MinMod: Extracting Models of Minerals from Knowledge
CriticalMAAS – TA2

Craig Knoblock (PI)
USC Information Sciences Institute

Table of Contents

1. Proposal Summary

USC Information Sciences Institute, in collaboration with InferLink Inc. and University of Minnesota, is proposing to develop an end-to-end system for creating grade and tonnage models and rich mineral site data from all relevant academic, governmental, and mineral industry datasets and reports. The resulting system, which we call MinMod, will automate much of the work in generating these models by building on the latest AI and machine learning techniques. MinMod will make it possible to create and maintain a comprehensive library of grade and tonnage models for minerals from the latest available information and produce mineral site data with a rich set of features.

The task of creating mineral models today is a manual, labor-intensive task requiring months of effort to assemble the relevant data and documents, extract the key pieces of information, and process the results into the final model. This information is then typically published as a report [Green et al., 2020], which becomes out of date as new mines go online or the existing mines publish their next technical or annual reports.

The core technical challenges and our plans to address them are as follows:

- **Linking and reconciling geospatial data sets:** we will build on our past work on geospatial entity linking using spatial language models to reconcile data on mineral sites from multiple sources.
- **Accurately extracting data from text in company reports:** we will deploy and extend our past work on using large language models for text extraction from documents to extract the production, reserve, and grade information from company mining reports.
- **Accurately extracting data from tables and structured data sources:** we will use semantic modeling techniques and build a mining knowledge graph to accurately extract and model the information in data sources and tables.
- **Automatically classifying deposit types:** we will use embedding techniques to represent the various deposit types in a vector space and use those embeddings to classify the data describing mines of unknown deposit type.
- **Building grade and tonnage models that can be vetted and updated:** we will represent and store the data and the source of that information in the mining knowledge graph, providing provenance information for all the data.

Our approach is both novel and unique by bringing together the latest machine learning techniques to solve the difficult challenges and using the latest work on knowledge graphs [Kejriwal, 2021] to provide a scalable and maintainable approach for managing the data. Upon success, MinMod will be able to rapidly create grade and tonnage models with minimal user interaction. These grade and tonnage models will then be used to populate rich mineral site data, which provides one of the important data inputs for prospectivity modeling. Beyond just creating these models, the ability to rapidly incorporate the latest reports from mining companies will make it possible to immediately update the grade and tonnage models and the mining site data with the very latest information.

We have assembled the ideal team to address this task with world experts in spatial data integration, text extraction, table understanding, knowledge graphs, and economic geology, and a PI with extensive experience in leading teams. This project will take twelve months and cost $700K to build an end-to-end system for creating grade and tonnage models and rich mineral site data, and it will take an additional six months and cost $300K to refine the system and apply it to a create the full set of grade and tonnage models for critical minerals.

2. Goals and Impact

Our goal is to semi-automatically build grade and tonnage models and rich mineral site data for a given mineral deposit type and reduce the time to build them from months to hours. The innovations that will make this possible are to build on the latest machine learning techniques, including 1) spatial language models to reconcile geospatial sources of data relevant to the minerals industry, 2) large language models to extract the needed attributes from minerals industry reports, 3) table learning to rapidly ingest and understand the contents of tables in mining reports and other sources, 4) neural embeddings to accurately classify and determine the deposit type of a given mineral deposit, and 5) knowledge graphs to represent, store and manage the data to construct the models, provide the ability to justify every value in the model, and quickly update the models as new data becomes available. In each step of the process, we will support a human-in-the-loop interface to allow a person to verify the results and make corrections as needed. The only previous work on automating creating grade and tonnage models is on generating the grade and tonnage graphs automatically [Singer & Bliss, 1990].

