

Mindat.org: The open access mineralogy database to accelerate data-intensive geoscience research

JOLYON RALPH¹, DAVID VON BARGEN¹, PAVEL MARTYNOV¹, JIYIN ZHANG², XIANG QUE², ANIRUDH PRABHU^{3,§}, SHAUNNA M. MORRISON^{3,||}, WENJIA LI², WEILIN CHEN², AND XIAOGANG MA^{2,3,*,*†,‡}

¹Hudson Institute of Mineralogy, Keswick, Virginia 22947, U.S.A.

²Department of Computer Science, University of Idaho, Moscow, Idaho 83844, U.S.A.

³Earth and Planets Laboratory, Carnegie Science, Washington, D.C. 20015, U.S.A.

ABSTRACT

The mindat.org website (Mindat) has been operating since October 2000 as a free, crowd-sourced, and expert-curated database particularly focused on mineral species and their occurrences worldwide. The project has transformed from a hobbyist site in the beginning into a resource that has found use in various scientific research projects and educational programs. Together with other open data resources, Mindat has helped accelerate scientific discoveries in many fields, such as mineral evolution, mineral ecology, and the co-evolution of the geosphere and biosphere. Recently, through open data efforts, machine interfaces and software packages have been established to enable flexible data discovery and download from Mindat. We assume that the data access and usage will further scale up in the next years. Although Mindat is curated by a team of geoscience and database experts across the world, the crowd-sourced records in Mindat possess some bias. In this paper, we first present an overview of the primary data subjects in Mindat and then give extensive details about the characteristics and partiality of three of the most popular data subjects: locality, mineral species, and mineral occurrence. In the discussion, we also give an outlook on appropriate data usage and future extension of data records. We hope users can obtain a more comprehensive view of the Mindat database through this paper and thus better plan their data use. We also hope more people will be inspired to contribute to the data curation work to make Mindat a sustained data ecosystem for geoscience research.

Keywords: Mineralogy, open data, data bias, data science, data-driven discovery, Mineral Informatics: Revolutionizing Mineralogy, Petrology, and Geochemistry

INTRODUCTION

Mindat.org (Mindat) is a large crowd-sourced database of mineral species and their distributions. It has been operating online since October 2000, and as of May 2024, the database contains about 27 TB of data, including information about ~6036 mineral species, 400 000 localities, 1.54 million geomaterial (mineral, rock, commodity, and other natural geological materials) occurrences at localities, and 1.27 million photographs. Of the 77 000 registered users, 650 have contributed to the locality records, 1425 have edited geomaterial occurrences, 7000 have uploaded photos, and 9000 have posted messages. The global base maps used on Mindat include OpenStreetMap, Google Maps (and satellite images), and Macrostrat (Peters et al. 2018) geologic maps. Regionally available are topographic and other specific maps, such as those made available by Ordnance Survey in the U.K. A team of about 100 mineralogical experts volunteer as data curators to review and cleanse the records and photos with input from other members. Unlike the Wikipedia system where

anonymous edits are possible, all contributions must be approved by registered users and are peer-reviewed by regional experts. Mindat is unusual for a large scientific database project in that it is almost entirely funded by voluntary contributions. It is owned by the Hudson Institute of Mineralogy, a 501(c)(3) not-for-profit organization, and is not directly affiliated with or sponsored by any academic institution. Donations range from a large bequest from the estate of Rock H. Currier, a past Mindat website manager to \$10 individual donations from visitors to the site.

Mindat is actively used by professionals and the general public, spanning from cutting-edge research on the evolutionary system of mineralogy (Hazen 2019), mineral exploration and mining (Dallaire-Fortier 2024), elementary to college education (Landicho 2021), to the hobbies of rock and mineral collecting (McGill 2020). The front-end website of Mindat has received about 60 million page views annually from more than 10 million unique visitors in the recent few years. Many researchers have requested access to the back-end database to query and download data sets of interest, but such an open data server had never been established due to a shortage of financial support, and the data requests could only be met case by case courtesy of the help from the Mindat technical team. Recently, through a grant from the U.S. National Science Foundation, we have implemented the OpenMindat initiative (Ma et al. 2024), and an application programming interface (API) for the database has been established. As the Mindat technical team is gradually making more data

* Corresponding author E-mail: max@uidaho.edu Orcid <https://orcid.org/0000-0002-9110-7369>

† Special collection papers can be found online at our website in the Special Collection section.

‡ Open access: Article available to all readers online under the following license terms: MSA rights usage.

§ Orcid <https://orcid.org/0000-0002-9921-6084>

|| Orcid <https://orcid.org/0000-0002-1712-8057>

subjects and records accessible through the API, we assume the usage will significantly increase soon. The aim of this paper is to help users better understand the database, be aware of the existing biases and concerns, and use the data appropriately. Ma et al. (2024) briefly described the technical structure of Mindat and the challenges it faced in building open data services. The following section below will give more details about its history of development, the current technical structure of the database, and the roles of different user types.

TECHNICAL STRUCTURES OF MINDAT AND ITS USER COMMUNITY

Mindat in its early years

Mindat was started by Jolyon Ralph as a personal database application written in ANSI C for personal computers in December 1993. The original aim was to combine a table of mineral names and their basic properties along with a table of the localities where these minerals were recorded. This initial database was not distributed at all and was only used by the original author in connection with his hobby of mineral collecting. In 1995, a newer version was created to take advantage of the Windows 95 operating system, and through the fledgling internet of the late 1990s, this version was used by, at most, two dozen other people worldwide. The major disadvantage of the system was that only Jolyon Ralph was able to add new data to the database, as it had no tools for merging in changes from different contributors. In the summer of 2000, the code for Mindat was transformed into a web database project, which was subsequently launched as a website (mindat.org) in October 2000. Unlike its previous stand-alone editions, this new version allowed for collaborative contributions, enabling users to add new data to the database. A system was also implemented to facilitate the review and approval of these submitted changes by experienced users. Over time, Mindat has undergone continuous development and evolution into its current state.

Current technical architecture and the OpenMindat initiative

Mindat currently (May 2024) runs on a single dedicated server using the LEMP (Linux, Nginx, MySQL, PHP) development stack. It is almost entirely built from custom code developed by Jolyon Ralph and others over the course of its lifetime and does not rely on any server-side software frameworks. Data are backed up daily and software code also resides on several servers. Caching is done locally on the server using an internal database cache of generated pages along with application-level caching at the web server level. In addition, the Cloudflare service is utilized to reduce the load on the server by providing a Content Delivery Network (CDN) for the content of the Mindat website. Data on the Mindat server are stored in a series of MySQL tables and normalized so that changes in mineral nomenclature [such as the mineral species list approved by the International Mineralogical Association (IMA)] are updated automatically across all uses/occurrences within the database. When a mineral name is changed, all the pages that use that mineral name in a structured format (such as a list of minerals at a locality) automatically update to the new name (subject to caching issues).

The OpenMindat API has been established for Mindat (Ma et al. 2024) to provide findable, accessible, interoperable, and reusable (FAIR) (Wilkinson et al. 2016) open data to interested users, especially researchers who need quick access and download of bulk data sets. The API currently enables access to three major groups of data: IMA mineral species list, geomaterials, and localities. The IMA mineral list is made a specific service because many researchers, including the curators of other geoscience data portals, need a formal and complete list of mineral species and their attributes. While the IMA website (Landucci and Bosi 2023) and the RRUFF database (Lafuente et al. 2015) both have the mineral species list, the former is a PDF document, and the latter cannot be downloaded. In comparison, through the Mindat open data API the mineral species list can be queried and downloaded in many formats. We also continuously collect users' feedback to upgrade the data subjects and attributes on the API. To further smoothen the data access workflow and promote open science, we have also developed open-source software packages in R and Python (Que and Ma 2023; Zhang et al. 2024). Our hope is that those packages will simplify the coding work of researchers for data retrieval from the API, such that they can have more time for data analysis and scientific discovery. More details about the API and the R and Python packages are available in the "Data and code availability" section at the end of the paper.

