Applied Computing and Geosciences 20 (2023) 100140

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences

The British Geological Survey Rock Classification Scheme, its representation as linked data, and a comparison with some other lithology vocabularies

Tim McCormick *, Rachel E. Heaven

British Geological Survey, Nicker Hill, Keyworth, Nottingham, NG12 5GG, United Kingdom

ARTICLE INFO

Keywords:
Ontology
Vocabulary
Lithology
Linked data

ABSTRACT

Controlled vocabularies are critical to constructing FAIR (findable, accessible, interoperable, re-useable) data. One of the most widely required, yet complex, vocabularies in earth science is for rock and sediment type, or 'lithology'. Since 1999 the British Geological Survey has used its own Rock Classification Scheme in many of its workflows and products including the national digital geological map. This scheme pre-dates others that have been published, and is deeply embedded in BGS' processes. By publishing this classification scheme now as a Simple Knowledge Organisation System (SKOS) machine-readable informal ontology, we make it available for ourselves and third parties to use in modern semantic applications, and we open the future possibility of using the tools SKOS provides to align our scheme with other published schemes. These include the IUGS-CGI Simple Lithology Scheme, the European Commission INSPIRE Lithology Code List, the Queensland Geological Survey Lithotype Scheme, the USGS Lithologic Classification of Geologic Map Units, and Mindat.org. The BGS lithology classification was initially based on four narrative reports that can be downloaded from the BGS website, although it has been added to subsequently. The classification is almost entirely mono-hierarchical in nature and includes 3454 currently valid concepts in a classification 11 levels deep. It includes igneous rocks and sediments, metamorphic rocks, sediments and sedimentary rocks, and superficial deposits including anthropogenic deposits. The SKOS informal ontology built on it is stored in a triplestore and the triples are updated nightly by extracting from a relational database where the ontology is maintained. Bulk downloads and version history are available on github. The RCS concepts themselves are used in other BGS linked data, namely the Lexicon of Named Rock Units and the linked data representation of the 1:625 000 scale geological map of the UK. Comparing the RCS with the other published lithology schemes, all are broadly similar but show characteristics that reveal the interests and requirements of the groups that developed them, in terms of their level of detail both overall and in constituent parts. It should be possible to align the RCS with the other classifications, and future work will focus on automated mechanisms to do this, and possibly on constructing a formal ontology for the RCS.

1. Introduction

In this paper, we describe the development of a machine-readable informal ontology for rock and sediment type (i.e., lithology), based on an existing classification scheme that has been in use within the British Geological Survey for the past two decades. We compare our classification scheme with some other available controlled vocabularies for lithology, highlighting similarities and differences.

Controlled vocabularies are a critical component in constructing 'FAIR' (findable, accessible, interoperable, re-useable) data resources.[1] A controlled vocabulary provides a discoverable, well-defined, authoritative list of terms for a field of thought, activity, or interest. Each term in the vocabulary is furnished with a persistent identifier, a definition, and often translations into different languages. The purpose is to provide a consistent, re-useable set of terms that can be used to attribute entities in data resources. There are several advantages to exploiting controlled vocabularies. First, if a well-known and well-defined set of terms are consistently used as attributes in a data set, then the data are much more readily searchable; the data become findable and accessible. Second, if disparate data sets use the same controlled vocabulary, then they 'speak

* Corresponding author.
E-mail address: tmcm@bgs.ac.uk (T. McCormick).

[1] https://www.go-fair.org/fair-principles/.

https://doi.org/10.1016/j.acags.2023.100140
Received 29 June 2023; Received in revised form 25 September 2023; Accepted 12 October 2023
Available online 17 October 2023
2590-1974/© 2023 British Geological Survey (C) UKRI 2023. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the same language' for that attribute and can be combined into aggregated systems; they become interoperable. Third, data sets that use well-defined controlled vocabularies can be much more easily understood and used by third parties; they become re-useable. Controlled vocabularies help to make data FAIR.

Controlled vocabularies can be of varying complexity from simple term lists, to taxonomies which may include hierarchical relationships, to thesauri which may also include synonyms and non-hierarchical relationships. These are all used to organise knowledge and can be expressed in machine readable form using W3C Simple Knowledge Organisation System (SKOS).[2] Ontologies can be a much richer and more complex structure defining not just classes and relationships but also properties and class instances, and they are robust enough that they can be used for reasoning; they are a model to represent the knowledge. Ontologies can be expressed in machine readable form using W3C Web Ontology Language (OWL),[3] in which case they are known as formal ontologies. For example, a SKOS vocabulary could contain the list of terms that constrain values of a database attribute; an ontology can represent the entire data model of the database and its contents. The SKOS and OWL standards are two of the technologies on which the semantic web is built.[4]

Much work has been done in recent years to develop controlled vocabularies for the earth sciences. Organisations like the International Union of Geoscientists (IUGS),[5] through its Commission for the Management and Application of Geoscience Information (CGI)[6] and its International Commission on Stratigraphy (ICS),[7] as well as the European Commission through its INSPIRE Directive,[8] and national and state geological surveys, have been instrumental in developing controlled vocabularies covering many subject areas. These include the subdivisions of geological time; mineable commodities; shape and form of mineral deposits; methods of mineral exploration, mining, processing, and evaluation; drilling and sampling methods; and remote sensing platforms and methods. Arguably, one of the most widely required, yet complex controlled vocabularies needed in the earth sciences, is lithology.

2. The requirement

In common with other geoscience organisations, BGS uses controlled vocabularies to standardise its data holdings for the reasons explained above. Since 1999 we have used a classification and naming scheme for rocks and sediments, developed in-house, to standardise our concepts and terminology in this area. It is called the Rock Classification Scheme (BGS RCS).[9] It is a hierarchical classification of concepts, so that the concept 'Granite', for example, belongs within the progressively broader concepts 'Granitic-rock', 'Coarse-grained normal crystalline rock', 'Normal crystalline igneous rock', 'Crystalline igneous rock', 'Igneous rock and sediment', 'Rock and sediment'. Each concept in the scheme has a name and a definition. A full description of the scheme is given in Section 4.

The RCS is a core component of our digital geological map of Great Britain (Smith, 2013) which in turn underpins almost all BGS activities and outputs. Additionally, the RCS is used in coding borehole logs, and thence in constructing 3D models of the subsurface (e.g., Wood and Kessler, 2021). It is used in the BGS System for Integrated Geoscience Mapping software (SIGMA) (Bow, 2015) which is now used for geological mapping in the UK and other countries (e.g., Thomas, 2010; Hughes, 2013; Jones and McCormick, 2020). McCormick (2021) highlighted the degree to which the RCS is embedded in BGS workflows and products.

Since its development the RCS, an informal ontology, has been available in full as printed reports (See Section 4). The knowledge organisation aspects (rock classes and hierarchies without associated properties) were managed in a relational database, as a searchable web application on the BGS website (Figs. 1–3) and as a flat term list in a non-standards based web API. However, it has not previously been published in a standards based machine-readable form. We now publish the RCS as a concept scheme using SKOS. This will enable us to use the RCS in modern semantic web applications and visualisations. It will also enable third parties to download and use the RCS in their own applications, and for client applications to interact with the machine-readable data. Ultimately, we intend to align our RCS with other lithology vocabularies, and use the SKOS representation to express those alignments, although we have not done this yet. Possible approaches are discussed below (see Section 7).

Table 1 lists some basic properties of the BGS RCS and compares them with some other lithology classifications that may be found on the Web. Note that the original publication of the RCS pre-dates the other schemes, although not its publication as a SKOS ontology which is happening now. This explains why, when it was decided that BGS needed a rock classification scheme in the late 1990s, it was necessary to develop one in-house. There was no other scheme available. It is known that the developers of some of the other schemes referred to the RCS when constructing their classifications.

As described above, the RCS is deeply embedded in BGS data sets and workflows. This means it would be a mammoth, and extremely difficult, task to replace the RCS in our activities with one of the other, newer schemes listed in Table 1 now. This is a further reason why we choose to publish the RCS as a SKOS ontology, so that ultimately, we can align it with other existing schemes.

3. Some online lithology vocabularies published by other organisations

Here we briefly describe some of the other controlled vocabularies for lithology that may be found on the Web, which were developed after the original publication of the RCS in 1999. We do this because it may be useful to readers interested in lithology vocabularies, and because it provides some context to our evaluation of the RCS SKOS ontology later in this paper (Section 6). The selection is by no means exhaustive.

3.1. CGI Simple Lithology Scheme

The CGI Simple Lithology Scheme[10], first published in 2009, is one of many controlled vocabularies developed and published by the IUGS-CGI. Their repository[11] currently includes 35 vocabularies developed for use with the GeoSciML data transfer standard.[12] These are mainly focussed on the types of observations that might be made when mapping regional geology, and include vocabularies for Alteration Type, Contact Type, Deformation Style, and Fault Type among others, as well as the Simple Lithology Scheme. In addition, there are currently another 19 vocabularies developed for use with EarthResourceML[13] which focus on mining and mineral resources, and include vocabularies for Commodity, Earth Resource Form and Shape, Environmental Impact, Exploration Activity Type, Mine Status, Mineral Occurrence Type, and Mining

2 https://www.w3.org/2004/02/skos/.
3 https://www.w3.org/OWL/.
4 https://www.w3.org/2001/sw/wiki/Main_Page.
5 https://www.iugs.org/.
6 https://cgi-iugs.org/.
7 https://stratigraphy.org/.
8 https://inspire.ec.europa.eu/.
9 https://www.bgs.ac.uk/technologies/bgs-rock-classification-scheme/.
10 http://resource.geosciml.org/classifier/cgi/lithology.
11 https://cgi.vocabs.ga.gov.au/vocab/.
12 http://geosciml.org/.
13 http://earthresourceml.org/.