The impact of this work will be to both construct and maintain the grade and tonnage models and rich mineral site data for all the critical minerals and their corresponding deposit types [Hofstra et al., 2021]. We plan to build a system that can process repositories of geospatial datasets and mining reports (SEDAR, EDGAR, etc…) from various countries, categorize the data based on mineral and deposit type, and extract and use the relevant data to populate the grade and tonnage models and mineral site data. The ability to automate much of this processing will make it possible to create new grade and tonnage models rapidly from the source data and documents. This means that existing grade and tonnage models and mineral site data can be quickly updated when new mines come online or new information about existing mines becomes available. Thus, providing up-to-date information for prospectivity modeling.

3. Technical Plan

In this section we describe each of the technical components as shown in Figure 1.

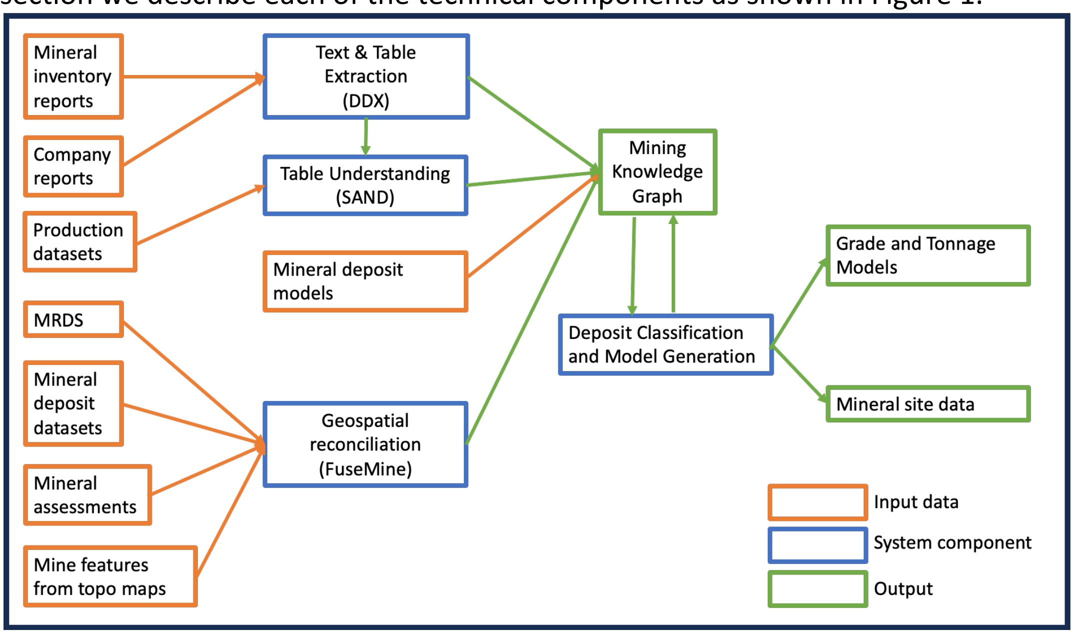


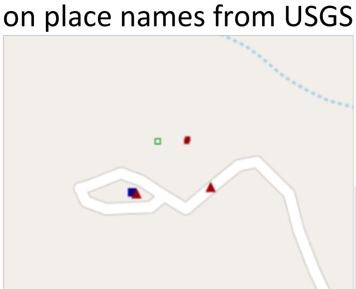
Figure 1: Architecture of the proposed MinMod system.

3.1 Reconciling Available Mineral Site Datasets

Many geospatial datasets describing mineral sites are publicly available, such as the datasets from the Mineral Resource Data System (MRDS), the USGS's mineral deposit database project (USMIN) (Mauk et al., 2021), geological maps (e.g., from TA1's output), and previous mineral assessments. These datasets provide valuable knowledge characterizing existing and potential mineral sites and are often costly to produce. However, the quality and accuracy of the reported location and attributes in them vary significantly and the records may not be up to date due to the continuous nature of mineral exploration and mining. For example, some MRDS records can have anonymous reporters (Schweitzer, 2019), and they typically contain only point locations. Further, significant data inconsistencies exist. For example, Figure 2 shows one working mine and two deposits from USMIN and one prospect and two mines from MRDS. Close examination of these records shows that there exist duplications and complicated M:M matches. Reconciliation of the two datasets is challenging because of the inconsistencies of the naming conventions (e.g., Meeke-Hogan Mine vs. Meeke Mine and SN vs. tin) and possible location determination methods (e.g., point locations of a pit entrance vs. the polygon center of a geographic feature).