Mindat user types and their rights in data sharing and curation

The Mindat website is open to all users for browsing whether they have a registered account or not. To contribute content to Mindat, a user account is necessary, and the accounts of different levels of access have different rights when entering data (Table 1). For a user to be upgraded to a higher level of access, they must demonstrate acceptable knowledge in both the mineralogy of the area they are adding data for and the technical tools on Mindat that are available for data entry/review. All entries are available for managers and experts to review within the change logs and daily

TABLE 1. Types of Mindat user accounts

Access Level	Access Rights	Number (approximate, as of September 2023)
Unregistered	Access the site in read-only mode only	954 000 (based on Google Analytics for April 2023)
Level 0 (default for new users)	Read-only access to site except message board where users can post messages and photos	64 650
Level 1 (approved accounts)	Users can propose additions to the locality database and upload photos pending approval	7318
Level 2 (trusted accounts)	Same as level 1 except that submissions are automatically approved; only for users with a proven track-record of high-quality entries	463
Level 3 (management team)	The management team is responsible for approval of entries, adding/editing geomaterial property entries, and deciding the general rules for data entry	55

reports, even those that are flagged for automatic approval. The management team is a diverse group of 55 amateurs and professionals worldwide who help with the day-to-day running of the site. Additionally, around 100 people are assigned as regional experts for a particular area due to their extensive knowledge of that region's mineralogy. These people can help the managers in approvals for items within their own areas of knowledge.

DATA SUBJECTS AVAILABLE IN MINDAT

The most characterized data subjects within Mindat are geomaterials (e.g., mineral/rock/variety) and localities (Fig. 1), together with other subjects such as occurrences, photos, articles, and glossary items.

Geomaterials

Mindat contains an entry for every mineral species approved by the IMA, along with data (where available) on crystallographic, physical, chemical, optical, and other properties. A list of well-known mineral classification and identifier systems, such as the Dana Classification (Dana 1868), the Strunz Classification (Strunz and Nickel 2001), and the Hey's Mineral Index (Clark 1993), are also included in the description of mineral species. The database also contains entries for other types of names—synonyms, varieties, groups, discredited species, mineraloids, synthetic analogs of natural minerals, incompletely described geomaterials, and, more recently, petrological names and meteorite classifications (Table 2). Each entry has its own unique Mindat ID (e.g., quartz - mindat:1:1:3337:0), which is resolvable on the Web (e.g., quartz - <https://www.mindat.org/1:1:3337:0>) (Ralph et al. 2024).

The primary entries are for the valid IMA mineral species. As of early May 2024, there are 6036 species listed on Mindat. Mindat's curators regularly update using the mineral species list PDF document published on the IMA website (Landucci and Bosi 2023) and the IMA Newsletters published in the *Mineralogical Magazine* and *European Journal of Mineralogy*.

Synonyms in Mindat are defined as a name that has a mapping in both directions to an IMA mineral species. For instance, "idocrase" is a synonym of vesuvianite by the understanding that anything that is called "idocrase" could be called vesuvianite, and everything that is vesuvianite could also be called "idocrase" if that name was preferred. The recording of synonyms on Mindat is for convenience and research and does not in any way indicate that such names should be used in preference to the official versions. If occurrence information about "idocrase" is added to a locality record or photograph, it will automatically be corrected to vesuvianite as it does not regard synonyms as being separate entities in these cases (although they do have their own internal geomaterial IDs and a page linking to the approved IMA name). Foreign language variants of names and historical variants of names are the primary use of synonyms in the database.

Varieties on Mindat describe a class of entries where there is a unidirectional mapping from a name to an IMA mineral species (or even another variety). A simple example of this is the term "amethyst." All "amethyst" is quartz, but not all quartz is "amethyst." When adding varietal names to a locality, the ID for the variety is kept and not converted to the parent mineral species directly, so a locality where "amethyst" is reported will have "Quartz Var. Amethyst" listed in the table of species occurrence. The issue of whether varietal names should be encouraged or discouraged is more complex than synonyms. While some would prefer sticking entirely to IMA-approved mineral names, the information that amethyst is found within a deposit would be lost if it were simply recorded as quartz.

Mixtures are names given for materials in the mineralogical realm that have proven to be fine-grained mixtures of different minerals. It is usually used to cover various names of historical significance for specific finds that were first thought to be a new mineral but were later discredited as mixtures of two or more minerals. For instance, "andrewsite" was historically regarded as a distinct mineral species but is now known to be a mixture of hentschelite and rockbridgeite (Dunn 1990).

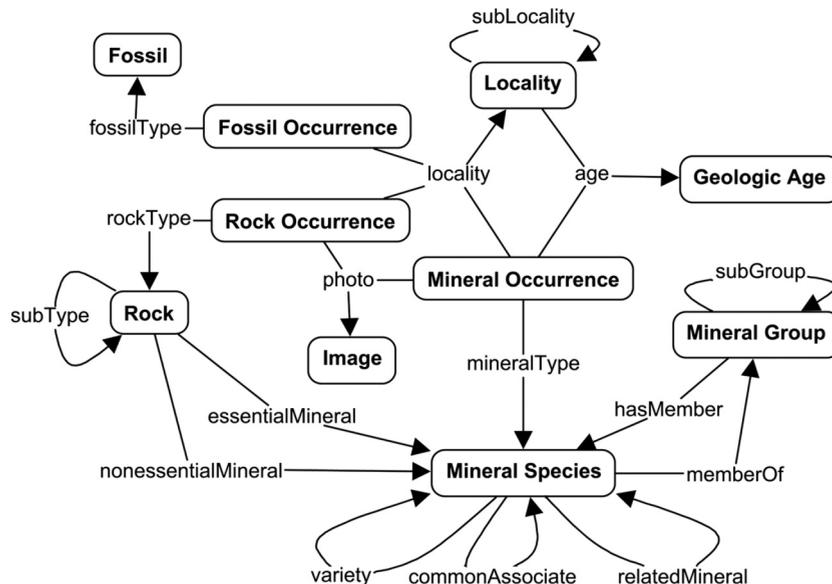


FIGURE 1. Major data subjects in Mindat. Please note the diagram is a simplified representation of the most representative data subjects of Mindat.

TABLE 2. Types of geomaterial records in Mindat

Type	Definition	Example(s)
IMA Mineral	A mineral species approved or accepted by the International Mineralogical Association	Quartz (mindat:1:1:3337:0) Calcite (mindat:1:1:859:4) Pharmacosiderite (mindat:1:1:3185:7)
Synonym	An alternative name for another entry; this could be a historical name or a name in another language	Chalybite (mindat:1:1:6234:9) = siderite (mindat:1:1:3647:0) Antimonglanz (mindat:1:1:6337:7) = stibnite (mindat:1:1:3782:8)
Variety	A named variety of an approved mineral or other variety	Amethyst (mindat:1:1:198:0) Agate (mindat:1:1:51:8) Ruby (mindat:1:1:3473:5)
Mixture	A named mixture of 2 or more minerals	Andrewsite (mindat:1:1:226:2)
Series	A series is two mineral species related by the substitution of two elements at a position in the unit cell	Calcite-rhodochrosite series (mindat:1:1:8682:6): Series Formula: CaCO ₃ to MnCO ₃
Group	A group is two or more mineral species related by similar structure and/or chemistry	Clinopyroxene subgroup (mindat:1:1:7630:8) Autunite group (mindat:1:1:29274:0)
Rock	A name for a type of rock	Syenite (mindat:1:1:48213:0) Sandstone (mindat:1:1:49438:4) Komatiite (mindat:1:1:48568:7)
Commodity	A special entry to record commodity records, which have their own namespace starting with "commodity" to differentiate them from other geomaterials with the same name	Commodity:copper (mindat:1:1:52442:7) Commodity:molybdenum (mindat:1:1:52479:1)
Meteorite classification	A name for a classification of meteorite type based on composition and alteration	CM2 chondrite meteorite (mindat:1:1:49557:8) IIIAB iron meteorite (mindat:1:1:49930:7)
Other	Names that do not fall conveniently into any other category; this can include non-crystalline materials, materials that are not (yet) approved as minerals, anthropogenic materials found within the geological environment, and discredited materials	Carbon dioxide ice (mindat:1:1:25561:5) Horsfordite (mindat:1:1:1932:9)

Series and groups are special classes of entries that describe either a series between two minerals or groups of two or more related minerals. Good examples of series include the plagioclase series (albite-anorthite) and wolframite (ferberite-hübnerite). Tourmaline is a good example of a group with 41 species (including dravite, elbaite, and schorl) and an additional four unnamed members. Groups are defined by the IMA-CNMNC (Mills et al. 2009) in the following way: "A mineral group consists of two or more minerals with the same or essentially the same structure, and composed of chemically similar elements." Groups can themselves be grouped into supergroups or divided into subgroups. Mindat follows the standard group definitions as published by the various IMA subcommittees, such as with the amphibole group (Leake et al. 1997; Hawthorne et al. 2012).