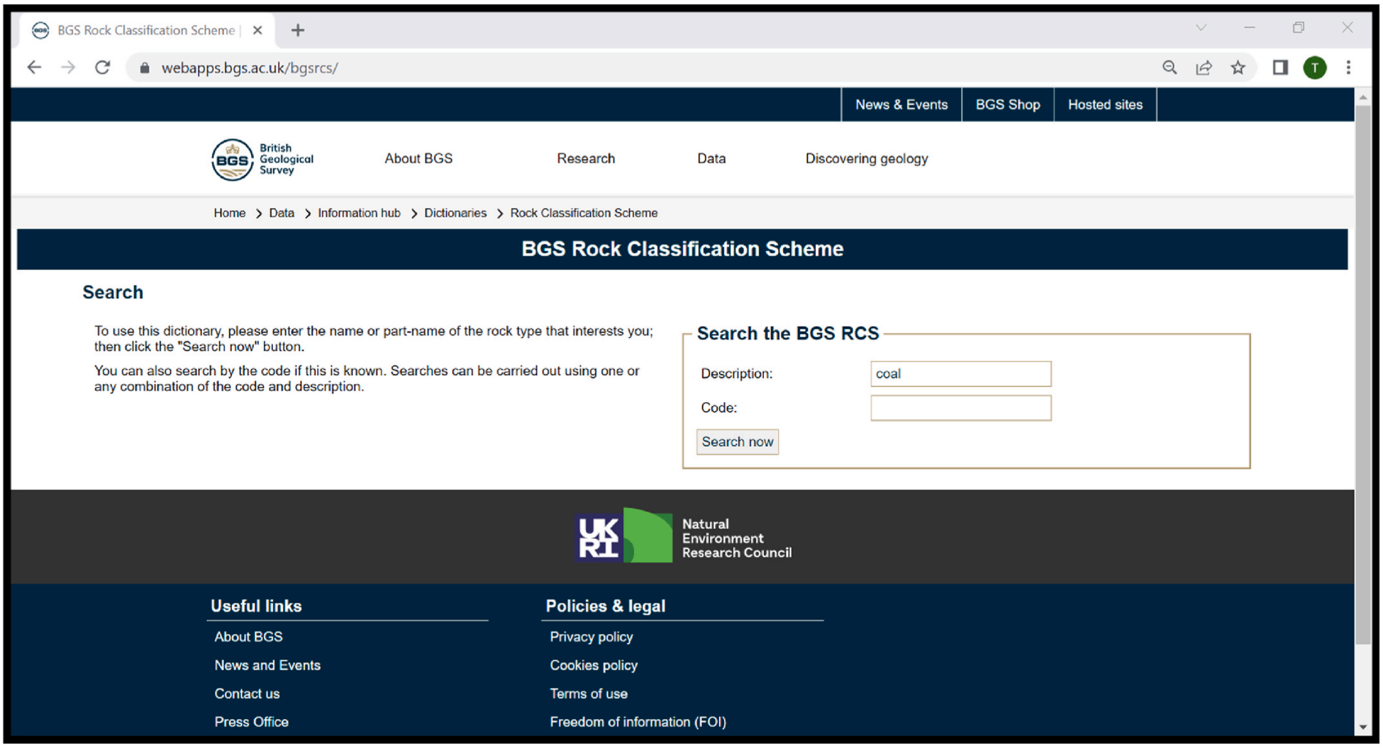

Fig. 1. Searching the RCS using the web app.

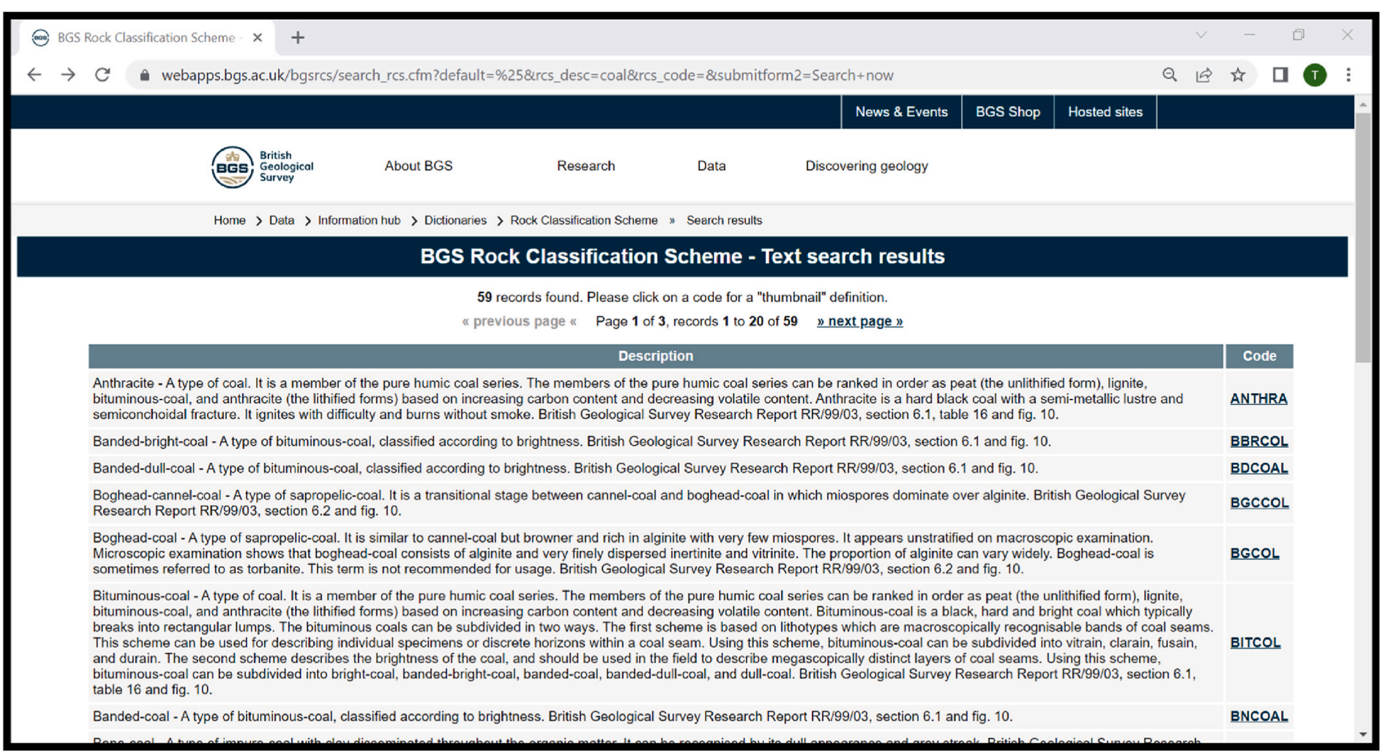

Fig. 2. Search results displayed by the RCS web app.

Activity among others. The vocabularies are presented using VocPrez[14] and are maintained by special interest groups within CGI.

The Simple Lithology vocabulary is a classification which may be browsed on screen or exported as SKOS in JSON, N-Triples, RDF/XML, and Turtle formats. Each concept is identified by a URI, and has Preferred Label, Definition, and Narrower and Broader relationships. The classification includes igneous rocks and sediments, metamorphic rocks, surficial deposits, and sedimentary rocks and sediments. There is no real inclusion of artificial deposits other than the concept

[14] https://github.com/RDFLib/VocPrez.

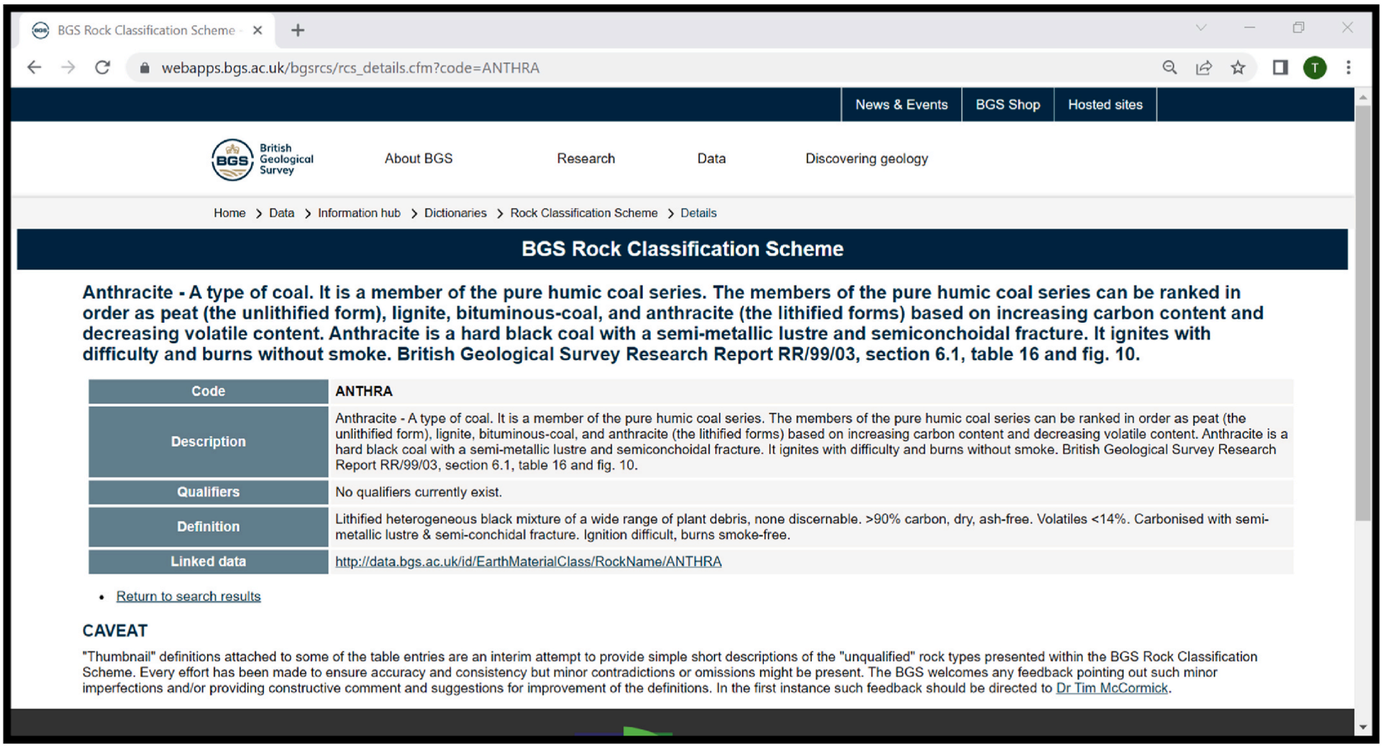

Fig. 3. An RCS entry displayed by the web app. Note the hyperlink pointing to the linked data representation of the same RCS entry.