We propose to build a machine learning approach called FuseMine, aiming to take all sources of geographic datasets describing mineral sites to generate a fused, integrated view of mineral sites. We will work with USGS's subject matter experts (SMEs) to select one or multiple "gold standard" datasets as the base layer (e.g., USMIN and previous mineral assessments) to link all other datasets to enrich their information.

Previously we built a spatial language model, called SpaBERT, to capture the spatial-varying context of points-of-interest (POI) and support POI linking across multiple datasets (Li et al., 2022). SpaBERT hypothesizes that we will know a POI's characteristics by its surrounding POIs, similar to knowing word meanings by their linguistic context. Specifically, SpaBERT extends BERT (Devlin et al., 2019) to encode a spatial neighborhood to obtain a contextualized representation of a POI. For example, SpaBERT first linearizes the neighboring POIs of the pivot POI, "University of Minnesota" into a sentence (Figure 3) and adds spatial embeddings to capture the order, distance, and orientation of each neighboring POIs to the pivot POI. During pre-training, SpaBERT learns to complete the full names of a POI from randomly masked partial names and spatial relations between subtokens (masked language modeling, MLM) and recovers the masked POI using its spatial relation to neighboring POIs (masked entity prediction, MEP). These tasks drive SpaBERT to learn a POI's surrounding characteristics and archieve significant performance improvement on POI linking over state-of-the-art approaches on place names from USGS historical topographic maps and WikiData.



Site name	Development	Commodities
Meeke-Hogan Mine	Producer	SN
Hogan-Mallery Tin Mine	Past Producer	SN
Tejon Tin Project	Prospect	SN

Site name	Feature type	Commodities
Meeke Mine	Underground Workings	tin
Meeke, West gossan	Deposit	tin
Meeke, East gossan	Deposit	tin

Figure 2. Data sources can have significant inconsistencies between sources and duplications within a source (map: USMIN and MRDS, top-right: MRDS, bottom-right: USMIN).

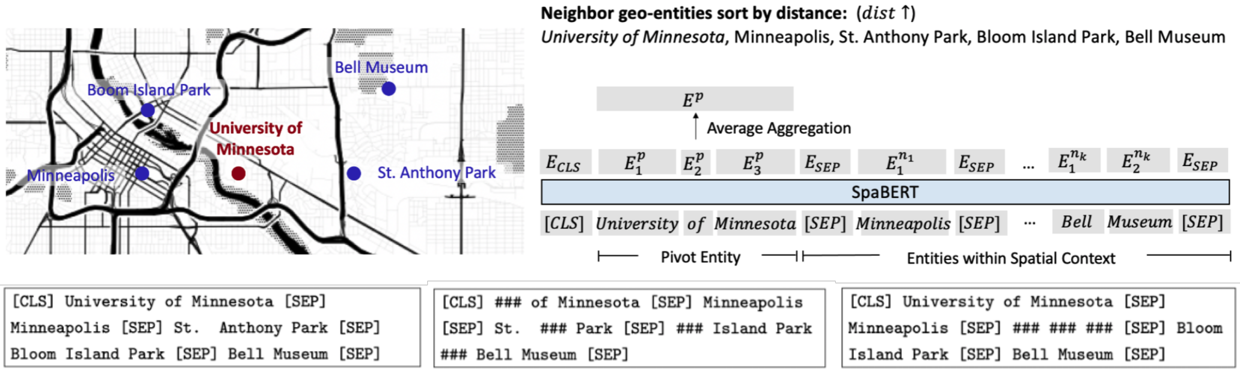


Figure 3. SpaBERT linearizes 2D POIs (top) and uses MLM (middle) and MEP (right) to learn contextualized embeddings for a POI

