user interfaces designed by geologists
- xDD a machine-reading platform for the scientific literature, with automated mechanisms for document acquisition and processing, access to over 16M articles, and the COSMOS AI knowledge-base construction system for extracting structured multi-modal information

1 Goals and Impact

This TA4 proposal seeks to augment the Macrostrat crustal data system (macrostrat.org) into an extensible platform that supports AI-enabled crustal characterization, which we dub "MacroMAAS." Macrostrat currently serves as a multiscale, contextual basemap that integrates a variety of sources of geological information [Fig. 2]. Here, we seek to harness a wider variety of information to tune and enrich this crustal framework for CMA. Close integration with the xDD–COSMOS document extraction pipeline will engage the geoscience literature in ML-assisted crustal characterization. The proposed MacroMAAS system will initiate this hybrid approach, adding new data ingestion and feedback tools to support AI-enabled CMA. This data platform will both assimilate AI-generated candidates with expert annotations and connect CriticalMAAS performers to available data distribution mechanisms.

If realized, the capabilities and integrations demonstrated by MacroMAAS will represent an advance in our ability to compile widely scattered geological knowledge into comprehensive, vetted description of the Earth's crust. This project will serve as a useful test case for applying AI techniques to extend an established "expert system" that implements geological reasoning. The data-compilation approach and infrastructure developed during this project will be applicable to many geoscience problems with broad societal impact (e.g., hazard assessment, groundwater modeling, carbon sequestration). Additionally, successfully integrating AI-generated datasets into Macrostrat, especially vectorized geologic maps, will allow rapid expansion of a widely used public geologic mapping platform.

2 Technical Plan

The MacroMAAS platform will support CMA in three ways: (1) ingesting geological map and rock unit-linked data from TA1-3 and other sources and overlaying them with "human-in-the-loop" (HITL) interfaces designed to harness user feedback to continuously improve TA extraction algorithms, (2) exposing the full complement of geological data and annotations in this system via APIs and custom visualization software designed for expert users, and (3) linking geological entities to scientific publications and integrating the multi-modal knowledge artifacts they contain into CriticalMAAS workflows.

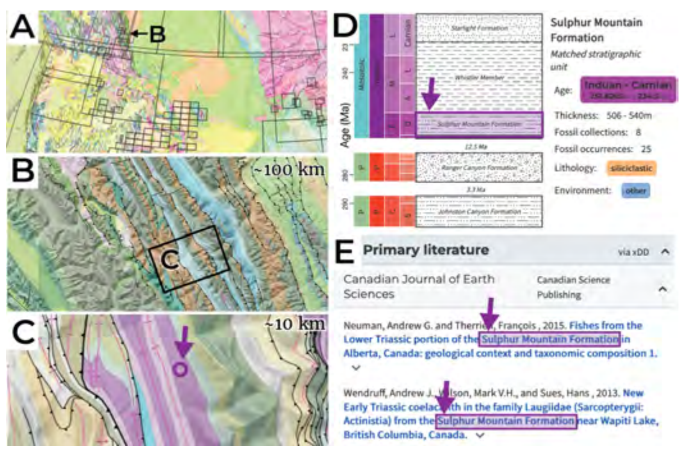

Figure 2: User interface elements for Macrostrat's map interface [1] showing data system attributes. **(a-c)** Multiscale map, with scale traversal by natural zooming. **(d)** Regional stratigraphic column and canonical unit properties. **(e)** Relevant linked literature entities from xDD. Arrows: a single geologic unit linked across all datasets.

During Phase 1 we will develop and test systems capable of handling multi-scale geological data, pipelines to feed raw data to and ingest data products from TA1-2, and APIs to expose the harmonized results to support mineral potential mapping and visualization within expert feedback interfaces. In Phase 2, we will scale and deploy our system and extend our integrated annotation capabilities to interpretative data products produced by TA3, with HITL workflows to enable iterative feedback at every stage of the CMA pipeline.

Project Components

This project will build on two existing data systems, aspects of which will provide immediate testbeds for our ability to facilitate activities in TA1-2 and serve as a geological base map for TA3.

The Macrostrat geological data platform Macrostrat is an NSF-initiated platform for curating geological data about the Earth's crust. It consists of a database of geologic maps and stratigraphic columns, compiled from many sources [2]. The software infrastructure atop this database provides capabilities for data fusion, integrating these distinct types of geological information across scales to characterize the aggregate properties of the crust (e.g., its mass, age, and composition). As such, the database is central to a wide range of research problems [3–5]; to date, Macrostrat has been used in more than 60 publications, including 8 in *Science* and *Nature* [6–13].