Table 1
Summary of lithology classifications discussed. *The BGS RCS is almost entirely mono-hierarchical with the exception of 'composite-concepts' described in the text.

Scheme	BGS Rock Classification Scheme	CGI Simple Lithology Scheme	INSPIRE Lithology Code List	QGS Lithotype Scheme	USGS Lithologic Classification of Geologic Map Units
First published	1999	2009	2012	2020	2002
Number of concepts	3454	265	276	2968	206
Depth of classification	11	6	7	11	5
Nature of classification	Mono-hierarchical*	Poly-hierarchical	Mono-hierarchical	Poly-hierarchical	Mono-hierarchical
Concepts have URIs	Yes	Yes	Yes	Yes	No

'anthropogenic material'. There are at present 265 concepts organised into a hierarchical classification up to 6 levels deep. The top-level concept is 'compound material'.[15]

The Simple Lithology Scheme is poly-hierarchical, meaning that a concept may belong to more than one immediate broader concept. For example, their concept 'granitoid' belongs to two immediate broader concepts: 'acid igneous rock' and 'phaneritic igneous rock'. The poly-hierarchical nature of the classification means that when it is fully expanded, entire sections are repeated, so that, for example, 'granite' and associated concepts appear in 5 different places in the classification; 'limestone' and associated concepts appear in 7 places.

3.2. *INSPIRE Lithology Code List*

The INSPIRE Lithology Code List,[16] published in 2012, is one of 34 controlled vocabularies currently published within the Geology Theme of the INSPIRE Code List Register[8], developed for the European Commission INSPIRE Directive. Other vocabularies within the INSPIRE Geology Theme include Aquifer Type, Borehole Purpose, Fault Type, Geochronologic Era, and Water Salinity. The vocabularies are presented using a bespoke web app.

The Lithology Code List is a classification which may be browsed on screen or downloaded in RDF/XML, JSON, CSV, ATOM, and ROR formats. Each concept is identified by a URI, and has Label, Definition, and Parent. Individual concepts may also be downloaded. The top-level concept is 'compound material'.[17]

In fact, the INSPIRE Lithology Code List is a copy of the CGI Simple Lithology Scheme with a small number of additional concepts, and the CGI working group have attributed every concept in their scheme with an exact match to the corresponding concept in the INSPIRE scheme. The INSPIRE scheme has two modifications compared with the CGI scheme. The first is that the identifiers have been modified to comply with the general rules applied for INSPIRE coding. The second is that the hierarchical classification of the INSPIRE version has been simplified to be mono-hierarchical, i.e., each concept is included in only one broader concept. This means each concept appears only once in the fully expanded classification. This accounts for the differing depths of the CGI and INSPIRE classifications.

15 http://resource.geosciml.org/classifier/cgi/lithology/compound_material.
16 https://inspire.ec.europa.eu/codelist/LithologyValue.
17 https://inspire.ec.europa.eu/codelist/LithologyValue/compoundMaterial.

3.3. Queensland Geological Survey Lithotype Scheme

The Queensland Geological Survey (QGS) Lithotype Scheme,[18] first published as linked data in 2020, is one of some 91 controlled vocabularies currently being published by the QGS, covering both geoscience and administration.[19] They are presented using VocPrez.

The QGS Lithotype Scheme is a classification which may be browsed on screen or downloaded as SKOS or 'Drop-Down List' in JSON, N-Triples, RDF/XML, and Turtle formats. Each concept is identified by a URI, and has Preferred Label, Definition, Broader Terms, and Alternative Label. There is also a Notation, or alphabetic code (note that these do not match the identifiers used by the BGS RCS, see Section 4). QGS has attributed many concepts with exact match to the corresponding concept in the CGI Simple Lithology Scheme. Each concept may be downloaded as SKOS in the formats mentioned above.

The more than 2900 concepts are organised into a poly-hierarchical classification. The top-level concepts are 'non rock observation'[20] (including miscellaneous observations that could be made at sites including contamination, fill, mine dump, mine workings, etc.) and 'rock'.[21] The QGS Lithotype Scheme includes concepts within 'meteorite and other extra-terrestrial rock and sediment', and 'ore', which do not appear in the other schemes described here, including the RCS.

The large number of concepts reflects the fact that this scheme goes into extraordinary detail. There are many 'term with qualifier' concepts that describe varieties of a rock type (for example biotite granite, orbicular biotite granite, Shap granite, brecciated granite, graphic granite). The meteorites section of the scheme appears to include more than 350 types of chondrite meteorite, some of which, owing to the poly-hierarchical nature of the classification, appear four times in the fully expanded hierarchy.

3.4. USGS Lithologic Classification of Geologic Map Units

The USGS Lithologic Classification of Geologic Map Units[22] is somewhat different to the other classifications discussed here. As well as rock and sediment types, it includes some generic litho-morphogenetic concepts like flood plain, levee, alluvial fan, talus, and till. It is stored as a relational database and presented using a bespoke web app. The vocabulary may be browsed on screen, and individual concepts may be downloaded as JSON or XML, or the full thesaurus as an SQLite database.[23]

Each concept has an Identifier, Label, Definition, and a list of Subtopics (narrower concepts). But unlike the other vocabularies discussed above, the identifier is not used to create a persistent URI. Rather, each concept is accessed via a URL whose structure exposes the technology and functionality of the web app; if the web app is re-coded using a different technology, then the URLs used for individual concept pages are likely to be different.

There are currently 206 concepts organised into a mono-hierarchical classification, 5 levels deep. The top-level concepts are unconsolidated material, sedimentary rock, volcanic rock (fine-grained igneous, pyroclastic and volcaniclastic rocks of other schemes), plutonic rock (coarse-grained crystalline igneous rocks of other schemes), and metamorphic rock. While it does include some litho-morphogenetic concepts, it does not include anthropogenic materials.

3.5. Mindat.org

Mindat[24] is not a controlled vocabulary in the same way as the others discussed here. Rather it is a comprehensive online resource of descriptions and illustrations of minerals, rocks, and their occurrences. But although it is not a vocabulary as such, it is a rich and widely used source of information about rocks and minerals. When an Internet search is carried out on the name of a rock or sediment type, Mindat is the only one of the resources discussed here that routinely appears on the first page of the result list, along with Wikipedia and various online dictionaries. While the primary focus of Mindat is minerals, pages for rocks and sediments have been added recently.

Mindat is presented as a large searchable set of static webpages. Each page provides a comprehensive introduction to the rock, sediment, or mineral, including photographs, classification, mineralogy, age and geographic distribution, and references. The pages are written by volunteers and checked and edited by an expert management team.

In the context of this discussion, the power of Mindat is to provide additional explanation, illustration, and context over and above what can be provided by the 'definition' attribute of a controlled vocabulary. The richness of a concept definition is greatly enhanced by adding a 'see also' link to the relevant Mindat page.

4. The BGS rock classification scheme

Here we describe the characteristics of the BGS RCS in some detail before explaining how the SKOS informal ontology was constructed in the next section.

The Rock Classification Scheme was initially developed by BGS in the late 1990s as a taxonomy of classes of rock types. The main driver at the time was the need for a consistent and comprehensive scheme for naming and classifying rocks and other deposits, to enable a digital geological map of the UK to be developed. At the time, the only published English-language taxonomy of the main rock types in international use was that published by the IUGS Subcommission on the Systematics of Igneous Rocks (Le Maitre et al., 1989). No overarching schemes existed for metamorphic or sedimentary rocks, and no other geological survey had a scheme like the RCS.

The RCS was first published in 1999 as four BGS research reports: Vol. 1 Classification of igneous rocks (Gillespie and Styles, 1999); Vol. 2 Classification of metamorphic rocks (Robertson, 1999); Vol. 3 Classification of sediments and sedimentary rocks (Hallsworth and Knox, 1999); and Vol. 4 Classification of artificial (man-made) ground and natural superficial deposits (McMillan and Powell, 1999) (Table 2). These reports provide the labels, definitions, and hierarchical classification of rocks and sediments used by BGS, including in places the property value ranges on which the rock classes are defined. They are available for free download from the BGS website.[25] During 2022, the four RCS reports were downloaded over 1600 times (the igneous report is the most popular by some distance). At the same time, the classes, descriptions, and hierarchical structure of the RCS was implemented in a relational database management system (RDBMS). In the RDBMS each

Table 2
The fundamental top-level subdivision of the Rock Classification Scheme.