We will build on SpaBERT's superior linking performance to develop FuseMine. FuseMine will first linearize neighboring mineral sites for individual mineral sites and learn a contextualized site embedding (CSE) for every site. FuseMine will also treat the attributes of each site as a document and encode them (e.g., using BERT) to represent a site in addition to using the site name. Then FuseMine will use MLM and MEP to generate CSEs for every site in every data source to capture the location and attribute of each site and their neighboring sites. With the CSEs, FuseMine will start the M: M matching process by clustering mineral sites from all sources using their CSEs and geographic locations (e.g., with Self-organizing Map). The mineral sites in the same cluster represent sites with similar attributes in close locations and should be linked. FuseMine will also add clustering information to the attributes of all gold standard records. These additional attributes inform downstream prediction tasks (TA3) about the linking and potential duplications and allow for necessary interventions or further investigations (TA4). FuseMine's final result is a gold standard dataset with enrichment from attributes sourced from various other data sources.

3.2 Harvesting Data from Mining Reports

There are a variety of mining reports that contain information that is useful for characterizing mineral deposits and for classifying mineral deposits into deposit types. For example, the CSA's SEDAR system in Canada and the SEC's Edgar system in the US contain company reports, including annual reports and technical reports, that have substantial relevant information. For instance, a mining company report may detail grade and tonnage data for specific deposit types at a given mine.

There are a substantial number of these reports on a site such as SEDAR, but identifying and extracting the appropriate data can be challenging for a variety of reasons. Relevant data may be present in only certain documents, and the documents can be hundreds of pages long, with relatively unstructured formats. The target data can be present in text and/or tables, and the descriptions can be complex.

For instance, Figure 4 shows a snippet of text in a 150-page annual report from Teck, a Canadian company, that indicates the copper production of Highland Valley Copper, an asset owned by Teck. Our goal is to extract structured information from the report, such as the tuple: (Relation=Production, Entity=Highland Valley Copper, Year = 2021, Mineral=Copper, Amount=130,800 tonnes). Note that involves identifying the right sentences in the report, identifying the specific data fields, and aligning them appropriately; for instance, the 2021 production was 130,800 tonnes, as distinct from the 2022 production.

Highland Valley Copper

Highland Valley Copper Operations is located in south-central B.C. Gross profit was $580 million in 2022, compared with $721 million in 2021 and $331 million in 2020. The decrease from 2021 was primarily the result of lower copper prices and production, and higher operating costs driven by inflationary pressures.

Highland Valley Copper's 2022 copper production was 119,100 tonnes, compared to 130,800 tonnes in 2021. The decrease in 2022 production was primarily a result of lower copper grades, coupled with a decrease in mill throughput

Figure 4: Text from a mining company technical report describing production.

To achieve this, InferLink will use the DDEx (Deep Data Extraction) NLP platform that we have developed in previous work for extracting information from technical documents. DDEx was originally developed in a DARPA-funded SBIR project and successfully commercialized, resulting in a spinoff company, Evid Science, that extracted information from the biomedical literature for the life sciences industry. (Evid Science was acquired in 2020.) DDEx is also currently being transitioned for use by DTRA and NGA (under separate non-SBIR phase III contracts) to extract information related to biosecurity and chemical weapons. As a result of these commercial and government successes, DARPA recently published an SBIR success story (see https://www.darpa.mil/attachments/InferLink-Corporation-Success-Report.pdf).

The DDEx system includes a harvester that can collect documents from web sources as well as API-accessible structured source. The documents are processed by an NLP pipeline that includes a series of ML-based extractors and classifiers that can be configured to extract entities, relations, locations, etc. The resulting data is entity resolved, and aggregated to create a knowledge based that can then be exported to a knowledge graph.