Macrostrat's data system is based on industry-standard, open-source software components. Geologic map and column data reside primarily within PostGIS-enabled PostgreSQL. The Macrostrat APIs, written in NodeJS, provide stratigraphic and geologic map information for analytical, reference, and public outreach purposes, with a CC-BY license [14]. Macrostrat's API supports content-rich, easy-to-use web interfaces and mobile apps with thousands of users weekly [e.g., 15]. Macrostrat has become a go-to resource for digital geologic map and stratigraphic data, primarily because it has amassed voluminous data that cannot be used in aggregate anywhere else. Over 300 different geologic maps, containing more than 2.5M rock unit polygons covering the world, are integrated into a unified, multi-scale, queryable environment [1]. More than 35,000 rock units from over 1,500 regional columns covering all of North America and several other regions globally, and ~2,000 columns in "projects" associated with separate, in-progress data compilation efforts, are also available. Macrostrat also maintains an archive of geochemical, geochronologic, and other sample-based data, with over 3M distinct data points available now. In some respects, Macrostrat is similar other systems that aggregate geological data (e.g., ScienceBase, EarthMRI, and NGMDB). However, Macrostrat seeks to build a unified, quantitative representation of the Earth's crust by bringing many geologic data sources into a common schema and linking shared geologic entities, allowing aggregate analysis and visualization [Fig. 2].

The xDD—COSMOS machine-reading platform xDD is a digital library and cyberinfrastructure that supports the location and extraction of information from PDF publications. xDD maintains publisher agreements that allow the system to access the full text of scientific articles. The system currently has access to over 16M PDFs, and it grows by thousands daily, making it one of the largest full-text archives of scientific literature. xDD automatically retrieves documents from publishers and applies basic indexing and machine-reading tools to the entire corpus; it also allows more complex ML/AI algorithms to be applied on-demand. The system has roots in geoscience, having originally been named "GeoDeepDive" and as such includes many geoscientific data sources. xDD is situated within the Center for High Throughput Computing (CHTC) at UW-Madison, which allows it to leverage substantial distributed computing capacity via HTCondor. xDD's public API allows full-text data mining using web requests, allowing other systems to readily establish links between scientific concepts/entities and the literature, leading to new science [e.g., 16, 17]. Recently demonstrated techniques to extract mineral-potential information from unstructured text [18] can be applied at scale when used over the xDD corpus.

One of the most powerful applications that has been built on top of xDD is COSMOS[1], a deep-learning system for extracting structured multi-modal information from PDFs [19]. COSMOS seg-

[1] COSMOS originated in the DARPA AIE "Automating Scientific Knowledge Extraction".

ments documents by content type (e.g., figures, tables, text) and builds an indexed knowledge-base over sets of documents. This system excels at extracting content of specific types from the literature, for instance allowing a search for "map" and returning the relevant content blocks as an image linked to bibliographic metadata and in-document semantic context. The AI-driven capabilities of COSMOS will allow deeper integrations with Macrostrat and other data platforms to surface relevant data from the scientific literature and technical reports and prepare them for inclusion in geological workflows.

T1 Supply geological data and literature artifacts to CriticalMAAS TAs 1-3

T1a Extend Macrostrat for TAs 1-3 The infrastructure supporting Macrostrat is well-oriented to ingesting and distributing geologic information from a variety of sources, and it can immediately begin providing data services in support of TA 1-3 workflows. A new scalable instance of Macrostrat, "MacroMAAS" (Macrostrat Mineral Assessment with AI Support) will be created in the CHTC infrastructure at UW–Madison. This system will be based on the current instance of Macrostrat, and will include all of its data. The data system will be containerized, converted to a service-oriented architecture, and established as a "code-as-infrastructure" GitHub repository. Individual services (e.g., the API, tile server, and backup subsystems) will be included as submodules or by reference to versioned, published Docker images. In Phase 2, this system will be deployed on a scalable cloud infrastructure [T4].