In many cases, there may be insufficient information about the record of a mineral at a locality to define it exactly to species level, and in these cases, it is sensible to record it at the group level (if that can be ascertained). For example, "garnet" is listed based only on visual observation (rather than guessing based on appearance/association as has often been done historically). In other cases, this is not possible. Older references to "eudialyte" may refer to one of the other members of the eudialyte group (such as kentbrooksites and ferrokentbrooksites) which cannot be distinguished visually, and without modern analytical data to correctly classify the species, the term "eudialyte group" should be used. This is a case where, when listed as present at a locality, the term "eudialyte group" specifies that the mineral in question is one (or possibly more) as-yet unclassified species within the eudialyte group; it does not indicate the entire eudialyte group family of minerals is present.

Mindat now has a database of 3094 petrological names in a complex hierarchical arrangement based primarily on the definitions published by the International Union of Geological Sciences (Le Maitre 2005; Fettes and Desmons 2007) and then supplemented by documentation from the British Geological Survey (BGS 2020). In particular, the BGS ruling for always using hyphens in names allows us to distinguish the difference in naming between

quartz-syenite (which is a root name) from biotite gneiss (which is a field name for a gneiss containing biotite). We also track updates in those community standards regularly (several times a year) to add new terms to the name list.

One of the key problems in building a hierarchical structure of rock types is that rocks are, by their nature, very much more complicated than minerals, and in many ways, classifying rocks can be highly subjective. There is also no controlling group that certifies names, unlike the IMA, which oversees mineral nomenclature. Volcaniclastic sediments can be classed as either sedimentary (if one is studying their deposition, for example) or igneous (if one is studying their composition). This is why Mindat allows for rock types to be placed in two different positions within the hierarchy at once (essentially, a child item can link back to two different parent branches). Metamorphic rock nomenclature is complicated by using different names depending on whether one is describing something as it is now (its mineralogy and texture, e.g., marble) or from what it came from (e.g., meta-limestone). As both names are entirely valid, the synonym structure used for minerals in Mindat does not apply to the rock names—not every meta-limestone is a marble.

Commodities on Mindat are geomaterials of economic value to human society. A special namespace prefix "commodity" is used for commodity records, so "commodity:gold" (mindat:1:1:52454:2) refers to gold as an ore, while "gold" refers to gold as a native mineral species (mindat:1:1:1720:2).

Meteoritic terminology is a special case of rock types, where standard coding as defined by the Meteoritical Society for meteorite types is used to both describe their composition and state of metamorphism (Meteoritical Bulletin Database 2023). For example, the CM2 chondrite meteorite has two parental branches: CM chondrite meteorite and Petrologic Type 2 chondrite meteorite, as shown in Figure 2.

In addition to the above-mentioned data subjects, Mindat also hosts many records that are difficult to group into existing types, such as the Carbon Dioxide Ice (mindat:1:1:25561:5) mentioned in the row "Other" in Table 2.

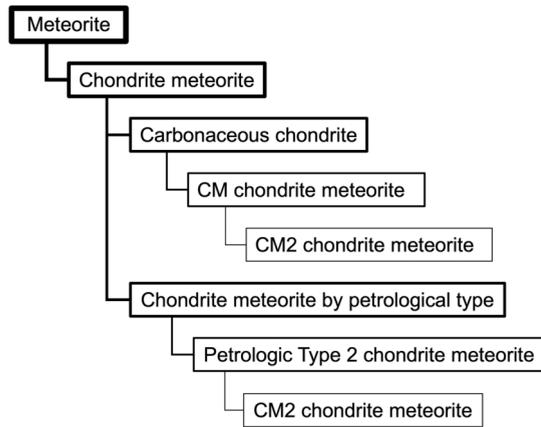


FIGURE 2. Tracking the hierarchy of meteorite classification with the CM2 chondrite meteorite as an example at the lowest level. The carbonaceous chondrite is a major class of chondrites that mostly have Mg/Si ratios near the solar value and oxygen isotope compositions that plot below the terrestrial fractionation line. The CM chondrite meteorite is the Mighei (CM) chemical group of carbonaceous chondrites, distinguished by small chondrules and refractory inclusions (0.3 mm), abundant fine-grained matrix (~70 vol%), and abundant hydrated minerals. The petrologic Type 2 chondrite meteorite designates chondrites that are characterized by abundant hydrated minerals and abundant fine-grained matrix, in which chondrules are present and sulfides are Ni-bearing.

Localities

A locality in Mindat is a point or region where geomaterial occurrences can be recorded. Unlike prior database systems, which frequently have different fields for mine name, town, state, and county, the structure of localities within Mindat is internally hierarchical but with no fixed meanings given to individual levels. Based on data returned from manned and unmanned missions and satellite reconnaissance, localities in Mindat are not limited to the Earth and include localities on the Moon, Mars, and other celestial bodies.

The normal Mindat hierarchy uses current political boundaries to categorize localities. As shown in Table 3, the major type of locality is a mine or prospect, but we can also use geological formations, geomorphological regions (e.g., mountains, lakes), and anthropogenic sites (e.g., road cuts, tunnels). Of the ~400 000 localities in Mindat, 246 000 have a latitude/longitude pair, and another 109 000 have boundaries, of which 91 000 are political entities. Of 252 000 localities, 12 000 have different associated minerals, rocks, or commodities. While localities such as individual outcrops and small mines are recorded essentially as a point record (with latitude and longitude where known), larger regions can be defined based on one or more polygons (using the

GeoJSON format for data storage). JSON (JavaScript Object Notation) is a data format used for structured information exchange on the Web. GeoJSON is a JSON-based format specifically designed for encoding geospatial data, including points, lines, and polygons, along with related attributes.

The complexities of how localities are defined deserve a specific study, but essentially, the problems of mapping recorded localities to the Mindat locality system can be demonstrated with examples of historical records based on old mining districts and from geographical regions. A historical label stating that a particular specimen was found in the “ore mountains” could be mapped today to either Germany or Czechia (Czech Republic) or cross political boundaries. Additional examples of localities are places such as parks, mountain ranges, and geological formations. To solve these problems, Mindat has added another way of entering what it describes as “non-hierarchical localities” that sits parallel to the hierarchical locality structure and can be defined with GeoJSON multipolygons or lists of localities to include or exclude. The pages for these non-hierarchical regions build up mineral lists based on all the point localities that fall within the defined polygon region.

Another unique characteristic of Mindat is that non-hierarchical localities (about 22 900 entries) are used to set up a hierarchy of geological regions. A standard (political) locality might be based on the hierarchy “Mine name, Town area, County, State, Country,” but the same locality could be described as “Mine name, Geological unit, Geological sequence, Basin/Range, Continental Plate.” This secondary hierarchy allows researchers to cluster localities by geology, which leads to better statistical results. Most of these localities are defined by boundaries, but those lacking boundaries can also be generated by including specific Mindat ID numbers. We also use some outdated political boundaries, such as mining divisions in British Columbia, Canada, mining districts in the Western U.S.A., and shires in Australia. A lot of the geological information in these areas uses these divisions. We can also have erratic localities where samples come from origins, which are some distance from where they are currently found. These include glacial deposits, alluvial deposits, and meteorites.