Level 0	Level 1	RCS Volume
Rock and sediment	Igneous rock and sediment	RCS Vol. 1
	Metamorphic rock	RCS Vol. 2
	Sediment and sedimentary rock	RCS Vol. 3
	Superficial deposit (natural and/or artificial)	RCS Vol. 4

[18] http://linked.data.gov.au/def/lithotype.
[19] https://vocabs.gsq.digital/vocabulary/.
[20] http://linked.data.gov.au/def/lithotype/non-rock-observation.
[21] http://linked.data.gov.au/def/lithotype/rock.
[22] https://apps.usgs.gov/thesaurus/thesaurus-full.php?thcode=4.
[23] https://www.sqlite.org/index.html.
[24] https://www.mindat.org/.
[25] https://www.bgs.ac.uk/technologies/bgs-rock-classification-scheme/.

rock and sediment type concept was given a unique alphanumeric persistent identifier (these are referred to as 'RCS codes' within BGS). The classification and identifiers have formed a core component of BGS workflows ever since, as described in Section 2.

In the RCS, rocks, sediments, and other types of deposit are named and classified on descriptive attributes which can be seen in hand specimen or thin section. The classification is (mostly) non-genetic, and it should not be necessary to see field relationships, or carry out geochemical analyses, to use the scheme (although see the reference to TAS Total Alkali versus Silica classification, Section 4.1). The classification is hierarchical and depending on factors including the nature of the study being undertaken, or the audience, the geologist may choose to use broader concepts near the root of the classification or narrower concepts near the leaves of the classification.

In the original RCS reports, rock and sediment names and concepts are attributed as either 'root' or 'approved'. Unfortunately, the usage of the term 'root' in the reports does not match modern tree-based terminology (the so-called 'root' concepts are on the leaves of the classification!). In Gillespie and Styles' (1999) description of how the scheme is organised, their use of 'root' and 'approved' has been replaced here with 'leaf node' and 'internal node' respectively:

> "In the BGS rock classification scheme, rock nomenclature is based on the principle that each distinct rock type has a unique [leaf node] name, which is assigned only when the geologist has all the modal and chemical or fragment size and origin information that is needed to classify it fully. Thus, in most cases the [leaf node] name represents the 'endpoint' in classification.
>
> [Internal node] names are also unique, and are assigned to a rock where there is insufficient information to classify it fully."

Among igneous rocks, examples of internal node concepts include 'Normal crystalline igneous rock', 'Coarse-grained normal crystalline rock', 'Granitic-rock' and 'Syenitic-rock'. Examples of leaf node concepts include 'Granite', 'Monzogranite', 'Syenogranite', and 'Syenite'.

The classification hierarchy in the RCS is almost entirely mono-hierarchical, so that each concept belongs to only one immediate broader concept and thus appears only once in the fully expanded classification. Exceptions include 'Diatomaceous-ooze', 'Radiolarian-ooze', and 'Sponge-spicular-ooze', which belong to both the broader concepts 'Non-clastic siliceous sediment' and 'Biological sediment', and the 'composite concepts' mentioned below. Compare poly-hierarchical classification schemes like the CGI Simple Lithology Scheme described in Section 3.1.

Although the four BGS research reports on which the classification is based have not been revised since they were published in 1999, the database implementation of the RCS used internally within BGS has had numerous ad hoc additions made to it over the years as required by the various workflows, products, and services that it supports. One example of such additions is so-called 'term plus qualifier' concepts, for example '*Biotite* granite', '*Sandy* limestone', and '*Ferruginous* sandstone'. These were added to the database as and when they were required for geological map production or borehole logging or some other purpose. Another example is so-called 'composite concepts' representing mixtures of rocks, for example 'Granite and diorite', 'Sandstone, siltstone and limestone', and 'Limestone with chert'. These composites were added to the database as a practical expedient to allow attribution of individual polygons in the digital geological map with multiple lithologies, because the digital map was developed using GIS software that only allowed each polygon to be attributed with one lithology code. These composite concepts, unlike most concepts in the RCS, belong to more than one broader concept, one for each component. Unlike the rest of the RCS, these parent terms are not super-classes, they are more generic constituent parts of the specialised complex term. It was also found necessary to add non-lithology concepts like 'Water', 'Void', and 'No core retrieval' to the vocabulary for use in describing borehole logs and cores. These concepts clearly do not describe a rock or sediment type and are referred to informally as 'dummy concepts'. They are another necessary addition to the vocabulary for practical use.

The entire scheme currently has 3454 valid terms organised into a classification 11 levels deep. When composite terms and terms with qualifiers are filtered out, there are 826 terms in a classification 9 levels deep.

A brief guide to the classification may be useful to geoscientists wishing to use it or compare it with other available classifications.

4.1. Igneous rocks and igneous sediments

The classification of igneous rocks and igneous sediments in the RCS is based on that established by the IUGS Subcommission on the Systematics of Igneous Rocks, with some refinements and modifications which the original authors felt would result in a more logical, consistent, and clearly defined scheme (Gillespie and Styles, 1999) (Table 3). Broadly, the subdivision is into fragmental igneous rock and sediment (including volcaniclastic sedimentary rock and sediment, tuffite, and pyroclastic rock and sediment), and crystalline igneous rock, which is subdivided into normal crystalline igneous rock (coarse-grained, medium-grained, and fine-grained), and exotic crystalline igneous rock.

Although the RCS is mostly designed to be applicable to rocks and sediments examined in hand specimen or thin section, within the 'Fine-grained normal crystalline rock' section of the RCS there are classes for rocks that can only be identified using geochemical analysis. Total Alkali versus Silica classification (TAS) is applied to chemically 'normal' fine-grained crystalline rocks where the modal mineralogy cannot be determined either because of the presence of glass, or the fine-grained nature of the rock (Gillespie and Styles, 1999).

4.2. Metamorphic rocks

The recommendations of the IUGS Subcommission on the Systematics of Metamorphic Rocks[26] had not yet been finalised when the metamorphic section of the RCS was developed (Robertson, 1999) (Table 4). In the RCS, the subdivision is into metasedimentary rock (which can be further classified on protolith, modal composition, or textural attributes), meta-igneous rock (also further classifiable on protolith, modal composition, or textural attributes), metamorphic rock with unknown protolith (further classifiable on modal composition or textural attributes), metavolcaniclastic rock, mechanically broken and reconstituted rock (including fault-breccia, cataclastic-rock, and mylonitic-rock), metasomatic-rock, hydrothermal-rock, migmatitic-rock, contact metamorphic rock, and impact metamorphic rock.

4.3. Sediments and sedimentary rocks

The subdivision of sediment and sedimentary rocks is based on several separate schemes developed by different authors for different types of sediments and rocks, many of which overlap (Hallsworth and Knox, 1999) (Tables 5 and 6). Broadly, the subdivision is into siliciclastic sediment and sedimentary rock (which can be further classified as rudaceous, arenaceous, or argillaceous), carbonate sediment and sedimentary rock (further classified into lime-sediment and limestone, dolomite-sediment, dolostone and magnesite-stone, and Na-carbonate sedimentary rock), phosphate-sediment and phosphorite, iron-sediment and ironstone, organic-rich sediment and sedimentary rock (which includes peat and the coal series), non-carbonate salt sediments and rocks (i.e., evaporites), non-clastic siliceous sediment and sedimentary rock (including various oozes, chert, and diatomite), miscellaneous hydroxide, oxide and silicate sediment and sedimentary rock (including illite-, kaolinite- and smectite-clay and claystone, and

[26] https://www2.bgs.ac.uk/scmr/home.html.

Table 3
Outline RCS classification of igneous rocks and sediments. Selected examples are at or below Level 4.

1	2	3	4	Selected Examples	RCS Vol. 1
Igneous rock and sediment	Fragmental igneous rock and sediment	Volcaniclastic igneous rock and sediment	Volcaniclastic sedimentary rock and sediment	Volcaniclastic-sand; Volcaniclastic-breccia; Volcaniclastic-sandstone	Section 4
			Tuffite	Tuffaceous-sand; Tuffaceous-breccia; Tuffaceous-sandstone	
			Pyroclastic rock and sediment	Ash; Lapilli-ash; Tuff; Lapilli-tuff	
	Crystalline igneous rock	Normal crystalline igneous rock	Coarse-grained normal crystalline rock	Anorthosite; Peridotite; Pyroxenite; Granite; Granodiorite	Section 5
			Medium-grained normal crystalline rock	Micro-anorthosite; Microperidotite; Micropyroxenite; Microgranite	
			Fine-grained normal crystalline rock	Rhyolite; Dacite; Trachyte; Andesite; Tephrite; Basalt; 'TAS' classified rocks	
		Exotic crystalline igneous rock		Kimberlite; Lamproite; Carbonatite	Section 6

Table 4
Outline RCS classification of metamorphic rocks. Selected examples are at or below Level 3.

1	2	3	Selected Examples	RCS Vol. 2
Metamorphic rock	Metasedimentary rock	… based on protolith name	Metalimestone	Section 3
		… based on modal composition	Psammite; Calcsilicate-rock	
		… based on textural attributes	Paraschist; Paragneiss	
	Metavolcaniclastic rock		Metavolcaniclastic-sandstone	Section 4
	Meta-igneous rock	… based on protolith name	Meta-anorthosite; Metagranite	Section 5
		… based on modal composition	Serpentinite; Metafelsite	
		… based on textural attributes	Orthoschist; Orthogneiss	
	Metamorphic rock with unknown protolith	… based on modal composition	Marble; Eclogite	Section 6
		… based on textural attributes	Schist; Gneiss	
	Mechanically broken and reconstituted rock		Fault-breccia; Mylonite	Section 7
	Metasomatic-rock		Skarn; Greisen	Sections 8, 9
	Hydrothermal-rock			
	Migmatitic rock			
	Contact metamorphic rock		Hornfels	
	Impact metamorphic rock			

Table 5
Outline RCS classification of sediments and sedimentary rocks. Selected examples are at or below Level 4.