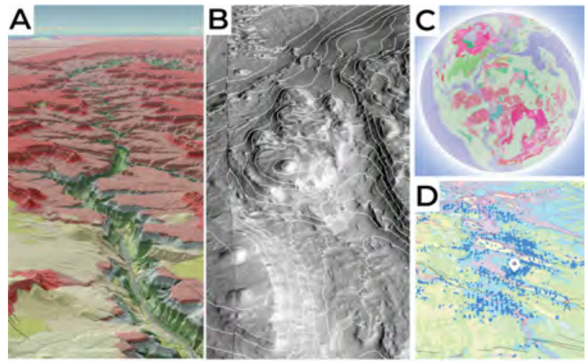

Figure 3: Visualization capabilities enabled by Macrostrat's tile server and web frameworks. **(a)** Vector tiles atop a high-resolution 3D basemap; *Grand Canyon*. **(b)** On-the-fly compositing of raster images and elevation maps; *Aeolis Dorsa, Mars*. **(c)** Paleogeographic rotation of Macrostrat vector tiles to 250 Ma. **(d)** Measurements layer showing USGS Mineral Resources Data System [20] sites.

During Phase 1, MacroMAAS will be primarily geared towards providing geologic map and column data to TA3. It will also be capable of supplying a variety of data to TA1-2 that can be useful for distant supervision and other model-training tasks. TAs will gain immediate API access to standardized data from over 300 geologic maps, which have been ingested, normalized into a common representation, enhanced to a minimum standard of age and lithologic information, and semantically linked to Macrostrat's stratigraphic archive. Point datasets, such as geochemical measurements, are also linked into this framework and are extensible as needed.

Raster data infrastructure Macrostrat maintains several public-facing, highly performant map data services as part of its core infrastructure. Its tileserver provides a global, multiscale geologic dataset to a large user base (~100,000 users/day). However, current systems are optimized for static vector map layers. Macrostrat has develop a prototype tile server infrastructure [21] that also supports raster data[2]. This system is derived from a new generation of "dynamic tilers" for serving geospatial data in cloud environments [e.g. 23]. It retains the performance and efficient caching of Macrostrat's existing tile server while allowing dynamic, on-demand tile generation from arbitrary datasets in multiple map projections [Fig. 3]. The system is optimized for distributed computing,

[2]Raster capabilities are designed to support both planetary-scale basemaps as part of PI Quinn's contributions to orbital Mars science [22] and future structural modeling efforts for Macrostrat.

allowing dynamic compositing of raster layers and referencing of remotely-stored (i.e., in object storage buckets) cloud-optimized GeoTIFFs [24]. This flexible system will allow pipeline inputs and outputs to be explored and queried without substantial data transfer between CriticalMAAS performers. By *Milestone 2*, this new tile server will be integrated into MacroMAAS, with a layer ingestion API to allow cloud-stored datasets to be registered for standalone or composited display. Steps to access tile layers in web interfaces and GIS software (ArcGIS; QGIS) will be documented.

Model feedback capabilities Several evolutions to Macrostrat's database schema and APIs will be needed to enable feedback on AI-generated datasets. Support will be added for "candidate" datasets for all TA target data, by extending existing models (maps, columns, and measurements) with fields to distinguish between model-generated and canonical datasets, and to record readiness levels (e.g. *candidate*, *corrected*, and *accepted*) for AI-generated data. A feedback subsystem will center around a single table maintaining polymorphic links to data models. This table will store quality flags, comments, specific locations or areas, and, in some cases, provenance information for corrected datasets. The resulting system will accommodate CriticalMAAS outputs alongside geologic information (maps, columns, and geochemistry) from other sources, and enable feedback atop all of these datasets.

New APIs will support ingestion and feedback for TA model results. Current Macrostrat ingestion pipelines are powerful but require manual data-cleaning steps; these will be streamlined for unsupervised use [T2a]. Documented API endpoints will allow ingestion of candidate outputs from TA pipelines [T2], and submission/retrieval of model feedback [T3] by TA1-3 pipelines. These proposed APIs will be the first Macrostrat data services to allow direct data entry by external users. Accordingly, they will be wrapped with a token- (for automated pipelines) and password-based (for expert users) authentication scheme based on xDD's current system.