Other popular data subjects among users

Geomaterial occurrences. Internally called “locentries,” these each are a record of a connection between a geomaterial and a locality, such as the record “Akdalaite from Emerald Creek, Latah County, Idaho, U.S.A.” (mindat:1:3:969699:0). There are various levels of confidence in this based on whether there is any doubt in the record. In cases where a geomaterial has been disproven at a particular locality (or, at least, that

TABLE 3. Mindat locality types

Type	Number (approximate) of localities	Example
Mines—producing mines, prospects, projects	201 000	Don León Mine, Angaco Department, San Juan Province, Argentina (mindat:1:2:434759:3)
Political—mainly used for rough locality also used in estimating latitude/longitude along with nearby localities	91 000	Idaho, U.S.A. (mindat:1:2:3724:5)
Geomorphic	25 000	Hoodoo Mountains, Idaho, U.S.A. (mindat:1:2:249752:8)
Geological—massifs, outcrops, non-mine areas	8000	Sawtooth Batholith, Idaho, U.S.A. (mindat:1:2:3745:0)
Meteorites—mainly fall sites	1600	Canyon Diablo meteorite, Arizona, U.S.A. (mindat:1:2:3317:9)
No type set—indeterminable from name for legacy localities prior to adding type as a database field	87 000	Switchback, Adams County, Idaho, U.S.A. (mindat:1:2:141009:2)

the existing recorded sample of a geomaterial from a locality was either proven to be something else or from somewhere else), it can be recorded as outdated and is displayed on Mindat with a strikethrough rather than left entirely off the list. This is important for ensuring that people do not re-add the incorrect data to the page by mistake when reading older references.

Mindat also links geomaterial occurrence records to references for the information source and identification methods for geomaterials not referenced from a published source. Mindat is now building a large database of reference files, including PDFs of content subject to permission. There are many different information sources, including professional journals, national and state/provincial geologic surveys (open file reports and information from websites), open technical reports from the mineral exploration industry, amateur publications, and individual observations—visual, dealers, and commercial mineral identification companies. Geomaterial occurrences can have additional information associated with them. This includes fields for mineral depositional environments (primary, secondary, post-mining/anthropogenic), quality, rarity, habit, color, luminescent properties, age, chemical analysis, and general comments.

Photos. The photo database on Mindat is the most visually appealing area of the website but is also information rich. Since its launch, Mindat has been allowing uploads of mineral specimens, fossil, and locality photographs. Mindat now has over 1.27 million photographs within the system, giving a powerful record of habits and associations of mineral species. While copyright is retained by the majority of photographers, a significant number are either under a Creative Commons license or in the public domain for open access and reuse.

DISCUSSION

Mindat, together with other open data resources such as RRUFF (Lafuente et al. 2015), Macrostrat (Peters et al. 2018), EarthChem (Lehnert et al. 2007), PetDB and GEOROC (Lehnert et al. 2000), have significantly facilitated the data-driven discoveries in geosciences in the past decade. Nevertheless, from our point of view as data curators, we present some characteristics and issues related to the Mindat records, including the bias of locality density in different areas across the world, the desire for new mineral species, and imbalanced records of geomaterial occurrence among localities. We hope Mindat data users are aware of these concerns, and we call for more community efforts to enrich Mindat records and mitigate those issues.

A boost to data-intensive geoscience discovery

As mentioned above, Mindat has been widely used on many research topics, spanning from mineralogy, mineral exploration, petrology, and geochemistry to planetary science. Here, we briefly list the recent progress made in mineral evolution, mineral ecology, and the co-evolution of geosphere and biosphere. Mineral evolution proposes that the mineralogy of planets and moons undergoes transformation through various physical, chemical, and biological processes, resulting in the formation of new mineral species (Hazen et al. 2008). Similar to other data-driven studies, mineral evolution research has faced challenges due to the complex synthesis of multidisciplinary data.

The three subjects of time, location, and redox state have been applied as axes to synthesize various data sets, including those from Mindat, to advance the research. Progress has been achieved on the mineral species of different elements and types (e.g., Grew et al. 2019) and the association between mineral evolution and the big events in Earth's history (Bradley 2015). Recently, Hazen and Morrison (2022) identified 57 paragenetic modes to illustrate the evolving formational environments across 5659 mineral species. Their findings suggest the potential to enhance current mineral classification systems.

Mineral ecology studies the diversity and distribution of mineral species on planetary bodies (Hazen et al. 2015). Using records from Mindat and other sources, Hystad et al. (2015) estimated the number of undiscovered mineral species on Earth. They observed that the frequency distribution of mineral species on Earth's surface follows the "Large Number of Rare Events (LNRE)" statistical pattern. This pattern indicates that most mineral species are rare, occurring in fewer than five localities, while a few, like quartz, are widespread. The LNRE model enables predictions about new species that may be discovered as mineral occurrence data expands. Hystad et al. (2019) refined this model by introducing a new Bayesian estimation, which improved accuracy and provided more reliable predictions.

The fields of mineral evolution and ecology contribute to a broader research theme—the co-evolution of the geosphere and biosphere, which explores the interactions between living and nonliving components throughout Earth's history (National Research Council 2008). Over the past two decades, data from various geoscience disciplines have increasingly incorporated deep-time attributes (4D Initiative 2018; Wang et al. 2021). These efforts have connected records of redox-sensitive elements across mineralogy, petrology, and geochemistry, linked relative ages of enzymes utilizing redox-sensitive transition elements, and integrated paleobiological and thermochemical data. This enriched data foundation allows scientists to gain new insights into the evolving oxidation states of Earth's near-surface environments, oceans, and atmosphere. For instance, Moore et al. (2018) explored the critical biological roles of cobalt and the geological and chemical factors influencing its utilization in early life evolution. In a subsequent study, Moore et al. (2022) applied network analysis to mineral element electronegativity and hard-soft acid-base properties, revealing that the evolving chemical composition of the Earth's crust was directly shaped by orogenic events, mantle redox state changes, planetary oxygenation, and climatic shifts.

These studies underscore the significance of well-curated data resources, advanced data science methods, and strategic research questions in driving scientific breakthroughs. As data-driven approaches continue to gain momentum in geosciences, Mindat and other open data resources will remain pivotal in various research areas in geoscience. That has also been a major driving force for us to build the Mindat open data API and the Python and R packages to allow scientists to quickly query and access the data in preferred formats. Recently, we have also become aware of the thriving of large language models (LLMs), and we have developed LLM-assisted experiments to help scientists with data queries and initial analysis. For example, a scientist can type data needs in a paragraph of natural language, and the LLMs will help

translate the needs into computer code for the query and run it toward the Mindat open data API to retrieve the data records. Similarly, LLMs can also be used as an assistant or guide in the process of data analysis. As the technologies mature, we will be able to integrate them as part of the Mindat user interface to make the Mindat open data service more friendly to end users.

Curators' thoughts and views on the bias of Mindat data

Bias of Mindat locality records. A locality in Mindat usually consists of a mine or rock outcrop that has been published in a journal or defined in some "official" report with descriptions of the mineralogy, petrology, or commodities mined. This would not include many prospects that were claimed but never entered into production or were described. There are also records of meteorite falls that record the types of meteorites found. Mines are much more likely to enter into official records and publications, while outcrops are more likely to be found in field notes and become a small part of a geological map. We normally do not further split a locality unless there are some distinct records associated with parts of it. In the locality of Tsumeb, Namibia (Gebhard 1999), studies have shown there were three distinct periods of gossan formation separated by depth. Areas of low density for localities include those covered by water, ice, glacial deposits, deserts, and tropical saprolites, as well as those of low topographic relief. Higher densities of localities will include areas of temperate climate, high topographic relief, greater population densities, and areas of denser infrastructure.

Figure 3 and Table 4 show that the U.S.A. and Western Europe have the greatest densities of localities in Mindat. The Northern Hemisphere contains some 68% of the Earth's landmass. Studies of regional mineralogy have been printed in English or Western European languages, and these areas tend to have more universities and state-run geological surveys, which produce more

TABLE 4. Patterns of Mindat localities by country

Country	State	Localities (as of September 2023)	Loc/1000 km ²
U.S.A.		155 394	18.2522
	California	28 422	67.0378
	Colorado	13 468	49.9117
	Arizona	10 727	36.3339
	Louisiana	54	0.3989
	North Dakota	50	0.2730
U.K.		17 997	96.2974
Australia		15 115	2.3726
China		7181	1.0378
Germany		5420	27.0497
Norway		5251	17.8333
Canada		4678	0.6534
France		4581	58.1836
Italy		4085	11.0228
Austria		3992	82.3922
Russia		2613	0.2321
Spain		2086	10.6108
Japan		2183	11.0228

literature. Historically, mining companies have also been more prevalent in these areas, having been active for longer periods. Moreover, these areas also have greater numbers of amateurs interested in mineral collecting. The density of localities in the U.S.A. (Fig. 4) has been aided by several Mindat members who have systematically added data to individual states, the availability of literature such as state mineralogy, hobby publications, and the relatively easy ability to import data from the USGS MRDS database (Schweitzer 2019). Most of the U.S.A. localities in Mindat are mines and prospects that were discovered between 1850 and 1910. Since Mindat started out as a site for mineral collectors, localities that produced mineral specimens for the trade have better coverage in the database. A similar condition exists in Europe with respect to people and literature. Type localities for minerals and complex mineral assemblages such as carbonatites and fumarole localities are more likely to be entered into Mindat.

