Computers & Geosciences 145 (2020) 104620

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

A new structure for representing and tracking version information in a deep time knowledge graph

Xiaogang Ma[a,*], Chao Ma[a], Chengbin Wang[b]

[a] Department of Computer Science, University of Idaho, 875 Perimeter Drive, MS 1010, Moscow, ID 83844-1010, USA
[b] State Key Laboratory of Geological Processes and Mineral Resources & School of Earth Resources, China University of Geosciences, Wuhan, 430074, China

ARTICLE INFO

Keywords:
Knowledge graph
Geologic time scale
Ontology
Version control
Open data

ABSTRACT

Ontologies and vocabularies are an effective way to promote data interoperability in open data and open science. The deep time knowledge graph is one of the most discussed and studied topics in geoscience ontologies and vocabularies. The continuous evolution of deep time concepts calls for a mechanism of version control and organization to reduce the semantic ambiguity. In this paper we propose a new structure for version control and tracking of concepts, attributes and topological relationships in the deep time knowledge graph. In our work we have reused the existing ontologies for geologic time scale and vocabularies for the International (Chrono) stratigraphic Chart (ISC). Through the new structure, we are able to represent the whole version history of the ISC charts (from 2004 to 2018) in a single knowledge graph. Moreover, the resulting knowledge graph is consistent with the existing ontologies and vocabularies. Experiments of SPARQL queries prove the efficiency of this structure for version tracking of concepts and attributes. We are now extending the knowledge graph with concepts from regional and local geologic time standards, such as North America, Europe, Britain, China, and Australia, and building a graphic user interface for the services. In a future work, we will implement the knowledge graph in data integration workflows. We hope this research will spur more discussion and development of methods for version control of knowledge graphs in geoscience and other disciplines.

1. Introduction

The open data and open science movements are transforming the modes of research in many fields of research, including geoscience (Kitchin, 2014; Nosek et al., 2015; Cutcher-Gershenfeld et al., 2017). With massive datasets being collected, rescued, curated, and shared online, geoscientists are now able to assemble big data and plan data-intensive research. However, many geoscientists often struggle to clean the data, connect the small data from different sources, and develop hypotheses for data analytics (Fox and Hendler, 2014; Yang et al., 2017; Zaslavsky et al., 2017). Deep time (i.e., geologic time) is a fundamental topic in geoscience research, and can be used as a common framework to join various data silos (NRC(National Research Council) , 2008; Hazen et al., 2019; Morrison et al., 2019). In recent years, several scientific initiatives have proposed to leverage open data to facilitate deep time research, such as the Deep-time Data Driven Discovery (4D) (4D Initiative Team, 2018) and the International Union of Geological Sciences' Deep-time Digital Earth (DDE) program (Normile, 2019). To gain insights into the evolutionary history of the Earth and its uncertain future, we need to collect and analyze diverse deep time data from disparate disciplines. Although "deep time" has been recognized as a central theme for assembling data from numerous sources, an effective strategy and the corresponding methods are needed. The authors are affiliate data science researchers for both the 4D and the DDE programs. Our previous discussion with various researchers in those two programs has shown the potential of a deep time knowledge graph for addressing the need for effective data integration. This paper will present one of the research outputs on the deep time knowledge graph.

In many of the existing open data facilities, the geologic time concepts are recorded in heterogeneous terminologies, which are hard for both human and machine to read, understand and use. Machine-readable knowledge graphs are a proven way to address the issue of concept heterogeneity (Ma et al., 2014; Gil et al., 2019). Sheth et al. (2019) and Hogan et al. (2020) have offered comprehensive reviews on the history of knowledge graphs and their applications. Hogan et al. (2020) gave an inclusive definition to knowledge graph: "a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations

* Corresponding author.
E-mail address: max@uidaho.edu (X. Ma).