T1b Extract literature artifacts using xDD-COSMOS and deliver to TA1-2 The xDD and COSMOS machine-reading and knowledge extraction systems provide another capability that can be harnessed to enable mineral potential mapping. xDD has access to a large subset of the geoscientific literature, and its ML-powered COSMOS overlay can provide extracted figures, tables, and other content types for document collections. Extractions will be generated for specific keywords relevant to CMA. The COSMOS pipeline will then extract and enable searchability of figures and tables within these sets. Relevant information (e.g., local-scale geologic maps or tables of element assays) will be identified and provided as data to the TA1 and TA2 modeling pipelines, via APIs as needed. Embedding models trained over geological corpora will also be made available via APIs and as stand-alone data products. This application of AI capabilities in the xDD infrastructure will augment Macrostrat's core data records with specific information that can be of immediate value to TA3 mineral mapping. This pipeline will be implemented by a graduate student (CS) working closely the project team. Because of its internal nature, the project will provide early and continuous opportunities for initial testing of data pipelines [T2] and feedback interfaces [T3], laying the groundwork for integration with more complex TA2 modeling pipelines.

T2 Ingest geologic data from TAs 1-3

Macrostrat capabilities to ingest structured geologic data, adapted to handle AI-generated candidates [T1], will be used to stage candidate datasets from TAs 1-3 and harmonize them with Macrostrat's other holdings. This process establishes the basis for expert feedback and correction workflows [T3] that will enable model iteration. In Phase 1, AI-generated candidate datasets produced by TA1-2 will be integrated into Macrostrat data holdings, linked to known geological entities (e.g., rock units from maps and/or columns), and resurfaced via APIs for use in TA3. TA3 interpretive datasets will be ingested in Phase 2 [T5].

T2a Ingest geologic maps from TA1 and link entities Geologic map candidates produced by TA1 will consist of vectorized geospatial data. These data can be direct inputs to TA2-3, but they are also amenable to ingestion into Macrostrat's geologic map system, which initiates a cascade of entity linking and standardization that can enrich maps and integrate them into multiscale products. Macrostrat's current map ingestion pipelines, will be adapted to MacroMAAS by automating several manual data-cleaning steps; possibly looser initial capture can be tolerated for AI-generated candidates, especially since quality flags and feedback tools will allow adjustment after import. Staging maps in MacroMAAS will make them accessible via Macrostrat APIs and tile servers, preparing them for deployment in visualization, annotation, and map editing interfaces to support expert feedback [T3a].

T2b Ingest geologic data from TA2 and link entities Documents processed by TA2 will yield a variety of data describing location-specific rock unit properties that are useful for critical mineral assessment (e.g., ore grade and tonnage). These data can be directly input to TA3, but in many situations Macrostrat will be able to provide additional linked information (e.g., spatial and temporal distribution of a given rock body, associated structures, etc.), potentially increasing context for the extractions. We will specify schemas for ingesting point-based geological data into Macrostrat and then surface the data in APIs accessible to all TAs. This will also stage TA2 data for visualization in feedback interfaces that will enable HITL iteration [T3b].

T3 Build human-in-the-loop interfaces for model and extraction improvement

Systems that enable efficient human-computer interaction are critical to digital data synthesis and can greatly enhance the efficacy of AI-driven workflows [e.g. 26]. From its outset, the Macrostrat team has designed the data system to support web interfaces that enable rich geological visualization and reasoning. Accordingly, the Macrostrat team maintains substantial capabilities for user interface development, having produced a variety of modern, optimized, and broadly accessible web and mobile applications [e.g., 1, 15]. This expertise will be drawn on to produce geologically rich model feedback workflows. Broadly, feedback interfaces will comprise "widgets" in map and stratigraphic context [Fig. 4b]; more substantial corrections will be made using existing map and column editing applications [Fig. 4c-d]. Editing interfaces will be produced by PI Quinn and CHTC staff in close collaboration, with internal testing and documentation support by a graduate student.

Figure 4: User interfaces for model feedback. **(a)** Existing training and feedback widgets for xDD–COSMOS entity extraction pipeline. **(b)** Prototype widgets for feedback on map-unit and deposit candidates, based in Macrostrat components and user interfaces [1, 25]. *(c-d)* Existing software which will be employed to correct AI extractions. **(c)** Mapboard GIS map editing app, showing capabilities for topology creation and editing. **(d)** Macrostrat stratigraphic column editor.