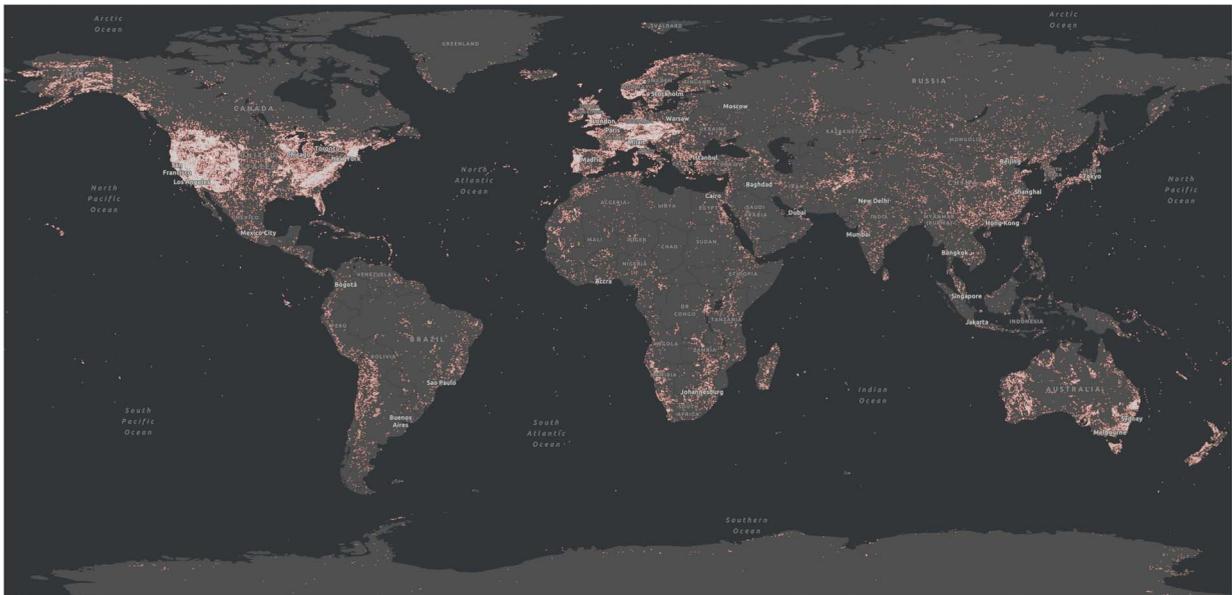


FIGURE 3. Mindat localities across the world. While Mindat has strong links with geological surveys, mining bureaus, museums, and similar organizations in the U.S.A., Europe, and Australia, it lacks such contacts in other areas, especially the Global South. There is a big potential for Mindat to enrich locality records in those areas.



FIGURE 4. Mindat localities in the U.S.A. The U.S.A. is a good example of what factors influence locality density. Areas of high density include the Appalachian Mountains and the mountains in the Western U.S.A. These areas are of high relief, with significant amounts of igneous and metamorphic rocks. The Great Plains is an area of low density except for some Mississippi Valley type deposits and the igneous/metamorphic rocks of the Black Hills. These areas have low relief and a dearth of outcrops. The Atlantic coastal plain is also an area of low density with similar conditions. Western areas are of low relief, such as the Central Valley in California, the Columbia River basalts, and the Four Corners region. On a smaller scale, the basins in the Basin and Range province are sediment-filled and low-relief areas that contain few localities since the depth of any ores is too deep to mine. Two areas in the West that have a high density of localities are the Sierra Nevada Mountains and the Colorado Mineral Belt. These are areas of high-relief and igneous rocks.

Comparison between Mindat and other sources of locality records. We can compare Mindat to other sources for localities in different regions, which will help us get a reasonable estimate of the number of localities that we are missing.

An interesting question we can ask is: How many localities are there across the world, and can any more be represented in Mindat? The USGS lists some 3000 major mines on Earth (Schweitzer 2019). Some official publications contain localities at the prospect level, which can provide about 200 000 to 300 000 potential new locality records to Mindat. Moreover, there can be other additional geological sampling points. We can also think about the potential locality records in another way. The Earth's land area, 148 326 000 km² with 395 000 Mindat localities (in September 2023), generates a density of 2.65 localities/1000 km². If we assume an average of 20 land localities/1000 km², it will raise the number of localities up to 3 000 000.

Of the “interesting” hydrothermal ocean vent sites, the total number of listings in InterRidge Vents Database, the completed Version 3.4 includes 721 vents (33 more than Version 3.3), with 666 confirmed or inferred active and 55 inactive (Beaulieu and Szafranski 2020). Mindat lists 110 localities in oceanic areas, some 15% of the world total.

Total drilling by the various expeditions for deep drilling of the ocean basins has been about 2400 holes with 400 000 m of core recovered. There have also been numerous drag samples recovered in the search for deposits of ferromanganese nodules and crusts. Significant wells have been drilled on the continental shelves for oil and gas production. The most interesting mineralogical (and biological) ocean sites have been the geothermal

vent fields. In contrast, Mindat still has a relatively low density of ocean locality records. Earth's ocean area of 361 132 000 km² with the 1000 Mindat localities (in September 2023) generates 0.003 localities/1000 km².

Canada has a similar distribution of mines compared with U.S.A. (Historical Canadian Mines 2024). The Rocky Mountain provinces have 2.91 mines/1000 km²; the Prairie provinces have 0.54 mines/1000 km²; the Eastern provinces in glaciated areas of Precambrian rocks have 2.77 mines/1000 km², which removed any sedimentary rock cover. In comparison, the Arctic regions of Canada above latitude 60° N have 0.09 mines/1000 km². This probably is because the Arctic regions have a relatively short prospecting season, low population density, and a dearth of infrastructure for moving equipment in and ore out, which requires rare rich deposits to justify prospecting and production. Overall, there is still potential for Mindat to import new locality records for Canada.

For China, Ottens (2005) stated there are 200 000 registered mineral deposits and 280 000 workings registered as privately or communally owned in China's 9 597 000 km² for 20–30 localities/1000 km². For instance, Jiangxi Province has superior metallogenic geological conditions and is extremely rich in mineral resources. There are 164 kinds of mineral resources found in China, 153 of which can be found in Jiangxi Province, with more than 5000 mineral-producing areas (Rong et al. 2023). The 167 000 km² of Jiangxi with 446 Mindat localities currently only generates a density of 6 localities/1000 km². There is a big potential to increase the locality density for many areas of China in Mindat.

The French Bureau de Recherches Géologiques et Minières (BRGM) created a database of some 1000 prospects in Mauritania (Marsh and Anderson 2015). At the time before this data was added to Mindat, there were some 10 Mindat localities. After the database entries were made, this resulted in a locality density of 1 locality for 1000 km² over 1.031 million km² of surface area in a desert environment. Similar records should also be available for many other African countries.

For the U.K., we can use the Peak District National Park as an example to show the potential. There are 3200 mineral veins (i.e., lead, fluor spar) of the Southern Pennine Orefield within the park captured as a single data set in 1983 from BGS 1:10 560 published maps with additional veins from referenced literature (Plant and Jones 2013). The data covers a small, very limited area and includes several pipe and flat deposits as well as mapped faults. The data set is approximately 99.5% complete. However, even with 90 localities for 1000 km² in the U.K., there are only 300 Mindat localities within the park (1440 km²). So, there is also potential to add more Mindat localities for the park as well as other areas in the U.K.

For meteorites, the percentage of meteorite localities that are in Mindat compared to the Meteorite Bulletin Database (2023) range from 0.32% for Antarctica, 0.69% for NWA, to 12.08% U.S.A., and 13.58% for martian meteorites. We average about 2.56% of the total number of meteorites in the Meteorite Database, with the more common localities having less coverage in the Mindat database. Ice-covered (Greenland and Antarctica) and desert landscapes provide good collecting sites for meteorites but seldom are sampled for their mineralogical contents.