https://doi.org/10.1016/j.cageo.2020.104620
Received 12 March 2020; Received in revised form 28 August 2020; Accepted 24 September 2020
Available online 28 September 2020

between these entities." The deep time knowledge graph discussed in this paper is derived from the concept of ontology in the Semantic Web (Fig. 1) as well as previous works on geologic time ontologies and vocabularies. Each ontology is the formal specification of a shared conceptualization of a domain (Gruber, 1995). For the topic of deep time, well-defined concepts and relationships will reduce the ambiguity of terminologies and facilitate data interoperability among data facilities where those terminologies are used (Fig. 1). Following the technical approaches of knowledge and ontology engineering (Paulheim, 2017; Kendall and McGuinness, 2019; Hogan, 2020), a knowledge graph in real-world practices often shows up as the mixture of TBox and ABox. The former is a list of classes and properties, and the latter is a list of instances of those classes. In a simplified understanding, one can perceive TBox as ontologies and ABox as vocabularies. Nevertheless, in real-world examples we can often see overlaps between the two, i.e. an ontology contains some instances, or a vocabulary includes definition of classes and properties. There have been many previous studies on ontologies and vocabularies for a formal digital representation of the geologic time scale. In this paper we use the term deep time knowledge graph as a general reference to those works. We define the deep time knowledge graph as a formal representation of concepts, attributes, and relationships in geologic time. When we introduce the technologies in detail in the subsequent sections, we also use the terms ontology and vocabulary.

Existing work on deep time knowledge graphs (Cox and Richard, 2005; Raskin and Pan, 2005; Perrin et al., 2011; Ma et al., 2011b; Ma and Fox, 2013) shows that there are two key classes in deep time: instant (i.e., a point in time) and interval (i.e., a period of time). For example, Jurassic is an instance of interval and the beginning or end of Jurassic is an instance of instant. In recent years, extra efforts have been made by Cox and Richard (2015) to refine and curate the deep time knowledge graph. They used Unified Modeling Language (UML) and Web Ontology Language (OWL) to incorporate the ISO/TC 211 modeling conventions for spatial and temporal data, which formalize a model for representing the geologic time scale. In their very recent work, the Time Ontology (Cox, 2016; Cox and Little, 2020) was incorporated into the deep time knowledge graph to improve the description of time reference systems (i.e. geometry) and the ordering relationships (i.e., topology) of deep time concepts. Within the knowledge graph, they have developed several core ontologies to set up the conceptual framework (i.e., TBox). Then, for each version of the International (Chrono)stratigraphic Chart (ISC) released by the International Commission on Stratigraphy (Cohen et al., 2013), a vocabulary (i.e., ABox) has been developed. Cox et al. (2016) have also applied the Spatial Information Services Stack Vocabulary Service (SISSVoc) to set up user-friendly access to those vocabularies. Through SISSVoc, the Uniform Resource Identifier (URI) of a concept automatically redirects to the information in the latest version of the vocabulary.

A well-organized deep time knowledge graph will add efficiency in data cleansing and integration because it reduces the ambiguity of geologic time concepts. For example, it was reported that in the 1980s the U.S. Geological Survey had reworked all of its vocabularies and maps due to the change to the definition of Pleistocene in the geologic time scale (Mascarelli, 2009). Such heavy rework can be avoided if the versions of geologic time ontologies and vocabularies are clearly defined and they are used in the digital map databases. The deep time knowledge graph will have the same benefit for the 4D and DDE research programs as well as many other projects. The work led by Cox and Richard has greatly promoted the machine-readability and standard-compatibility of the deep time knowledge graph. Yet, it is tedious to create a new vocabulary for each version of the ISC chart, and it is complicated or even impossible to query the updated concepts, attributes, relationships or their version history across those vocabularies. To address that challenge, we design and implement a new structure in this paper to capture and display different versions of deep time concepts in the ISC chart. Our aim is not to reinvent the wheel. Instead, the new structure is based on the ontologies developed by Cox and Richard and it reuses the existing vocabularies, which makes the result fully compatible with the existing work. In Cox and Richard's work, each version of the vocabulary has a unique scheme identifier. The key idea of our work is to use the vocabulary scheme identifiers in the description of concepts and attributes to record the history of updates. Through this new design, we have captured and represented all the versions of the ISC charts (from 2004 to 2018) in a single knowledge graph. A service of the new knowledge graph has been set up, and result of test queries shows that it is functional to provide the version information of concepts, attributes and relationships between concepts. In the future, when a new version of the ISC chart is released, we just need to update the descriptions in this single knowledge graph, which will reduce the time and efforts spent on knowledge graph maintenance.