1	2	3	Selected Examples	RCS Vol. 3
Sediment and sedimentary rock	Siliciclastic sediment and sedimentary rock	Siliciclastic rudaceous sediment and rock	Silicate-gravel; Silicate-conglomerate	Section 2
		Siliciclastic arenaceous sediment and rock	Silicate-sand; Silicate-sandstone; Arenite; Wacke	
		Siliciclastic argillaceous sediment and rock	Silicate-mud; Silicate-silt; Silicate-clay; Silicate-mudstone; Silicate-siltstone; Silicate-claystone	
		Diamicton		
		Diamictite		
	Carbonate sediment and sedimentary rock	Lime-sediment and limestone	Lime-gravel; Lime-sand; Lime-mud; Limestone; Ooid-limestone; Lime-grainstone; Chalk	Section 3
		Dolomite-sediment, dolostone and magnesite-stone	Dolomite-gravel; Dolomite-sand; Dolomite-mud; Dolostone	
		Na carbonate sedimentary rock	Natron; Thermonatrite; Trona	
	Phosphate-sediment and phosphorite	Phosphate-sediment	Phosphate-gravel; Phosphate-sand; Phosphate-mud	Section 4
		Phosphorite	Ooid-phosphorite; Phosphate-grainstone; Guano	

bauxite), and generic sedimentary rocks classified by grain-size.

4.4. Superficial deposits

Superficial deposits are fundamentally subdivided into artificial deposits (including aggregate and wastes of various kinds) and natural superficial deposits (McMillan and Powell, 1999) (Table 7). This part of the RCS differs from the other parts in being, in part, genetic in nature. It is subdivided into residual deposit (including duricrust and regolith), biological deposit (including bioclastic sand deposit, and various oozes and biological rocks, so that this part of the classification overlaps with the non-clastic siliceous sediments and rocks mentioned above), and chemical deposit (including bog iron-ore and tufa).

A revised system was developed within BGS in 2006 for naming, classifying, and coding the common unconsolidated natural deposits (clay, silt, sand, gravel, cobbles, boulders, and peat) that was designed to be more consistent with UK civil engineering industry usage. This scheme has largely replaced the entries for these deposits created as part of the original 1999 RCS volumes 3 and 4 (Cooper et al., 2006).

Table 6
Outline RCS classification of sediments and sedimentary rocks (continued). Selected examples are at or below Level 4.

1	2	3	Selected Examples	RCS Vol. 3
Sediment and sedimentary rock	Iron sediment and ironstone	Iron-sediment Ironstone	Iron-gravel; Iron-sand; Iron-mud Ooid-ironstone; Iron-grainstone	Section 5
	Organic-rich sediment and sedimentary rock	Sediment rich in organic matter Sedimentary-rock rich in organic matter	Sediment rich in sapropel (coorongite); Peat Sapropelite; Coal; Anthracite	Section 6
	Sedimentary rock and sediment composed of non-carbonate salts	Non-carbonate salt	Gypsum-gravel; Gypsum-sand; Gypsum-stone; Halite-stone; Anhydrite-stone	Section 7
	Non-clastic siliceous sediment and sedimentary rock	Non-clastic siliceous sediment Siliceous rock	Siliceous-ooze Diatomite; Radiolarite; Chert	Section 8
	Miscellaneous hydroxide, oxide and silicate sediment and sedimentary rock	Monomineralic aluminium-silicate Hydroxides and oxides of alumina and iron	Kaolinite-clay; Illite-clay; Kaolinite-claystone; Illite-claystone Ferricrete; Bauxite; Lithomarge	Section 9
	Sedimentary rock based on grain-size or crystal size	Conglomerate Sandstone Mudstone		Section 10

Table 7
Outline RCS classification of superficial deposits. Selected examples are at or below Level 4.

1	2	3	Selected Examples	RCS Vol. 4
Superficial deposit (natural and/or artificial)	Artificially modified ground	Artificial deposit	Brick; Concrete; Domestic/garden refuse; Blast-furnace slag; Industrial waste; Radioactive waste	Section 3
	Natural superficial deposit	Residual deposit	Calcrete; Ferricrete; Gossan; Saprolite; Weathered rock	Section 4
		Biological deposit	Diatomaceous-ooze; Radiolarian-ooze; Sponge-spicular-ooze; Diatomite; Radiolarite; Spiculite	
		Chemical deposit (natural superficial)	Bog iron-ore; Tufa	
		Peat	Bouldery peat; Clayey peat	Unpublished
		Boulders	Clayey boulders; Gravelly cobbly clayey boulders	
		Cobbles	Bouldery cobbles; Clayey bouldery cobbles	
		Gravel	Bouldery gravel; Clayey bouldery gravel	
		Sand	Bouldery sand; Cobbly bouldery sand	
		Silt	Bouldery silt; Peaty bouldery silt	
		Clay	Bouldery clay; Cobbly bouldery clay	

5. Construction of the SKOS concept scheme

The entire Rock Classification Scheme has been modelled in linked data as a SKOS concept scheme composed of 3 sub-schemes.[27]

The principal RCS sub-scheme is a taxonomic hierarchy defining concepts of single types of rocks and sediments, i.e., all the classes described in full in the 4 original published RCS reports. This is represented in the linked data by the sub-scheme 'RockName'.[28] The SKOS 'broader' and 'narrower' relations[29] between concepts in this sub-scheme represent superclass-subclass relations. As such, this sub-scheme can be considered as an informal ontology. Each child concept inherits all the properties of its broader concept (its super class, or more general parent term), and any instance of a narrower concept (the subclass or more specific child term) must also be an instance of its parent term. Some terms may also have SKOS 'narrower' relations to terms in the composite sub-scheme, these relations are described below.

The second RCS sub-scheme 'RockComposite'[30] contains the 'composite concepts' i.e., the concepts that describe a bulk material that is composed of more than one single rock type (Section 4). This sub-scheme meets the technical requirement in BGS legacy systems of having at most one RCS code per map polygon or logged interval, whilst respecting there are limits to the spatial resolution of mapping or logging observations, and that the real world cannot always be neatly categorised. The concepts in this sub-scheme are all sibling concepts at the same hierarchical level, but each contains SKOS 'broader' relations to one or more concepts in the RockName scheme. This reflects the hierarchical modelling in the relational database model, whereby a highly specialised composite concept is modelled as a child of a less specialised single rock type. Strictly speaking this SKOS broader-narrower relationship is a whole-part relationship, where a composite rock type whole has multiple relations to its single rock type constituent parts. A child RockComposite concept in this scheme may inherit some properties of its multiple parent RockName concepts such as compositional properties, but would not inherit bulk properties such as strength or density.

The third and final RCS sub-scheme 'RockDummy'[31] consists of the handful of 'dummy concepts' for materials that are discontinuities and fillings, rather than rocks or sediments (Section 4). All concepts are currently siblings at the same hierarchical level. Any future hierarchical relations would be expected to be subclass-superclass relations. There are no relations to RockComposite or RockName sub-scheme terms.

27 https://data.bgs.ac.uk/.
28 https://data.bgs.ac.uk/ref/EarthMaterialClass/RockName.html.
29 https://www.w3.org/TR/skos-primer/#sechierarchy.
30 https://data.bgs.ac.uk/ref/EarthMaterialClass/RockComposite.html.
31 https://data.bgs.ac.uk/ref/EarthMaterialClass/RockDummy.html.

The identifiers for the linked data representation were created with care to ensure they would be persistent and technology agnostic. The W3C principles of “Cool URIs”,[32] the UK Government recommendation “Designing URI Sets for the UK Public Sector”[33] and the UK Ordnance Survey’s implementation (Goodwin, 2013) were taken into account. Identifiers for concepts take the form https://{domain}/id/{scheme}/{subscheme}/{reference}, e.g. https://data.bgs.ac.uk/id/EarthMaterialClass/RockName/AREN. Representations are handled by content negotiation, https://{domain}/id/{scheme}/{subscheme}/{reference}/{doc.file-extension}, e.g. https://data.bgs.ac.uk/id/EarthMaterialClass/RockName/AREN.rdf. Identifiers for schemes and predicates take the form https://{domain}/ref/{scheme}/{subscheme}/{reference}, e.g. https://data.bgs.ac.uk/ref/EarthMaterialClass/RockName. References for concept schemes use camel case with a capital letter, e.g. EarthMaterialClass. References for predicates use camel case with a lower-case letter, although no self-defined predicates are used in the RCS linked data scheme.

We use the domain http(s)://data.bgs.ac.uk. We host the linked data ourselves and have full control of this domain, so we don’t need to use an external Persistent Identifier service such as https://w3id.org/. We did consider the risk that the domain name might change in the future if our organisational name changed (see[34] for BGS’ history of name changes), but decided to accept that risk, noting that Ordnance Survey - a leading adopter of linked data at the time - used http://data.ordnancesurvey.co.uk. We opted not to include a date or version number with the persistent identifiers. The URI always returns the latest available data.

Plain SQL scripts are used to extract the data from the RDBMS where they are maintained (Section 4) and render them as N-Triples which are saved in plain text files. The Rdflib library[35] is used to validate the triples before loading them into a Fuseki Jena triplestore.[36] The triples are also uploaded to the Vocabularies project in the British Geological Survey github repository.[37] This provides the public versioning history and allows bulk download of the triples.

Because the RCS and its alphabetic identifiers (the ‘RCS codes’) were conceived and designed in the 1990s long before any intentions to use on the Web (Section 4), some of the identifiers include special characters that are invalid in a URI.[38] During export to create a concept identifier for linked data publishing, these illegal characters have been substituted for web safe characters in a predictable and reversible translation as shown in Table 8.