Bias of Mindat mineral records. Common minerals have large numbers of records in Mindat, but there should be more. Common minerals in the Mindat database can be classified into three groups: (1) common rock-forming minerals—quartz, calcite, muscovite, magnetite, hematite, chlorite; (2) minerals that are important ore minerals—hematite, magnetite, pyrite, gold, chalcocopyrite, sphalerite, galena; and (3) gangue minerals—limonite, calcite, quartz, pyrite. Globally, these are still underrepresented in Mindat, since many localities not on Mindat will contain these common minerals. For minerals recorded only from a single locality, Mindat has 100% of the worldwide localities. Rare minerals are overrepresented in the mineral literature since common minerals rarely justify their study and publication. Regional mineral books also emphasize the rare at the expense of the common minerals, such as type localities and where a mineral was first found in the area.

More work needs to be done on specimen records in Mindat, including species identification and locality information. Except for easily identified minerals such as quartz, calcite, pyrite, chalcocopyrite, sphalerite, galena, etc., specimens need to be sampled and identified from localities. The prospects and mines have to be available for sampling by collectors (not reclaimed or subjected to heap leaching or converted to housing) and have people with free time to search and collect within an area. Minerals that are colored, have elements with economic value, or form decent-sized crystals that are desired by collectors tend to have more records in Mindat. Mines that have large amounts of rocks excavated are more likely to contain a diversity of minerals. The generation of secondary minerals will increase the number of

minerals in a locality. Localities that form interesting deposits with unusual chemistry and unique formation temperatures and pressures, such as fumaroles, alkalic rocks, carbonatites, and meteorites, tend to be more diverse. There are a limited number of scientists who are willing to work on specimen identity and have the desire, time, training, and equipment to study new minerals. Researchers attending mineral shows such as Tucson often develop groups of amateur collectors who bring them specimens of minerals that they have not been able to identify.

Studies for theses and dissertations tend to add minerals to a locality since the locality can be more thoroughly sampled, analyzed, and published on the Internet. Many collectors are curious enough about their specimens to use commercial identification services to confirm the mineralogy where visual identification is not definitive. Studies are not usually done in areas with simple mineralogy, even though they move large volumes of rock, such as banded iron formation, unless mining or beneficiation problems are involved. Amateur mineralogists arise when members of an educated citizenry who are not at subsistence levels develop an interest in minerals. One of the most mineralogically diverse localities on Mindat is the Clara mine, Germany (mindat:1:2:1782:7), with 474 valid minerals (Fig. 5; Table 5). Collecting minerals there is encouraged by bringing in truckloads of ore for collectors to search for a fee.

Personal preference in mineral species studies and reports also leads to bias in the mineral occurrence records. Among the current (May 2024) approximately 400 000 localities on Mindat, most have a very small number of unique mineral species, while a few have a relatively large number of records. Figure 5 shows the global distribution of the top 500 localities with the most recorded mineral species. The numbers of reported mineral species span from 474 on the top (at Clara Mine, Baden-Württemberg, Germany, mindat:1:2:1782:7) to 49 at the bottom (at Gacun Mine, Sichuan, China, mindat:1:2:146933:5) among these 500 localities. Table 5 gives details of the top five in those 500 localities, including the Mindat identifiers and the links to access more information about them on the Mindat website.

IMPLICATIONS

Mindat has been running for more than two decades and has demonstrated its great potential for geoscience research and education. In the next decade, we will leverage Mindat's best practices on data collection and quality control and embrace the open science campaign to connect Mindat's open data to tools and platforms for data analytics. Our aim for Mindat is an active ecosystem of open data, code, samples, and science by and for the community.

We will sustain the “crowd-sourcing and expert-curating” mode of Mindat in database construction and quality control. As one of the early Internet projects in “citizen science,” Mindat had to create procedures for the management and approval of data contributed by the public. These approval rules have changed over the years, but they rely on the following general guidelines: (1) To add data to the database, a user must be registered and request approval to add data. This will include a description of what entries they want to contribute and some background on

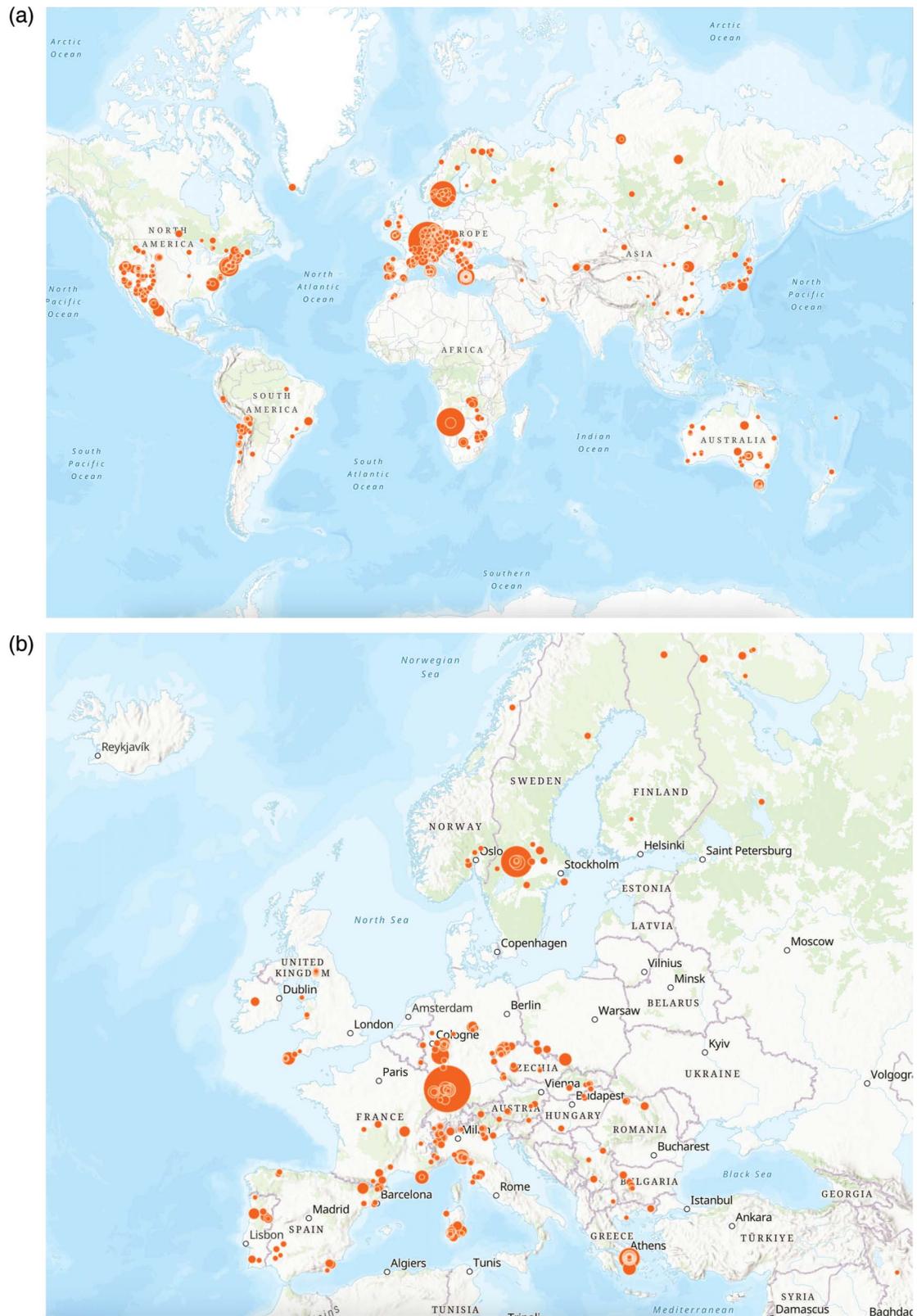


FIGURE 5. The top 500 Mindat localities with the most recorded mineral species. (a) An overview of the 500 localities across the world and (b) the localities in Europe. In both maps, each spot is a locality record in Mindat, and the size of the spot represents the number of mineral species found at that locality. Table 5 gives more details of the top five localities.