Our work on this project will focus on configuring DDEx to extract relevant information about mines and testing the results. Our goal is to obtain both high precision (e.g., >99%) and recall (e.g., >97%) with respect to the target tuples. We note that due to the challenging nature of the text, we expect that we will need to employ some new extraction techniques based on generative models (e.g., GPT4), which we are currently incorporating into DDEx, and this will require fine tuning. Furthermore, we also plan to upgrade the system to identify relevant tables in documents, that can be then parsed as described next (Section 3.3).

3.3 Ingesting Tabular Data from Resource and Production Datasets

Much of the information about production and reserves is contained in tabular format embedded in reports or contained in spreadsheets or databases. For example, Figure 5 shows

	Production[1]			Sales[1]		
(thousand tonnes)	**2022**	2021	2020	**2022**	2021	2020
Highland Valley Copper	**119**	131	119	**127**	124	119
Antamina	**102**	100	86	**101**	99	85
Carmen de Andacollo	**39**	45	58	**39**	45	59
Quebrada Blanca	**10**	11	13	**9**	12	14
Total	**270**	287	276	**276**	280	277

Figure 5: Example table with production data from a mining company technical report.

the production data for the Highland Valley Copper Mine along with three other mines. Also, note that this table contains both production and sales for years 2020, 2021, and 2022. This production data needs to be accurately extracted with the corresponding mine and year of production associated with it.

To accurately extract data from tables, USC will build on our previous work on building semantic models of tabular data [Vu et al., 2021]. In our previous work, we developed an approach for semantic modeling, implemented in a system that we call SEED, that uses graphs to represent possible (n-ary) relationships in the tables and collective inference to eliminate spurious relationships on the graphs. Specifically, we construct a candidate graph containing relationships between table columns and its context values by leveraging possible connections between data in the table and existing knowledge in a knowledge graph (e.g., Wikidata). Then, incorrect relationships in the candidate graph are detected and removed using a Probabilistic Soft Logic (PSL) model. Through collective inference, the PSL model favors links with high confidence, more informative, and consistent with constraints in the ontology and existing knowledge in the knowledge graph.

In order to apply SEED to the problems of extracting mining related data from tables, we will construct a specialized knowledge graph with the relevant data about mines (e.g., from FuseMine output). This will include the set of critical minerals, the deposit types for each of those minerals, and the known mines for each of those minerals and deposit types (initially populated from sources, such as the USGS Bulletin 1693 [Cox and Singer, 1992], which has organized thousands of mines by mineral and deposit types). For the example above, SEED would then use this knowledge graph to identify the first column as a list of mines and link each row to the corresponding mine in the knowledge graph. Then it would identify both the production and sales figures and separate those into the appropriate years. Given this semantic model of the table, the system would be able to associate the production data for each of those mines with the appropriate years of production. This semantic representation is critical to be able to make sense of data coming from multiple sources and reports. To ultimately create the grade and tonnage models, we will need to pull data from many sources in order to get the production data for each year (perhaps spread over many technical reports and/or annual reports) and combine all of the production data with the reserve data from the most recent annual report.

3.4 Deposit Classification, Model Generation, and Mineral Site Data Production

In this section we describe how we combine the reconciled geospatial data sets and the extracted data from the tables and reports to build the grade and tonnage models and produce the mineral site data for training and validation. Given the set of documents and data for a particular critical mineral, such as all the datasets and reports related to tungsten (these could be assembled by scanning the relevant repositories for mining reports and classifying them by mineral type), the system would incremental process the data populating the knowledge graph with the extracted information.

Deposit Classification: As each report or dataset is processed by the system, the deposit type of the given mine needs to be determined if it is not already known. In the knowledge graph, we will have a comprehensive set of deposit types for critical minerals using the recent work on a deposit classification scheme [Hofstra et al., 2021], which organizes the deposit types into a hierarchy. The knowledge graph will also be augmented with descriptive models of minerals from USGS Bulletin 1693 as well as other more recent sources. We will then create a neural network embedding of the descriptive models for each of the deposit types and then place the input document that we need to classify into the same embedding space and find the closest deposit type for that mineral. Before associating a mine with a deposit type in the knowledge graph, the information would first be presented to an expert for confirmation to maintain the high quality of information in the knowledge graph.