The remainder of the paper is organized as follows. Section 2 presents more background information about the ontologies and vocabularies in the existing deep time knowledge graph and their limitation on version control and tracking. Then the method of our version-tracking work is described in detail. Section 3 demonstrates the resulting knowledge graph and results from experiments on querying the version information. Section 4 discusses the advantage of this work with comparison to existing ontologies and vocabularies, and describes our plan for future work. Finally, Section 5 summarizes the main contributions of this paper and offers suggestions for similar work in other fields.

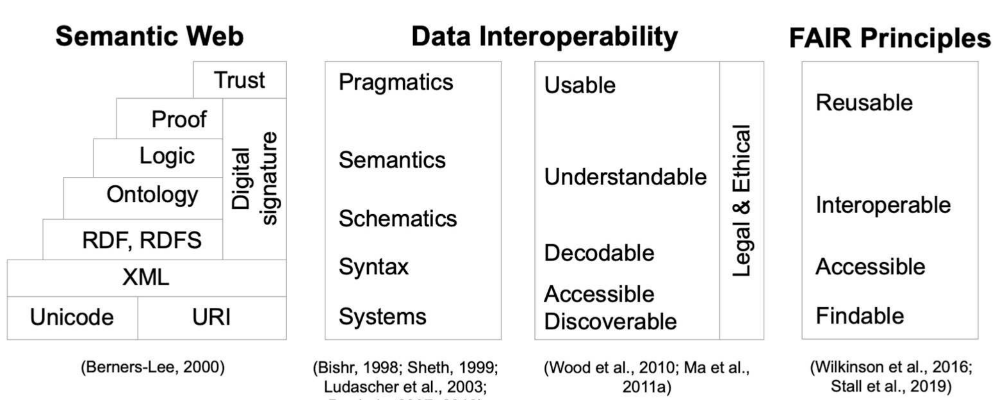

Fig. 1. Layers of data interoperability and comparison to Semantic Web technical architecture and FAIR data principles (from Bishr, 1998; Sheth, 1999; Berners-Lee, 2000; Ludascher et al., 2003; Brodaric, 2007, 2018; Wood et al., 2010; Ma et al., 2011a; Wilkinson et al., 2016; Stall et al., 2019).

2. A version control structure for the deep time knowledge graph

This section will first give an overview to the existing ontologies and vocabularies for geologic time, analyze their limitation on version control, and then introduce the designed new structure. The consistency of the new structure to the existing ontologies and vocabularies will also be presented.

2.1. Existing ontologies and vocabularies in the deep time knowledge graph

In the earlier work of the deep time knowledge graph (Cox and Richard, 2015), two core ontologies were created: the Temporal Hierarchical Ordinal Reference System (THORS) ontology and the Geological Time Scale (GTS) ontology. The former reflects the temporal reference system and topology model in ISO 19108 (ISO/TC-211, 2002), and the latter implements THORS, ISO 19156 (ISO/TC-211, 2011) and several other ontologies to set up a unique conceptual structure for deep time. In the recent updates (Cox and Richard, 2015; Cox, 2016), several Open Geospatial Consortium (OGC) standards and World Wide Web Consortium (W3C) recommendations, such as the Time Oncology (Cox and Little, 2020) and GeoSPARQL (OGC, 2012) were used to further formalize the deep time knowledge graph. The adoption of community standards improves the interoperability of the resulting knowledge graph with the broad open knowledge network on the Web. Table 1 gives a list of the core ontologies in the current deep time knowledge graph and briefly describes their roles. The source code of the knowledge graph is hosted on the GeoSciML website (Cox, 2019), where details about all the ontologies and schemas used in the knowledge graph can be found.