Figs. 4–6 show the BGS linked data website and the linked data representation of a single term in the RCS. Using the schemas RDF (Cyganiak et al., 2014), RDFS (Brickley and Guha, 2014), SKOS (Miles and Bechhofer, 2009), VS (Brickley, 2009) and DCT (Albertoni et al., 2020), the attributes presented in the linked data for each RCS concept are: rdf:type, skos:inScheme, skos:notation, rdfs:label, skos:prefLabel, skos:definition, skos:broader, skos:narrower, and vs:term_status. The attributes presented to describe the RCS scheme and sub-schemes are: rdf:type, dct:title, rdfs:label, dct:description, dct:isFormatOf, rdf:about, dct:hasPart or dct:part of, dct:created, dct:modified, dct:creator, dct:publisher, dct:identifier, and dct:bibliographicCitation.

The RCS concepts are themselves the object of triples in other BGS linked data schemes, namely The Lexicon of Named Rock Units[39] and the linked data representation of the BGS 1:625 000 scale geological map[40].

An experimental public SPARQL endpoint on the Fuseki Jena triplestore is made available through an instance of VocPrez; link available on request to the authors. Future work on this web API and SPARQL endpoint is discussed below.

Table 8
Substitution characters used to replace illegal characters in RCS identifiers during construction of linked data URIs.

Illegal character	Substitution character for linked data identifier
+	P
#	H
*	S
ˆ	C
~	T
@	A

6. Evaluation

Clearly if BGS was setting out today to implement a rock classification scheme to support its digital geological map production and other workflows, we would use one of the published schemes that now exists, for example the CGI or INSPIRE lithology schemes. However, in the late 1990s when the need for a controlled vocabulary for use with our digital workflows became apparent, no such scheme yet existed. Moreover, the standards that now guide things like development of persistent identifiers did not exist yet or were in their infancy. So, it was necessary to devise a scheme from scratch.

The RCS can now be considered a mature classification. It has been used for more than 20 years to drive the workflows within the UK’s national geological survey. While the source RCS reports have not been updated since their original publication in 1999, the database implementation of the classification has had many additions made as and when necessary. These additions have almost always been in the form of ‘term plus qualifier’ concepts (e.g., biotite granite) and ‘composite concepts’ (e.g., granite and diorite) as described in Section 4.

While the RCS is a simple classification in principle, its size and level of detail has been criticised; it is likely many entries have never been used by any BGS geologist. This highlights the importance of being able to subset a large vocabulary in a way that is suitable to the requirement. In most situations, the geologist knows which relatively small set of rock types they want to refer to, and do not need or want to be burdened with the entire vocabulary.

Compared with the classifications described in Section 3, the RCS goes into greater detail on unconsolidated sediments and deposits, and anthropogenic deposits. This reflects its development and use in the British Isles, whose shallow subsurface, the ‘zone of human interaction’, is heavily impacted by glaciation, and by industrialisation and other human activities, and where a lot of construction work takes place on floodplains.

The CGI and INSPIRE vocabularies probably match the RCS most closely in scope, but do not go into as much detail, in terms of the number of levels of classification. The QGS lithotype scheme goes to great levels of detail, at least as much as the RCS, particularly in its inclusion of detailed concepts with qualifiers representing local variants of rock types (for example Shap granite). It is notable that it includes classifications of ore rocks and meteorites, not included in the other schemes described here. Like the RCS focus on unconsolidated and artificial deposits, the QGS inclusion of ores and meteorites appears to reflect aspects of the geology of the country where it was developed and the interests of those who developed it. The USGS Lithologic Classification is in some ways the least detailed of the vocabularies described,

[32] https://www.w3.org/TR/cooluris/.
[33] https://www.gov.uk/government/publications/designing-uri-sets-for-the-uk-public-sector.
[34] https://www.bgs.ac.uk/about-bgs/our-work/our-history/.
[35] https://github.com/RDFLib/rdflib.
[36] https://jena.apache.org/documentation/fuseki2/.
[37] https://github.com/BritishGeologicalSurvey/vocabularies.
[38] https://www.rfc-editor.org/rfc/rfc3986.
[39] https://data.bgs.ac.uk/doc/Lexicon.html.
[40] https://data.bgs.ac.uk/doc/625KGeologyMap.html.

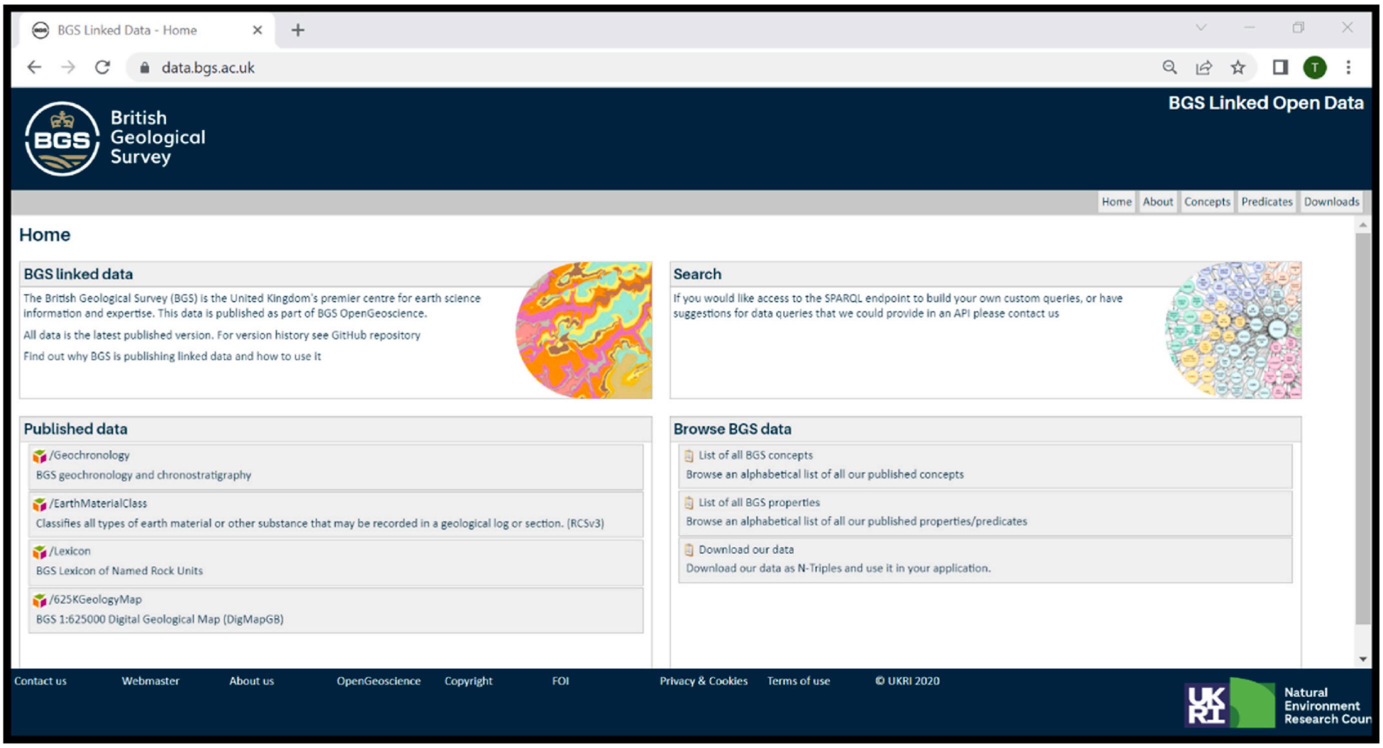

Fig. 4. BGS linked data home.

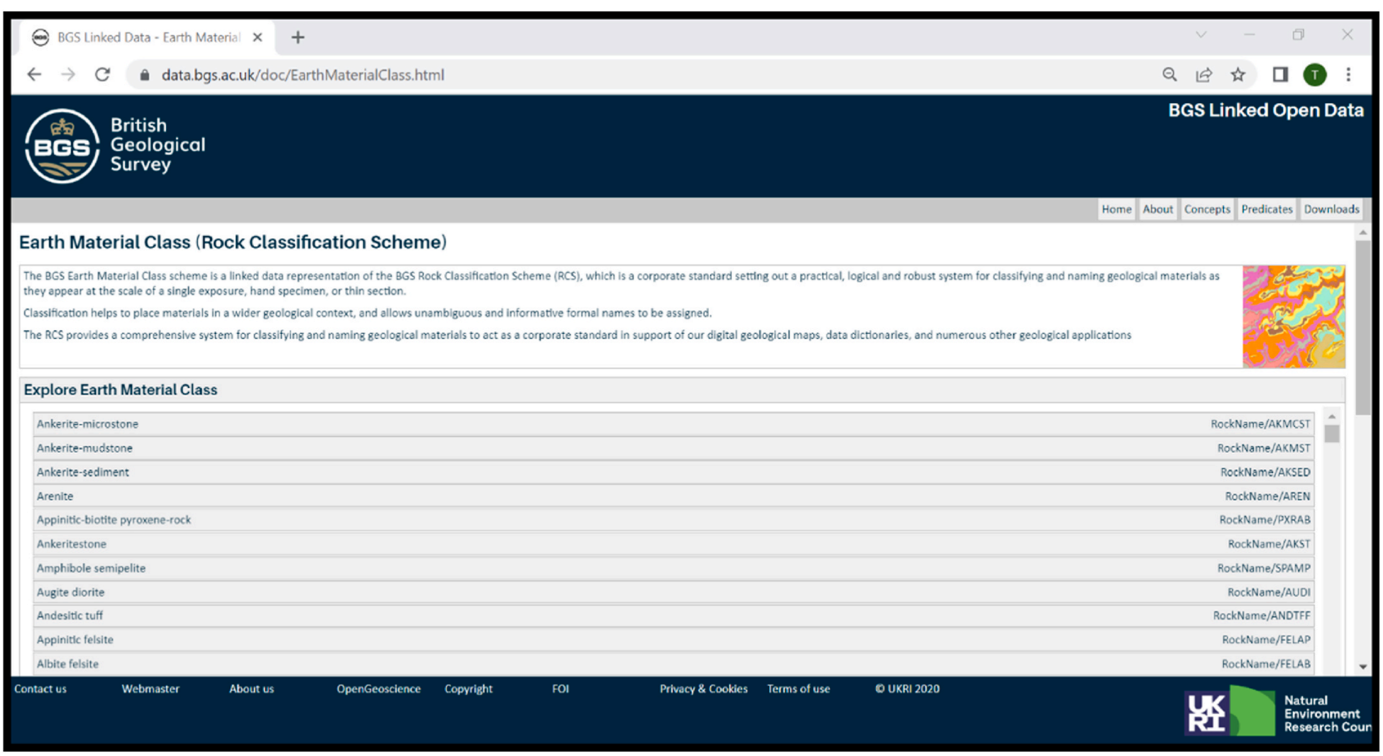

Fig. 5. Rock Classification Scheme as linked data.