Model Generation by Deposit Type: After all the input reports and datasets have been processed, the next step is to produce the actual grade and tonnage models. Since these need to be computer readable models, we will produce these as tabular data with the deposit name, Ore (in metric tons), and grade (in percent), and other relevant features including the source information for every value. Note that producing the grade in percent may require some additional reasoning since some mineral deposits use grams/tonne (= parts per million, where 1 ppm = 0.0001%), and other deposits may report, say lithium, in either lithium % or ppm or lithium carbonate equivalent in % or ppm. We will need to understand the units and then add appropriate conversions to ensure like is compared with like. In our previous work, we built a general system for recognizing and converting units [Shbita et al., 2019], which can be applied directly here. This resulting tabular information can then be easily shared and used to render the standard graphs showing the tonnage against the proportion of deposits, the grade against the proportion of deposits, and a log chart with the tonnage against the grade.

We can create the grade and tonnage models for each of the deposit types for a given mineral by using all the data compiled in the knowledge graph. This can be done by querying the knowledge graph for the data of each of the mines of each deposit type of a critical mineral. The tonnage can be calculated by summing the production over all years and adding the most recent reported reserve numbers. The grade would come directly from the most recent report with appropriate conversions to normalize the grades into a common unit. One of the compelling features about this organization of the information in the knowledge graph is that when new information is added, creating a new grade and tonnage model is simply a query to the knowledge graph to appropriately aggregate all the old and new information. In addition, this organization will allow us to preserve all the source information and provide a natural way for a person using the TA4 capabilities to both review and correct the data in the knowledge graph in the case of extraction errors.

Production of Mineral Site Data Including Mappable Criteria: We will join the reconciled mineral site data with the tabular data organized by deposit type and site name to determine the grade and tonnage attributes for each mineral site. To overcome the challenge that the name might not be inconsistent across the dataset, we will use a fuzzy matching strategy with a pre-trained language model (e.g., BERT). We will also use the possibly duplicated records detected by FuseMine to fine-tune the language model to improve the fuzzy matching robustness on mineral site data.

To identify the mappable criteria, we will first compile a list of names and their synonyms of mappable criteria (e.g., geochemical properties). Then we will use the deposit type for each mineral site to determine appropriate mineral deposits models and mineral systems models and find the mappable criteria in these documents and add them to the mineral site data. We will also detect mappable criteria in each mineral site's attribute data (which often exist as natural language text) and add them as structured attributes to the site.

3.5 Evaluation

We will conduct quarterly evaluations of each of the individual components as well as the performance of the integrated system. For the text extraction, table understanding, and geospatial data reconciliation modules, we will build test sets with ground truth data and report precision, recall, and F1-measure for each component. For deposit classification, we will measure the accuracy of performing deposit classification on known models. And for the construction of the grade and tonnage models, we will evaluate MinMod's ability to reproduce existing grade and tonnage models as well as the precision and recall in producing new models.

4. Capabilities/Management Plan

Capabilities: Dr. Craig Knoblock (PI) is a professor of both computer science and spatial sciences and has managed many large projects over the years. He has extensive experience working on building knowledge graphs from diverse sources of data as well as working with geospatial data. He has worked extensively on projects for extracting data from maps and turning that extracted data into a knowledge graph. He also has conducted several projects on automatically understanding tabular data and linking of that information across sources.

Dr. Yao-Yi Chiang (UMN PI), Ph.D., GISP is an Associate Professor of Computer Science and Engineering at the University of Minnesota. He is an expert in Spatial AI and on ML methods for solving real-world problems with geospatial data. With a unique background and training in AI and GIS, he has led many research projects on spatial data extraction from images, spatial data alignment and linking, and spatiotemporal predictions in several application domains.

Dr. Goran Muric (Inferlink PI) is a Principal Scientist at InferLink and is an expert in network theory, machine learning, deep learning, and agent-based modeling. Dr. Muric has led projects that involve analyzing vast amounts of longitudinal data revealing hidden patterns in the text and images shared by users.