A primary reason for us to design a new structure for version control and tracking in the deep time knowledge graph is that we found a limitation in the existing knowledge graph. Currently, there is a separate vocabulary created for each version of the ISC chart (Cox, 2019). Each vocabulary is designed as a concept scheme with a unique URI. For example, the 2017 version of the ISC chart has a URI ts:isc 2017. Each deep time concept has a persistent URI across those different versions of vocabularies. For example, the URI of Jurassic is isc:Jurassic. In the current knowledge graph, there is a triple describing if a time concept appears in a certain vocabulary scheme. For example, isc:Jurassic skos:inScheme ts:isc 2017. However, for other attributes of the time concept, there is no mechanism for such version control. For example, the age at the base boundary of Jurassic was 199.6 ± 0.6 Ma in the ISC charts before 2010, and it was changed to 201.3 ± 0.2 Ma in the 2012 ISC chart and has remained the same in subsequent versions. Although those numeric values were recorded in the corresponding versions of the vocabulary schemes (Fig. 2), there is no easy way to track the history of changes if we want to make a single query starting from the URI of Jurassic. In other words, we need to query Jurassic many times across the services of all those vocabulary schemes to find the attributes of Jurassic and arrange the version history by ourselves. Intuitively, we can see that a potential improvement to the code in Fig. 2 is to add version information of the corresponding vocabulary scheme in the specification of isc:BaseJurassicTime and isc:BaseJurassicUncertainty.

Table 1
Core ontologies in the existing deep time knowledge graph (Cox and Richard, 2015; Cox, 2016).

Prefix	Namespace	Role in the deep time knowledge graph
dc	<http://purl.org/dc/elements/1.1/>	Specify metadata of vocabulary schemes and concepts
dcterms	<http://purl.org/dc/terms/>	Specify metadata of vocabulary schemes and concepts
geo	<http://www.opengis.net/ont/geosparql#>	Specify the location of golden spikes
gts	<http://resource.geosciml.org/ontology/timescale/gts#>	Based on THORS and ISO 19156; Specify the structure of core classes and relationships in the geological time scale
isc	<http://resource.geosciml.org/classifier/ics/ischart/>	Specify the deep time concepts in the ISC charts
sf	<http://www.opengis.net/ont/sf#>	Specify spatial feature types
skos	<http://www.w3.org/2004/02/skos/core#>	Specify hierarchical structure and multilingual labels of deep time concepts
thors	<http://resource.geosciml.org/ontology/timescale/thors#>	Based on ISO 19108; Specify the temporal hierarchical ordinal reference system of deep time concepts
tm	<http://def.seegrid.csiro.au/isotc211/iso 19108/2002/temporal#>	Based on ISO 19108; Specify temporal objects and the reference system
time	<http://www.w3.org/2006/time#>	Specify the reference system and topological relationships of deep time concepts
ts	<http://resource.geosciml.org/vocabulary/timescale/>	Specify the different versions of vocabulary schemes for the ISC charts

2.2. A new structure to represent versions of concepts, attributes and relationships

To address the above-mentioned issues of version control in the existing deep time knowledge graph, we propose a new structure to represent versions of deep time concepts as well as their attributes and inter-relationships. Our aim is not to redesign the wheel, so we have applied two guidelines in our work: 1) Do not make changes to the ontologies and 2) Reuse the existing vocabularies as much as possible. Our design is based on the latest version of the vocabulary, i.e. version 2017 in Cox (2019), as it has adopted the state-of-the-art community standards, such as GeoSPARQL and Time Ontology.

We categorize the changes and updates in the ISC charts into several types. First, the changes in concepts. These include two topics. a) New time concepts based on updated classification at a certain part of the geologic time scale. For example, the three Age/Stage concepts Greenlandian, Northgrippian, and Meghalayan first appeared under Holocene in the 2018–07 version of the ISC chart. b) New Golden Spikes (i.e., ratified Global Boundary Stratotype Sections and Points). For example, the golden spike at the base of Jiangshanian first appeared in the 2012 version of the ISC chart. Second, the changes in attributes. These include updates to the label of a concept, the number of age at a boundary and its uncertainty. For example, the label for the Age/Stage concept Stage 10 was changed from "Stage 10" in the 2010 ISC chart to "Jiangshanian" in the 2012 ISC chart when the golden spike at the base of it was formally ratified. For updates of numbers, a good example is shown in Fig. 2. Third, changes in the relationship between concepts (i.e., topology). For example, the removal, return, and updates of Quaternary and its relationships with other concepts under Cenozoic across several ISC charts. Our investigation of the ISC charts between 2004 and 2018 shows that the majority of changes are in the attributes, following by changes in concepts and relationships. The target of this research is a comprehensive structure to capture the version information of all those types of changes and updates.