but the level of detail it goes to and the types of concepts it includes reflect its primary use, which is to attribute geological map units. As an 'unfussy' classification, particularly for use in the field, it appears very successful.

Mindat.org is the only one of the resources described here that is undergoing continuous update and addition so far as we are aware and is the only one that can be said to be 'crowd-sourced', at least in so far as its pages are authored by volunteers. The CGI, INSPIRE, and USGS vocabularies were developed fully formed by expert groups. The RCS was also initially published in 1999 as a fully formed scheme, albeit that additions have been made to the database version subsequently.

Should the user wish to navigate a hierarchical classification of rock

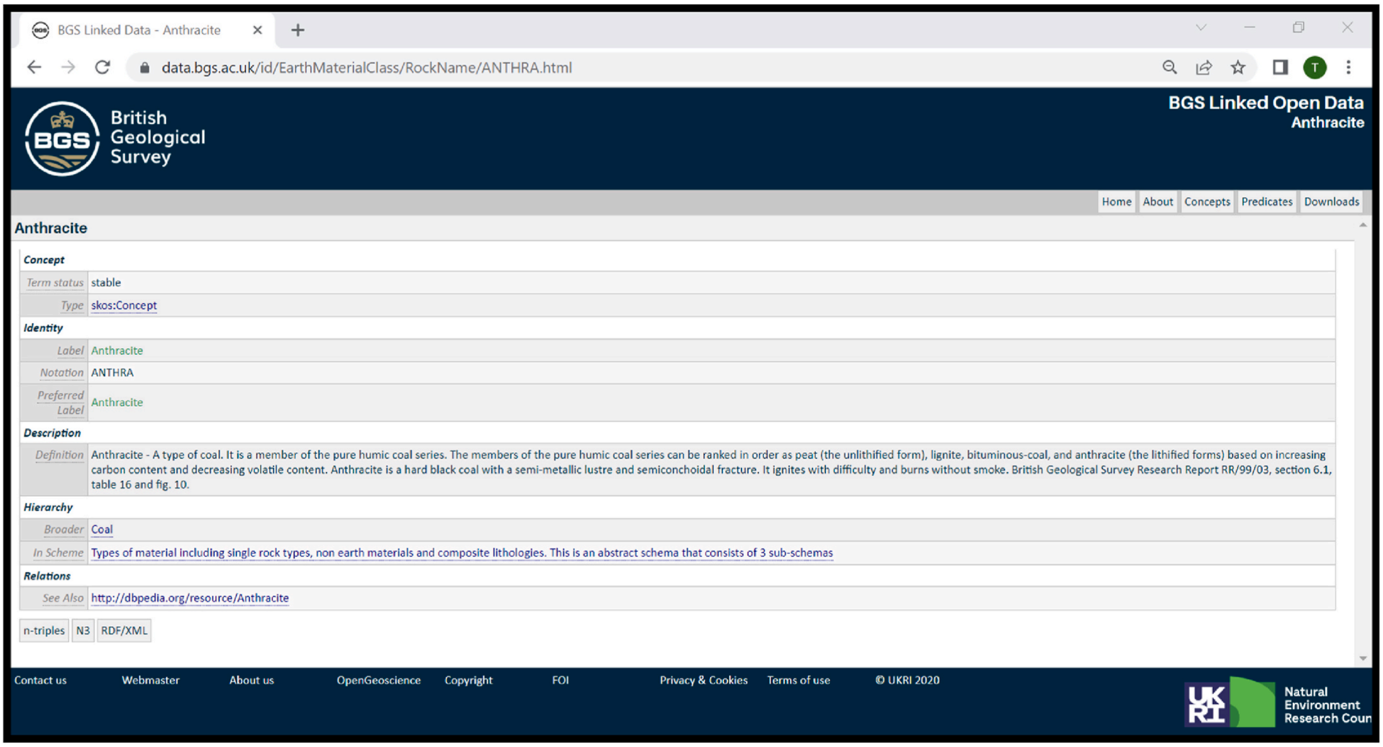

Fig. 6. An RCS entry displayed as linked data.

and sediment types, moving from broader to narrower concepts and vice versa, as might be the case when navigating the classification in a drop-down list in a software application, this is simpler with the mono-hierarchical classifications like the RCS, INSPIRE and USGS classifications. The poly-hierarchical nature of the CGI and QGS classifications makes such navigation problematic, or at least confusing, because segments of the classification are repeated, sometimes multiple times. The multi parent structure is designed to make entries more discoverable because it provides more than one route into the classification, for example by providing the ability to discover a concept classified by either grain size or chemistry. But it can make the classification, when expanded, baffling.

Considering the linked data representation of the RCS, its organisation into three sub-schemes (RockName, RockComposite, and RockDummy, see Section 5) reflects the rules on how the concepts are organised hierarchically. One drawback with having the three sub-schemes is that when an RCS code is used in a dataset it is not immediately known which of the sub-schemes it belongs to, without first doing a look up – either a trial and error lookup of the three possible URI roots using the linked data API, a SPARQL query using the SKOS:notation attribute, or (if an internal user) a lookup in the relational database. One solution to this could be to present a unified scheme in parallel to the sub-schemes, with SKOS:exactMatch relations to the equivalent sub-scheme term. Work to resolve this is currently underway.

7. Future work

Several controlled vocabularies for lithology are now available, each following the same broad outline but differing from each other in detail. Lithology is a large and complex ontology to build. However, the fact that there is a finite set of sources on which it can be based tends to result in broad similarity between vocabularies, both in their overall arrangement and in their detailed term definitions. This should make the task of aligning lithology vocabularies easier than it might otherwise be.

Alignment of taxonomies or ontologies has an important role in integration of knowledge from different sources but can be laborious and requires domain expertise to perform manually. Automated ontology alignment (or ontology matching) techniques are an active area of research, as evidenced in The Ontology Alignment Initiative (OAEI)[41] which provides an annual evaluation and benchmarking event for alignment/matching systems. The European Commission's Interoperable Europe Initiative provides guidelines for how to approach alignment, depending on the expressivity of the taxonomy languages involved.[42]

Classic systems such as LogMap (Jiménez-Ruiz and Cuenca Grau, 2011) and AgreementMakerLight (AML) (Faria et al., 2013) use lexical similarity, structural matching, and logical reasoning. More recent systems augment those approaches with Machine Learning (ML) techniques of word embedding which can use contextual knowledge to infer semantic similarity (e.g., Chen et al., 2021; He et al., 2022). Nkisi-Orji et al. (2019) produced an algorithm for ontology alignment based on word embedding and random forest classification; this can be applied to SKOS vocabularies. Another approach could be to model the RCS as a formal ontology with additional properties and use the property ranges of classes (grain size, mineralogy, etc.) as the basis of automated alignment with other similarly expressive ontologies. The Geoscience Ontology (Brodaric and Richard, 2021) could be used as a starting point for such an ontology. To date the RCS has not been aligned to other taxonomies or ontologies. This will be the subject of future work which will have the benefit of linking the RCS scheme into a larger fused geoscience knowledge graph.

Since the CGI/INSPIRE vocabularies broadly resemble the RCS, it seems likely that alignment of many core RCS concepts (i.e., excluding 'term plus qualifier' concepts and 'composite concepts') with their CGI/INSPIRE equivalents can be made at least on 'close match' basis. For example, the RCS definition of granite as "a coarse-grained crystalline igneous rock whose mineral mode plots in field 3 of the quartz - alkali

[41] http://oaei.ontologymatching.org/.

[42] https://joinup.ec.europa.eu/sites/default/files/document/2011-12/guidelines-and-good-practices-for-taxonomies-v1.3a.pdf.

feldspar - plagioclase - feldspathoid (QAPF) diagram" is closely similar to the CGI definition "Phaneritic crystalline rock consisting of quartz, alkali feldspar and plagioclase (typically sodic) in variable amounts, usually with biotite and/or hornblende. Includes rocks defined modally in QAPF Field 3" ('phaneritic' means having crystal size big enough to be visible to the naked eye).

The same may be true for alignment between the RCS and QGS vocabularies, although direct alignment may not be possible for the most highly detailed terms in both classifications, and obviously not for the ore and meteorite classes that are present in QGS but not in RCS (but could be added). One might expect that 'close match' alignment between the RCS and the equivalent terms in the USGS vocabulary should be possible, but note that, as described in Section 3.4, the purpose of the USGS classification is slightly different to the others and it includes lithomorphogenetic terms that do not appear in the RCS or other vocabularies.