Dr. Simon Jowitt (Consultant) is a tenured Associate Professor of Economic Geology at the University of Nevada, Las Vegas, Nevada, USA. His research focuses on understanding a variety of different mineralizing systems and use of this knowledge in exploration targeting (including prospectivity modeling), mineral economics, global metal resources and the security of supply of the critical elements, and the "economic" side of economic geology.

Management Plan: Dr. Knoblock will serve as the PI and manage the overall project and the coordination across the work at USC, UMN, and Inferlink. He will hold weekly research meetings with the entire team to ensure that there is close coordination across groups. He will have a project Slack to allow easy communication across the entire team and use a shared github repository to have a single place where all the code is shared. He has collaborated extensively in the past with both UMN and Inferlink on previous projects.

Dr. Knoblock will lead the work at USC on table understanding to extract and process production data from datasets and reports as well as the work on classifying deposits and creating grade and tonnage models from the data. Dr. Yao-Yi Chiang will lead the work on geospatial reconciliation and producing the mineral site data. Dr. Goran Muric will lead the work on extracting attributes from text documents. Dr. Simon Jowitt will provide expertise on creating grade and tonnage models and on classifying mineral deposits by deposit type. He will also help find relevant data and documents and provide feedback on the working system.

5. Bibliography

Chiang, Y.-Y., Duan, W., Leyk, S., Uhl, J. H., and Knoblock, C. A. (2020). Using Historical Maps in Scientific Studies: Applications, Challenges, and Best Practices, Springer (ISBN 978-3-319-66908-3)

Chiang, Y.-Y., Leyk, S., and Knoblock, C. A. (2014). A Survey of Digital Map Processing Techniques. ACM Computing Surveys, 47(1):1–44. doi: 10.1145/2557423

Chiang, Y.-Y., Chen, M., Duan, W., Kim, J., Knoblock, C. A., Leyk, S., Li, Z., Lin, Y., Namgung, M., Shbita, B., & Uhl, J. H. (2023 (in press)). GeoAI for the Digitization of Historical Maps. In Handbook of GeoAI. Springer.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT. http://arxiv.org/abs/1810.04805

Green, C. J., Lederer, G. W., Parks, H. L., & Zientek, M. L. (2020). Grade and tonnage model for tungsten skarn deposits—2020 update (No. 2020-5085). US Geological Survey.

Hofstra, A. H., Lisitsin, V., Corriveau, L., Paradis, S., Peter, J., Lauzière, K., ... & Kreiner, D. C. (2021). Deposit classification scheme for the Critical Minerals Mapping Initiative Global Geochemical Database (No. 2021-1049). US Geological Survey.

Kejriwal, M., Knoblock, C. A., & Szekely, P. (2021). *Knowledge graphs: Fundamentals, techniques, and applications*. MIT Press.

Li, Z., Kim, J., Chiang, Y.-Y., & Chen, M. (2022). SpaBERT: A Pretrained Language Model from Geographic Data for Geo-Entity Representation. In arXiv [cs.CL]. arXiv. http://arxiv.org/abs/2210.12213

Mauk, J. L., Karl, N. A., San Juan, C. A., Knudsen, L., Schmeda, G., Forbush, C., Van Gosen, B. S., Mullins, M., & Scott, P. (2021). The critical minerals initiative of the US geological survey's mineral deposit database project: USMIN. Mining, Metallurgy & Exploration, 38(2), 775–797. https://doi.org/10.1007/s42461-020-00372-w

Mudd, G. M. & Jowitt, S. M. (2022) The New Century for Nickel Resources, Reserves and Mining: Reassessing the Sustainability of the Devil's Metal. Economic Geology, 117, 1961–1983.

Mudd, G. M., & Jowitt S. M. (2018) Growing Global Copper Resources, Reserves and Production: Discovery is Not the Only Control on Global Metal Resources. Economic Geology 113, 1235–1267.