TABLE 5. Details of the top five Mindat localities with the most recorded mineral species

Locality name	Locality identifier	Access URL	Number of mineral species
Clara Mine, Oberwolfach, Ortenaukreis, Freiburg Region, Baden-Württemberg, Germany	mindat:1:2:1782:7	https://www.mindat.org/1:2:1782:7 or https://www.mindat.org/loc-1782.html	474
Tsumeb Mine (Ongopolo Mine), Tsumeb, Oshikoto Region, Namibia	mindat:1:2:2428:3	https://www.mindat.org/1:2:2428:3 or https://www.mindat.org/loc-2428.html	345
Långban Mine, Långban Ore District, Filipstad, Värmland County, Sweden	mindat:1:2:3167:0	https://www.mindat.org/1:2:3167:0 or https://www.mindat.org/loc-3167.html	315
Sterling Mine, Sterling Hill, Ogdensburg, Sussex County, New Jersey, U.S.A.	mindat:1:2:3948:5	https://www.mindat.org/1:2:3948:5 or https://www.mindat.org/loc-3948.html	241
Hilarion Mine, Kamariza Mines (Kamareza Mines), Agios Konstantinos (Kamariza), Lavrion Mining District, Lavreotiki, East Attica, Attica, Greece	mindat:1:2:1936:4	https://www.mindat.org/1:2:1936:4 or https://www.mindat.org/loc-1936.html	212

their interest and experience. Geomaterial entries must either contain a literature reference or details of how the identification of mineral species was carried out. (2) When new entries are submitted, they are put into an approval queue. Regional experts will review these submissions, and the data provided before they are added to the core data sets. In cases where there is little doubt (e.g., a report of microcline feldspar from an area where granitic pegmatites are known) then these can be approved rapidly. In comparison, submissions of occurrences of rare and unusual minerals or unexpected minerals within the region, require a greater level of verification and can frequently involve communication with the submitter to ensure the information is correct. (3) In addition, because of the public reach of the Mindat website within the scientific, commercial, and amateur communities, possible mistakes in existing data can be discovered, reported, investigated, and, if necessary, fixed. Many functions have been built in Mindat to support the procedures of data entry, review, and approval, as well as communications between data contributors, end users, and the management team. We wish any interested users to register as data contributors, submit new data records, and help curate the database.

Gradually, Mindat has been transitioning from a database website to an online community of data sharing, curation, distribution, and analytics. Along with the development of the Mindat open data API in the recent few years, we have created channels on the Slack platform to allow users to discuss directly with the Mindat data managers. The number of members in those channels has increased significantly since the summer of 2022. Various topics, from data standards, new data subjects, data cleansing, and API documentation to R/Python package development and data reuse in critical mineral exploration, have been mentioned and addressed. We have also incorporated the latest data analytics research from the geoscience community and built new functions on the Mindat website. For example, the works of mineral association analysis (Morrison et al. 2023) and paragenetic modes of minerals (Hazen and Morrison 2022) have now been applied to each locality on Mindat. Based on the existing mineral species list at a locality, predictions can be generated on new minerals to be discovered as well as the paragenetic modes of the deposit.

Mindat is also leading in comprehensive semantic descriptions of minerals and lithologies and facilitates the best practices of open data service among scientific communities. For the IMA mineral species list, we have received help from the IMA's Mineral Informatics Working Group on user needs, and we

have also referred to the RRUFF data portal to learn the best practices for providing information on mineral species. The R/Python packages developed by Mindat have functions to output the mineral species list with detailed attributes, which benefit not only individual users but also other databases that need such a standardized list. We have also collaborated with the OneGeochemistry community and the Deep-time Digital Earth (DDE) program of the International Union of Geological Sciences (IUGS) on the standards of rock classification and the associated synonyms by leveraging widely used standards developed by the BGS as well as the crowd-sourced information on Mindat.

With those continuous endeavors on data curation, function enrichment, and community building, our long-term vision for Mindat is a healthy ecosystem of data, software packages, scientists, and innovative ideas. Mindat has supported many impressive data-intensive geoscience studies in the past years, and we believe more ground-breaking discoveries are yet to come.

DATA AND CODE AVAILABILITY

The Mindat website ([mindat.org](https://www.mindat.org)) is free for anyone to use and a registered user can see more features and functions. The documentation of the Mindat open data API is available at: <https://api.mindat.org/schema/redoc/>. The tutorial on obtaining and using the API token is accessible at: https://www.mindat.org/a/how_to_get_my_mindat_api_key. The OpenMindat R and Python packages are open source under MIT License and Apache Software License, respectively. The documentation, source code, and demos of the OpenMindat R package are accessible at: <https://cran.r-project.org/web/packages/OpenMindat>. Similar materials of the OpenMindat Python package are accessible at: <https://pypi.org/project/openmindat>.

ACKNOWLEDGMENTS AND FUNDING

This work was supported by the U.S. National Science Foundation (Grant No. 2126315). We thank Robert M. Hazen for many constructive discussions on the recent data-driven discoveries in mineralogy. We appreciate Marthe Klöcking and another anonymous reviewer for their detailed comments on an earlier version of this paper, which greatly increased the quality of the writing during the revision.

REFERENCES CITED

- 4D Initiative (2018) White Paper of The 4D Initiative: Deep-time Data Driven Discovery (Online). https://4d.carnegiescience.edu/sites/default/files/4D_materials/4D_WhitePaper.pdf (accessed on: April 25, 2024). Carnegie Science.