T3a Annotate and edit geologic maps The complex spatial and temporal relationships encoded by geologic maps present particular challenges for their digital authoring; these same challenges complicate feedback for AI-generated maps. Macrostrat has sought to streamline the time-consuming process of creating geologic maps with editing tools that allow rapid digital manipulation of map boundaries and stratigraphic relationships. These tools will be integrated into MacroMAAS to apply corrections to TA1 vector geologic maps. Edits will support both feedback and retraining for TA1 models, with finishing edits to bring AI-generated datasets up to a uniform quality standard. Widget-based annotation tools will also be provided for candidate maps, focusing on validating unit identifications and confirming accurate capture of spatial boundaries [T3b].

Stratigraphic editing Macrostrat's column acquisition/editing user interface supports the creation and manipulation of a temporal layout of geologic units. A modernized React-based web app [Fig. 3d] awaits testing and integration with user-management capabilities [T1a]. An additional capability will be added to display column information styled as a "correlation of map units," with colors and labels synced to geospatial representations. This output will allow geologists to apply typical map quality criteria, including the temporal layout of units, to AI-generated maps.

Geospatial map editing The Mapboard GIS system [27], created by PI Quinn, complements Macrostrat with tools for efficient creation of geologic maps. Geological contacts are assembled into a space-filling map using iterative topological algorithms implemented in PostGIS [28]; an iPad app [29] supports fluid map editing that mimics pen-and-paper workflows. Mapboard GIS server components will be integrated into MacroMAAS by *Milestone 4* to support editing of AI-generated candidate vector maps. This system will allow geologists to rapidly and intuitively correct map spatial information. Compatibility with GIS workflows (ArcGIS; QGIS) will be maintained alongside optimized, iPad-based map editing.

Cooperative map vectorization Rich editing interfaces for map-associated data form the centerpiece of the MacroMAAS workflow. These new editing interfaces will be documented and made available to external users by *Milestone 4*. In the second half of Phase 1, these interfaces will be aligned with ingestion APIs to tighten model-feedback iteration, leading to a reactive "mixed-initiative" system [26] for AI-supported map digitization.

T3b Annotate geologic data extractions and linked entities Feedback on AI-synthesized mineral-assessment information (TA2) will be chiefly provided using "widget"-based tools [Fig. 4b], which will allow annotation of model quality, and, where appropriate, basic correction of links to Macrostrat vocabularies (e.g., geologic units, lithologies, minerals). Views adapted from Macrostrat's web interface [1], based on React and the Mapbox mapping environment, will situate these tools in map and stratigraphic context. The web codebase is partially transitioned to a system of reusable React components (e.g., column renderers, map pages) that will simplify addition of new pages and tools [25]. Rich geological syntheses and model inputs will be presented alongside candidate extractions to increase the quality and efficiency of feedback. Interfaces will be constructed at *Milestone 2*, with results integrated to support TA2 pipelines at *Milestone 4*.

Phase 2 (option) tasks

In Phase 2, an end-to-end system will be deployed to support feedback workflows for TA1-3. This will allow MacroMAAS to support mineral assessment workflows from the ingestion of raw maps and literature to the final generation of mineral potential maps.

T4 Deploy MacroMAAS in Kubernetes Unlike the current deployment of Macrostrat, a production MacroMAAS system will support direct access by AI pipelines and feedback providers while managing potentially large numbers of AI-generated data revisions. Macrostrat services containerized in Phase 1 [T1a] will be deployed in a Kubernetes cluster on CHTC infrastructure, allowing modern capabilities for scaling, replication, and load balancing. Maintenance load will be shared with Macrostrat and xDD core services, and colocation in CHTC will tighten integration with COSMOS and other high-throughput AI-based processing workflows. CHTC staff will work closely with the project team to optimize and deploy the MacroMAAS stack in this infrastructure.

T5 Ingest interpretive data products from TA3 TA3 ML/AI mineral potential mapping pipelines will synthesize data generated by TA1-2, evidence layers served by MacroMAAS, and other GIS data products into mineral potential maps, which are likely to be geospatial rasters defining mineral occurrence probabilities and uncertainties. Although most needed functionality will be established in Phase 1, we target Phase 2 for fully ingesting these AI-generated data products into MacroMAAS for visualization within an ecosystem of applications and model-feedback interfaces [T6].

T6 Annotate mineral potential model predictions TA3 mineral potential model predictions and uncertainties ingested into Macrostrat's raster infrastructure [T5] will be made available in widget-based feedback interfaces for during Phase 2. Feedback tools will allow geologists to identify and correct false positives and tune inputs by adjusting model-appropriate parameters. Requirements will be gathered from TA3 performers by *Milestone 6*, and feedback interfaces will be constructed by *Milestone 8*, supporting active iteration on TA3 models by the end of the project timeline.