Other future work could be to represent the RCS as a formal web ontology using OWL[43]. Stephen et al. (2023) present a design pattern for formalising scientific taxonomies as ontologies which they consider is "crucial in their utilization for data integration, discovery, and exploration purposes within KGs" (Knowledge Graphs). One of the advantages of doing so would be that the different semantic meaning of the current SKOS hierarchical relationships would be translated to more precise ontological relations, enabling reasoning. Furthermore, it's possible that additional knowledge presented in the original RCS research reports could be expressed in a formal ontology in addition to the basic taxonomy structure. This would enable additional reasoning and integration using rock and sediment properties such as grain size and mineral content.

Work is in progress to update the API and user interface that presents the RCS linked data at https://data.bgs.ac.uk, using the more recent Prez[44] and Prez-UI[45] software. The SPARQL endpoint will be made available as a production service and we intend to share example SPARQL queries, for example to obtain all child terms of a given term, or to use the associated linked data datasets to find all map polygons that have been attributed with a particular RCS code. The user interface will allow drill down navigation of the classification hierarchy. As part of this work, some new predicates are being added, and the '.ttl' versioned files on the github repository will be re-organised so that individual concept schemes such as the RCS can be easily downloaded in entirety.

8. Conclusions

Several controlled vocabularies for lithology have been published over the last two decades, and some of them are available as linked data. The BGS RCS was the first classification scheme to be published but it has not so far been available in a modern machine-readable format. This we have now done.

The degree to which the RCS is embedded in our workflows and products makes it impossible to replace it in our processes with another vocabulary at this time. But by publishing it as an informal SKOS vocabulary, we make it available for use in modern semantic applications and visualisations both by ourselves and third parties, and we also make it possible to use the tools SKOS provides to formally align the RCS to other published lithology classifications, thereby embedding it in the wider geoscience knowledge graph.

Comparing the RCS with the other schemes, while they are broadly similar, each has characteristics that reveal the interests and requirements of the groups that developed them, in terms of level of detail and the concepts represented. The BGS RCS for example places emphasis on unconsolidated superficial deposits and anthropogenic deposits, reflecting its development in a country where human activities often take place on floodplains or glaciated landscapes, and where those human activities have transformed the landscape to a great degree.

No alignment of the RCS with any of the other classification schemes has yet been done. This should certainly be possible but doing it manually will be difficult and time-consuming. Several approaches to automated or semi-automated alignment can be considered and the RCS affords us an opportunity in future to test these approaches.

CRediT authorship contribution statement

Tim McCormick: Writing – original draft, Data curation, Conceptualization. **Rachel E. Heaven:** Writing – original draft, Methodology, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

We thank two anonymous editors and two anonymous reviewers of an early draft of this paper whose comments helped us to improve the clarity of the paper. This work was funded by British Geological Survey National Capability funding. We gratefully acknowledge the many BGS geologists and data managers who have contributed to the content of the RCS over the years.

References

Albertoni, R., Browning, D., Cox, S., Gonzalez Beltran, A., Perego, A., Winstanley, P., 2020. Data Catalog Vocabulary (DCAT) - Version 2. W3C Recommendation 04 February 2020. https://www.w3.org/TR/vocab-dcat-2/. (Accessed 8 May 2023).

Bow, J., 2015. BGS SIGMA 2015: External User Guide for BGS SIGMA Mobile and Desktop Toolbars. British Geological Survey Open Report, Keyworth, UK, p. 63pp. OR/15/029.

Brickley, D., 2009. Term-centric semantic web vocabulary Annotations. In: (Editor's Draft of a Potential) W3C Interest Group Note 31 December 2009. https://www.w3.org/2003/06/sw-vocab-status/note.html. (Accessed 8 May 2023).

Brickley, D., Guha, R.V., 2014. RDF Schema 1.1. W3C Recommendation 25 February 2014. https://www.w3.org/TR/rdf-schema/. (Accessed 8 May 2023).

Brodaric, B., Richard, S.M., 2021. GeoScience Ontology: Release 1.0.2 (1.0.2). Zenodo. https://doi.org/10.5281/zenodo.4750707.

Chen, J., Jiménez-Ruiz, E., Horrocks, I., Antonyrajah, D., Hadian, A., Lee, J., 2021. Augmenting ontology alignment by semantic embedding and distant supervision. In: The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings 18. Springer International Publishing, pp. 392–408.

Cooper, A.H., Kessler, H., Ford, J., 2006. A Revised Scheme for Coding Unlithified Deposits (Also Applicable to Engineering Soils). IR/05/123. British Geological Survey Internal Report, Keyworth, UK, p. 45 (Unpublished).

Cyganiak, R., Wood, D., Lanthaler, M., 2014. RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation 25 February 2014. https://www.w3.org/TR/rdf11-concepts/. (Accessed 8 May 2023).

Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I.F., Couto, F.M., 2013. The agreementmakerlight ontology matching system. In: Meersman, R., Panetto, H., Dillon, T., Eder, J., Bellahsene, Z., Ritter, N., Leenheer, P., Dou, D. (Eds.), On the Move to Meaningful Internet Systems: OTM 2013 Conferences: Confederated International Conferences: CoopIS, DOA-Trusted Cloud, and ODBASE 2013, Graz, Austria, September 9-13, 2013, Proceedings. Springer, Berlin, Heidelberg, pp. 527–541.

Gillespie, M.R., Styles, M.T., 1999. BGS rock classification scheme volume 1. Classification of igneous rocks. British Geol. Surv. Res. Report 52, 2nd edition) RR/99/06, Keyworth, UK.

Goodwin, J., 2013. New ordnance survey linked data service. September 2013 Geoinformatics 16 (6), 30–31. Emmeloord, Netherlands.

Hallsworth, C.R., Knox, R.W.O.'B., 1999. BGS Rock Classification Scheme Volume 3. Classification of Sediments and Sedimentary Rocks. British Geological Survey Research Report RR/99/03, Keyworth, UK, p. 44pp.

[43] https://www.w3.org/OWL/.
[44] https://github.com/RDFLib/prez.
[45] https://github.com/RDFLib/prez-ui.

He, Y., Chen, J., Antonyrajah, D., Horrocks, I., 2022. BERTMap: a BERT-based ontology alignment system. Proc. AAAI Conf. Artif. Intell. 36 (5), 5684–5691.

Hughes, L., 2013. Digital geological mapping of UK Quaternary deposits using the BGS SIGMA system. Bull. Soc. Cartogr. 47, 17–20.

Jiménez-Ruiz, E., Cuenca Grau, B., 2011. LogMap: logic-based and scalable ontology matching. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (Eds.), The Semantic Web–ISWC 2011: 10th International Semantic Web Conference. Springer, Berlin, Heidelberg, pp. 273–288. Bonn, Germany, October 23-27, 2011, Proceedings, Part I 10.

Jones, D., McCormick, T., 2020. Strengthening geoscience in developing nations. Geoscientist 30 (5), 14–17.

Le Maitre, R.W., Bateman, P., Dudek, A.J., Keller, M.J., 1989. A Classification of Igneous Rocks and Glossary of Terms. Blackwell, Oxford, p. 193.

McCormick, T., 2021. The British Geological Survey Rock Classification Scheme. Paper presented at ESIP Summer Meeting, July 19-23, 2021, Virtual Online. https://2021esipsummermeeting.sched.com/event/jMO5/vocabularies-for-rock-type-categories. (Accessed 21 September 2023).

McMillan, A.A., Powell, J.H., 1999. BGS Rock Classification Scheme Volume 4. Classification of Artificial (Man-made) Ground and Natural Superficial Deposits – Applications to Geological Maps and Datasets in the UK. British Geological Survey Research Report RR/99/04, Keyworth, UK, p. 65pp.

Miles, A., Bechhofer, S., 2009. SKOS Simple Knowledge Organization System Namespace Document - HTML Variant. 18 August 2009 Recommendation Edition. https://www.w3.org/2009/08/skos-reference/skos.html. (Accessed 8 May 2023).

Nkisi-Orji, I., Wiratunga, N., Massie, S., Hui, K.Y., Heaven, R., 2019. Ontology alignment based on word embedding and random forest classification. In: Berlingerio, M., Bonchi, F., Gärtner, T., Hurley, N., Ifrim, G. (Eds.), Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part I 18. Springer International Publishing, pp. 557–572.

Robertson, S., 1999. BGS rock classification scheme volume 2. In: Classification of Metamorphic Rocks. British Geological Survey Research Report RR/99/02, Keyworth, UK, p. 24pp.

Smith, A., 2013. Digital Geological Map of Great Britain, Information Notes, vol. 2013. British Geological Survey Open Report OR/13/007, Keyworth, UK, p. 54pp.

Stephen, S., Shimizu, C., Schildhauer, M., Zhu, R., Janowicz, K., Hitzler, P., 2023. A Pattern for Representing Scientific Taxonomies. Paper presented at. 13th Workshop on Ontology Design and Patterns, WOP 2022. Virtual Online, 24/10/22.

Thomas, R., 2010. Digital geological mapping at the British Geological Survey with the SIGMA system: examples from the United Arab Emirates and Tanzania. Johannesburg, South Africa, 17 Sept 2010. In: Modern Mapping Techniques. Geological Society of South Africa, Johannesburg, South Africa (Unpublished).

Wood, B., Kessler, H., 2021. Model creation based on digital borehole records and interpreted geological cross-sections. In: Turner, A.K., Kessler, H., van der Meulen, M.J. (Eds.), Applied Multidimensional Geological Modeling. Wiley Blackwell, pp. 235–246.