Schweitzer, P. N. (2019). Record quality tables for the mineral resources data system. US Geological Survey Data Release.

Shbita, B.; Rajendran, A.; Pujara, J.; and Knoblock, C. A. (2019), Parsing, Representing and Transforming Units of Measure. In Proceedings of the Conference on Modeling the World's Systems.

Singer, D.A. and Bliss, J. D. (1990), Program to Prepare Standard Figures for Grade-Tonnage Models on a Macintosh, Editor(s): J. Thomas Hanley, Daniel F. Merriam, In Computers and Geology, Microcomputer Applications in Geology 2, Pergamon, https://doi.org/10.1016/B978-0-08-040261-1.50026-X

Vu, B. and Knoblock, C. A. (2022), SAND: A Tool for Creating Semantic Descriptions of Tabular Sources, Demo at ESWC '22 European Semantic Web Conference.

Vu, B.; Knoblock, C. A.; Szekely, P.; Pham, M.; and Pujara, J. (2021), A Graph-Based Approach for Inferring Semantic Descriptions of Wikipedia Tables. The Semantic Web – ISWC 2021, pages 304–320, Springer International Publishing.

Vu, B., Knoblock, C. A., and Pujara, J. (2019), Learning Semantic Models of Data Sources Using Probabilistic Graphical Models, In Proceeding WWW '19 The World Wide Web ConferencE.

Werner, T. T., Mudd, G. M., Jowitt, S. M., & Huston, D. (2023) A global assessment of rhenium resources. Resources Policy, 82, Article 103441.

Dr. Craig Knoblock is the Executive Director of the USC Information Sciences Institute and a Research Professor of both Computer Science and Spatial Sciences at the University of Southern California. He received his Ph.D. from Carnegie Mellon University in computer science. He has published more than 400 journal articles, book chapters, and conference and workshop papers on these topics and has received 7 best paper awards on this work. He is the co-founder of Geosemble Technologies, which developed geospatial data integration techniques to link text documents to locations and was acquired by TerraGo in 2011. Dr. Knoblock is a Fellow of the Association for the Advancement of Artificial Intelligence (AAAI), the Association of Computing Machinery (ACM), and the Institute of Electrical and Electronic Engineers (IEEE).

Dr. Yao-Yi Chiang is an Associate Professor in the Computer Science & Engineering Department at the University of Minnesota. Previously, he held the position of Associate Professor (Research) in Spatial Sciences at the University of Southern California. He earned his Ph.D. in Computer Science from the University of Southern California and his bachelor's degree in information management from the National Taiwan University. Before pursuing his Ph.D., Dr. Chiang worked as a research scientist for Geosemble Technologies and Fetch Technologies in California, where he co-invented a patent on geospatial data fusion techniques.

Dr. Goran Muric, is a Principal Scientist at InferLink, is an expert in network theory, machine learning, deep learning, and agent-based modeling. Dr. Muric has led projects that involve analyzing vast amounts of longitudinal data revealing hidden patterns in the text and images shared by users. Muric co-led the technical efforts in DARPA's SocialSim program, whose goal was to develop a scalable agent-based simulation framework, calibrated on social signals data. His team won one of the first DARPA SocialSim challenges by outperforming other teams in accurate agent-based simulation of online social environments.

Dr. Simon Jowitt is a tenured Associate Professor of Economic Geology at the University of Nevada, Las Vegas, Nevada, USA. His research focuses on understanding a variety of different mineralizing systems and use of this knowledge in exploration targeting (including prospectivity modeling), mineral economics, global metal resources and the security of supply of the critical elements, and the "economic" side of economic geology. His expertise in the latter area is demonstrated by a number of recent publications on global base, precious, and critical metal and mineral resources and the impact of COVID-19 on the global minerals industry. He has published more than 110 scientific papers and peer-reviewed book chapters since 2010, is currently the Vice-President for Student Affairs for the Society of Economic Geologists (SEG) and was awarded the SEG's Waldemar Lindgren Award in 2014.