Deliverables

This system will be delivered as a set of containerized software and data components: (1) The MacroMAAS system, which will augment Macrostrat capabilities for organizing crustal data along with data extracted from the geological literature via xDD; (2) pipelines to ingest and coordinate AI assessment and feedback atop MacroMAAS; and (3) human-in-the-loop feedback tools for rich annotation of AI-generated datasets, drawing on Macrostrat linking capabilities and designed by geologists. All software infrastructure developed during this project will be open source, and the entire system will be portable to USGS and other organizations. Geologic map datasets produced during this project will be made available to the general public via Macrostrat's mapping API, producing an immediate broader impact.

3 Capabilities and management plan

The PIs are geo- and computer scientists who are experts in digital stratigraphy, geologic mapping, large-scale machine learning, and cyberinfrastructure. PI Quinn is an early-career geoscientist who has worked broadly in crustal structure, remote sensing and geologic mapping, and is now focused on building end-to-end software to enhance geological science [30]. Quinn leads efforts to broaden the Macrostrat data platform [31], and conceptualized and built Macrostrat's column visualizations, 3D/paleogeographic map interfaces [1], tileserver, the Mapboard GIS mapping system [29], and other software/visualization building blocks [32–36]. Research Scientist Quinn will oversee and coordinate project activities. Co-PI Peters is a geoscientist and the leader of the Macrostrat project, having inaugurated its approach and used it to produce high-impact science. Professor Peters is also a leader in summarizing data from the geologic literature, and was instrumental in developing the xDD and COSMOS systems; Peters will assist with project management and oversee Macrostrat-related work. Co-PI Venkataraman is an expert on developing software systems for large-scale machine learning and data analysis. Assistant Professor Venkataraman is a founding Project Management Committee member for Apache Spark and more recently developed Marius, an engine for training large scale graph embeddings; he will supervise graduate students and oversee AI method development for locating and extracting geologically-linked information from publications. Co-PI Bockelman has extensive experience in operating scientific services using "cloud native" technologies. Bockelman will oversee xDD and MacroMAAS infrastructure.

The PIs will be supported by already in-place CHTC staff who will provide software development, data management, and infrastructure support for the project. Graduate students in computer science and geoscience will help construct literature data pipelines and validate feedback workflows. Project participants will meet weekly with PI Quinn and collaborate regularly on software development tasks. PI Quinn will additionally host weekly meetings open to participants in all TAs to coordinate their use of the MacroMAAS platform.

References

[1] *Macrostrat map interface*, `https://macrostrat.org/map`.

[2] S. E. Peters, J. M. Husson, and J. Czaplewski, Macrostrat: A platform for geological data integration and deep-time earth crust research, *Geochemistry, Geophysics, Geosystems* 19.4 (2018), 1393–1409.

[3] S. E. Peters, Macrostratigraphy of North America, *The Journal of Geology* 114.4 (2006), 391–412.

[4] S. E. Peters et al., Macrostratigraphy: Insights into Cyclic and Secular Evolution of the Earth-Life System, *Annual Review of Earth and Planetary Sciences* 50.1 (May 2022), 419–449, issn: 0084-6597, 1545-4495, doi: `10.1146/annurev-earth-032320-081427`, (visited on 07/08/2022).

[5] D. C. Segessenman and S. E. Peters, Macrostratigraphy of the Ediacaran System in North America, *Laurentia: Turning Points in the Evolution of a Continent*, Jan. 2023, isbn: 9780813712208, doi: `10.1130/2022.1220(21)`, url: `https://doi.org/10.1130/2022.1220(21)`.

[6] J. Wang et al., The evolution of the marine carbonate factory, *Nature* (2023), 1–5.

[7] R. Close et al., The spatial structure of Phanerozoic marine animal diversity, *Science* 368.6489 (2020), 420–424.

[8] T. T. Isson and N. J. Planavsky, Reverse weathering as a long-term stabilizer of marine pH and planetary climate, *Nature* 560.7719 (2018), 471–475.

[9] S. E. Peters, D. C. Kelly, and A. J. Fraass, Oceanographic controls on the diversity and extinction of planktonic foraminifera, *Nature* 493.7432 (2013), 398–401.