- Beaulieu, S.E. and Szafranski, K.M. (2020) InterRidge Global Database of Active Submarine Hydrothermal Vent Fields (ver. 3.4.) (Online). Available: <https://doi.org/10.1594/PANGAEA.917894>. PANGAEA.
- Bradley, D.C. (2015) Mineral evolution and Earth history. *American Mineralogist*, 100, 4–5, <https://doi.org/10.2138/am-2015-5101>.
- British Geological Survey (BGS) (2020) The BGS Rock Classification Scheme (Online). Available at: <https://www.bgs.ac.uk/technologies/bgs-rock-classification-scheme/>. Keyworth, Nottingham.
- Clark, A.M. (1993) *Hey's Mineral Index: Mineral Species, Varieties and Synonyms*, 3rd ed., 848 p. Chapman and Hall.
- Dallaire-Fortier, C. (2024) A comprehensive historical and geolocalized database of mining activities in Canada. *Scientific Data*, 11, 307, <https://doi.org/10.1038/s41597-024-03116-3>.
- Dana, J.D. (1868) *A System of Mineralogy*, 5th ed., 894 p. Wiley.
- Dunn, P.J. (1990) Andrewsrite and laubmannite formally discredited. *American Mineralogist*, 75, 1197–1199.
- Fettes, D. and Desmons, J., Eds. (2007) *Metamorphic Rocks: A Classification and Glossary of Terms: Recommendations of the International Union of Geological Sciences Subcommission on the Systematics of Metamorphic Rocks*, 256 p. Cambridge University Press.
- Gebhard, G. (1999) *Tsumeb II. A Unique Mineral Locality*, 328 p. GG Publishing.
- Grew, E.S., Hystad, G., Toapanta, M.P., Eleish, A., Ostroverkhova, A., Golden, J., and Hazen, R.M. (2019) Lithium mineral evolution and ecology: Comparison with boron and beryllium. *European Journal of Mineralogy*, 31, 755–774, <https://doi.org/10.1127/ejm/2019/0031-2862>.
- Hawthorne, F.C., Oberti, R., Harlow, G.E., Maresch, W.V., Martin, R.F., Schumacher, J.C., and Welch, M.D. (2012) Nomenclature of the amphibole supergroup. *American Mineralogist*, 97, 2031–2048, <https://doi.org/10.2138/am.2012.4276>.
- Hazen, R.M. (2019) An evolutionary system of mineralogy: Proposal for a classification of planetary materials based on natural kind clustering. *American Mineralogist*, 104, 810–816.
- Hazen, R.M. and Morrison, S.M. (2022) On the paragenetic modes of minerals: A mineral evolution perspective. *American Mineralogist*, 107, 1262–1287, <https://doi.org/10.2138/am-2022-8099>.
- Hazen, R.M., Papineau, D., Bleeker, W., Downs, R.T., Ferry, J.M., McCoy, T.J., Sverjensky, D.A., and Yang, H. (2008) Mineral evolution. *American Mineralogist*, 93, 1693–1720, <https://doi.org/10.2138/am.2008.2955>.
- Hazen, R.M., Grew, E.S., Downs, R.T., Golden, J., and Hystad, G. (2015) Mineral ecology: Chance and necessity in the mineral diversity of terrestrial planets. *Canadian Mineralogist*, 53, 295–324, <https://doi.org/10.3749/canmin.1400086>.
- Historical Canadian Mines (2024) *The Historical Canadian Mines Data Hub and Visualization Centre* (Online). Available: <https://www.arcgis.com/apps/dashboards/780a4bc0aa524cc38e10a4699bc3511e> (accessed on April 25, 2024). Historical Canadian Mines.
- Hystad, G., Downs, R.T., Grew, E.S., and Hazen, R.M. (2015) Statistical analysis of mineral diversity and distribution: Earth's mineralogy is unique. *Earth and Planetary Science Letters*, 426, 154–157, <https://doi.org/10.1016/j.epsl.2015.06.028>.
- Hystad, G., Eleish, A., Hazen, R.M., Morrison, S.M., and Downs, R.T. (2019) Bayesian estimation of Earth's undiscovered mineralogical diversity using noninformative priors. *Mathematical Geosciences*, 51, 401–417, <https://doi.org/10.1007/s11004-019-09795-8>.
- Lafuente, B., Downs, R.T., Yang, H., and Stone, N. (2015) The power of databases: The RRUFF project. In T. Armbruster and R.M. Danisi, Eds., *Highlights in Mineralogical Crystallography*, p. 1–30. De Gruyter.
- Landicho, C.J.B. (2021) Changes, challenges, and opportunities in teaching senior high school Earth science amidst the COVID-19 pandemic. *Journal of Learning and Teaching in Digital Age*, 6, 55–57.
- Landucci, C. and Bosi, F. (2023) International Mineralogical Association - Commission on New Minerals, Nomenclature and Classification. IMA-CNMC (Online). <http://cnmc.units.it/> (accessed September 08, 2023). International Mineralogical Association.
- Le Maitre, R., Ed. (2005) *Igneous Rocks: A Classification and Glossary of Terms: Recommendations of the International Union of Geological Sciences Subcommission on the Systematics of Igneous Rocks*, 2nd ed., 256 p. Cambridge University Press.
- Leake, B.E., Woolley, A.R., Arps, C.E.S., Birch, W.D., Gilbert, M.C., Grice, J.D., Hawthorne, F.C., Kato, A., Kisch, H.J., Krivovichev, V.G., and others. (1997) Nomenclature of amphiboles; Report of the Subcommittee on Amphiboles of the International Mineralogical Association, Commission on New Minerals and Mineral Names. *American Mineralogist*, 82, 1019–1037.
- Lehnert, K., Su, Y., Langmuir, C.H., Sarbas, B., and Nohl, U. (2000) A global geochemical database structure for rocks. *Geochemistry, Geophysics, Geosystems*, 1, 1012, <https://doi.org/10.1029/1999GC000026>.
- Lehnert, K.A., Walker, D., and Sarbas, B. (2007) EarthChem: A geochemistry data network. *Geochimica et Cosmochimica Acta*, 71, A559, <https://doi.org/10.1016/j.gca.2007.06.020>.
- Ma, X., Ralph, J., Zhang, J., Que, X., Prabhu, A., Morrison, S.M., Hazen, R.M., Wyborn, L., and Lehnert, K. (2024) OpenMindat: Open and FAIR mineralogy data from the Mindat database. *Geoscience Data Journal*, 11, 94–104, <https://doi.org/10.1002/gdj.3204>.
- Marsh, E. and Anderson, E. (2015) Database of mineral deposits in the Islamic Republic of Mauritania (phase V, deliverables 90 and 91). In C.D. Taylor, Ed., *Second projet de renforcement institutionnel du secteur minier de la République Islamique de Mauritanie (PRISM-II)*. U.S. Geological Survey Open-File Report 2013-1280-S, <https://doi.org/10.3133/ofr20131280S>.
- McGill, M. (2020) Young mineral collectors and the future of the mineral collecting hobby. *Rocks and Minerals*, 95, 537–547, <https://doi.org/10.1080/00357529.2020.1791628>.
- Meteoritical Bulletin Database (2023) *Meteoritical Bulletin Database* (online). Available: <https://www.lpi.usra.edu/meteor/metbullclass.php?sea=CM2> (accessed on July 13, 2023). The Meteorological Society.
- Mills, S.J., Hatert, F., Nickel, E.H., and Ferraris, G. (2009) The standardisation of mineral group hierarchies: Application to recent nomenclature proposals. *European Journal of Mineralogy*, 21, 1073–1080, <https://doi.org/10.1127/0935-1221/2009/0021-1994>.
- Moore, E.K., Hao, J., Prabhu, A., Zhong, H., Jelen, B.I., Meyer, M., Hazen, R.M., and Falkowski, P.G. (2018) Geological and chemical factors that impacted the biological utilization of cobalt in the Archean eon. *Journal of Geophysical Research. Biogeosciences*, 123, 743–759, <https://doi.org/10.1002/2017JG004067>.
- Moore, E.K., Golden, J.J., Morrison, S.M., Hao, J., and Spielman, S.J. (2022) The expanding network of mineral chemistry throughout Earth history reveals global shifts in crustal chemistry during the Proterozoic. *Scientific Reports*, 12, 4956, <https://doi.org/10.1038/s41598-022-08650-x>.
- Morrison, S.M., Prabhu, A., Eleish, A., Hazen, R.M., Golden, J.J., Downs, R.T., Perry, S., Burns, P.C., Ralph, J., and Fox, P. (2023) Predicting new mineral occurrences and planetary analog environments via mineral association analysis. *PNAS Nexus*, 2, pgad110, <https://doi.org/10.1093/pnasnexus/pgad110>.
- National Research Council (2008) *Origin and Evolution of Earth: Research Questions for a Changing Planet*, 150 p. The National Academies Press.
- Ottens, B. (2005) *Mining in China: A 3000-year tradition*. The Mineralogical Record, 36, 4–11.
- Peters, S.E., Husson, J.M., and 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>.
- Plant, J.A. and Jones, D.G. (2013) *Metallogenic models and exploration criteria for buried carbonate-hosted ore deposits—A multidisciplinary study in eastern England*, 187 p. Springer.
- Que, X. and Ma, X. (2023) OpenMindat R Package (v.1.0) (Online). Available: <https://github.com/quexiang/OpenMindat> (accessed on April 25, 2024).
- Ralph, J., Martynov, P., Ma, X., Von Bargen, D., Li, W., Huang, J., Golden, J., Profeta, L., Prabhu, A., Morrison, S., and others. (2024) Identifier Service in the Mindat Database: Persistent and Structured Access to Massive Records of Minerals and Other Natural Materials. *Data Intelligence*, in press, <https://doi.org/10.3724/2096-7004.di.2024.0021>.
- Rong, S., Wu, J., Liu, J., Li, Q., Ren, C., and Cao, X. (2023) Environmental magnetic characteristics and heavy metal pollution assessment of sediments in the Le'an River, China. *Minerals*, 13, 145, <https://doi.org/10.3390/min13020145>.
- Schweitzer, P.N. (2019) Record quality tables for the Mineral Resources Data System [Data set]. U.S. Geological Survey, <https://doi.org/10.5066/P9DYLWMP>.
- Strunz, H. and Nickel, E.H. (2001) *Strunz Mineralogical Tables*, 9th ed., 870 p. E. Schweizerbart'sche Verlagsbuchhandlung.
- Wang, C., Hazen, R.M., Cheng, Q., Stephenson, M.H., Zhou, C., Fox, P., Shen, S.Z., Oberhänsli, R., Hou, Z., Ma, X., and others. (2021) The Deep-Time Digital Earth program: Data-driven discovery in geosciences. *National Science Review*, 8, nwab027, <https://doi.org/10.1093/nsr/nwab027>.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., and others. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018, <https://doi.org/10.1038/sdata.2016.18>.
- Zhang, J., Clairmont, C., and Ma, X. (2024) OpenMindat Python Package (Online). Available: <https://github.com/ChuBL/OpenMindat> (accessed on April 25, 2024).

MANUSCRIPT RECEIVED JUNE 5, 2024

MANUSCRIPT ACCEPTED OCTOBER 22, 2024

ACCEPTED MANUSCRIPT ONLINE OCTOBER 29, 2024

MANUSCRIPT HANDLED BY DANIEL HUMMER