[10] I. Halevy, S. E. Peters, and W. W. Fischer, Sulfate burial constraints on the Phanerozoic sulfur cycle, *Science* 337.6092 (2012), 331–334.

[11] S. E. Peters and R. R. Gaines, Formation of the "Great Unconformity" as a trigger for the Cambrian explosion, *Nature* 484.7394 (2012), 363–366.

[12] B. Hannisdal and S. E. Peters, Phanerozoic Earth system evolution and marine biodiversity, *Science* 334.6059 (2011), 1121–1124.

[13] S. E. Peters, Environmental determinants of extinction selectivity in the fossil record, *Nature* 454.7204 (2008), 626.

[14] *Macrostrat API*, `https://macrostrat.org/api`.

[15] *Rockd: Learn, explore and document your geologic world*, `https://rockd.org`.

[16] J. F. Emmings et al., Pyrite mega-analysis reveals modes of anoxia through geological time, *Science Advances* 8.11 (2022), eabj5687.

[17] S. E. Peters, J. M. Husson, and J. Wilcots, The rise and fall of stromatolites in shallow marine environments, *Geology* 45.6 (2017), 487–490.

[18] C. J. M. Lawley et al., Applications of Natural Language Processing to Geoscience Text Data and Prospectivity Modeling, *Natural Resources Research* (June 2023), issn: 1520-7439, 1573-8981, doi: `10.1007/s11053-023-10216-1`, (visited on 06/03/2023).

[19] *COSMOS: Knowledge base construction from raw scientific documents*, https://github.com/UW-COSMOS/Cosmos.

[20] US Geological Survey, *Mineral Resources Online Spatial Data*, https://mrdata.usgs.gov/, Online; accessed 3-June-2023, 2023.

[21] *Macrostrat's tileserver: A dynamic tile server for geologic maps*, https://github.com/UW-Macrostrat/tileserver.

[22] D. P. Quinn and B. L. Ehlmann, The Deposition and Alteration History of the Northeast Syrtis Major Layered Sulfates, *Journal of Geophysical Research: Planets* 124.7 (2019), 1743–1782, issn: 2169-9100, doi: 10.1029/2018JE005706, (visited on 10/05/2019).

[23] *TiTiler: A modern dynamic tile server built on top of FastAPI and Rasterio/GDAL*, https://github.com/developmentseed/titiler.

[24] *Cloud Optimized GeoTIFF*, https://www.cogeo.org/.

[25] *Macrostrat web components: A library of React-based UI components for composing data-rich geological applications*, https://github.com/UW-Macrostrat/web-components.

[26] E. Horvitz, Principles of Mixed-Initiative User Interfaces, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99*, Pittsburgh, Pennsylvania, United States, 1999, 159–166, isbn: 978-0-201-48559-2, doi: 10.1145/302979.303030, (visited on 06/05/2023).

[27] D. P. Quinn, The Mapboard, Reimagined Digitally — Building a User-Interface Tailored for Field Geological Mapping, *GSA 2020 Connects Online*, Oct. 2020, (visited on 03/09/2023).

[28] *Mapboard topology manager: A PostGIS-based iterative system for building geologic maps*, https://github.com/mapboard/topology-manager.

[29] *Mapboard GIS*, https://mapboard-gis.app.

[30] *Daven P. Quinn's CV*, https://davenquinn.com/cv.

[31] D. P. Quinn, C. I. Idzikowski, and S. E. Peters, Building a multi-scale, collaborative, and time-integrated digital crust: the next stage of the Macrostrat data system, *Geoscience Data Journal* in press (2023).

[32] *COSMOS Visualizer: User interfaces for COSMOS training-data generation and results visualization*, https://github.com/UW-COSMOS/cosmos-visualizer.

[33] D. P. Quinn et al., Sparrow: A Data Management System and Cyberinfrastructure Component Targeted at Geochronology Laboratories, *GSA Annual Meeting in Phoenix, Arizona, USA-2019*, 2019.

[34] D. P. Quinn, *FGDC Geologic Patterns for the Web*, https://davenquinn.com/projects/geologic-patterns, 2017.

[35] *Cesium-Martini: On-the-fly meshing of raster elevation tiles for the CesiumJS virtual globe*, https://github.com/davenquinn/cesium-martini.

[36] *Jezero Crater's context within northeast Syrtis Major — a multiscale interactive explorer*, https://dev.macrostrat.org/mars/syrtis-jezero/.