The core idea in our structure for version control is using the URIs of vocabulary schemes as version identifiers in the specification of concepts, attributes and relationships. The diagrams in Fig. 3 depict a comparison between the original and the updated structures in the specification of Jurassic and its attributes. The four small diagrams on the left side of Fig. 3 are corresponding to the code shown in Fig. 2, which are taken from the existing vocabularies. The single diagram on the right side of Fig. 3 shows the updated structure. The major change here is to add the corresponding vocabulary scheme URI to each version of the attribute. In the original structure, the version information of the attributes is implicit, although the version of the concept Jurassic is

Jurassic in 2004 vocabulary scheme

```
isc:Jurassic
    thors:begin isc:BaseJurassic ;
    skos:inScheme
    <http://resource.geosciml.org/classifierscheme/
    ics/2004/ischart> .

isc:BaseJurassic
    tm:temporalPosition isc:BaseJurassicPosition .

isc:BaseJurassicPosition
    tm:value "199.6"^^xsd:float ;
    thors:positionalUncertainty
    isc:BaseJurassicUncertainty .

isc:BaseJurassicUncertainty
    basic:value "0.6"^^xsd:float .
```

Jurassic in 2010 vocabulary scheme

```
isc:Jurassic
    thors:begin isc:BaseJurassic ;
    skos:inScheme
    <http://resource.geosciml.org/classifierscheme/
    ics/2010/ischart> .

isc:BaseJurassic
    tm:temporalPosition isc:BaseJurassicTime .

isc:BaseJurassicTime
    tm:value "199.6"^^xsd:float ;
    thors:positionalUncertainty
    isc:BaseJurassicUncertainty .

isc:BaseJurassicUncertainty
    basic:value "0.6"^^xsd:float .
```

Jurassic in 2012 vocabulary scheme

```
isc:Jurassic
    thors:begin isc:BaseJurassic ;
    skos:inScheme
    <http://resource.geosciml.org/classifierscheme/
    ics/ischart/2012> .

isc:BaseJurassic
    tm:temporalPosition isc:BaseJurassicTime .

isc:BaseJurassicTime
    tm:value "201.3"^^xsd:float ;
    thors:positionalUncertainty
    isc:BaseJurassicUncertainty .

isc:BaseJurassicUncertainty
    basic:value "0.2"^^xsd:float .
```

Jurassic in 2017 vocabulary scheme

```
isc:Jurassic
    time:hasBeginning isc:BaseJurassic ;
    skos:inScheme ts:isc2017.

isc:BaseJurassic
     time:inTemporalPosition isc:BaseJurassicTime .

isc:BaseJurassicTime
     time:numericPosition 201.3 ;
     gts:positionalUncertainty
     isc:BaseJurassicUncertainty .

isc:BaseJurassicUncertainty
     time:numericDuration 0.2 .
```

Fig. 2. Specification of Jurassic and its base boundary age in four different vocabulary schemes in the deep time knowledge graph (code from Cox, 2019).

specified. In particular, when a user queries across those different versions of vocabularies, a very careful design is needed to mark which BaseJurassicTime and BaseJurassicUncertainty values are from which vocabulary scheme. In the updated structure, the version information is made explicit by adding the vocabulary scheme URIs in the specification of attributes. Fig. 4 shows the code corresponding to BaseJurassicTime and BaseJurassicUncertainty in Fig. 3 for their attribute versioning. It is noteworthy that Fig. 4 also includes other vocabulary scheme URIs that are not depicted in Fig. 3. Moreover, although BaseJurassic is just a bridging concept between Jurassic and BaseJurassicTime, we also record version information for it, i.e. a detailed list of URIs to show in which vocabulary schemes the concept BaseJurassic exists. Given a specified vocabulary scheme URI (e.g., ts:isc 2017-02), the value and uncertainty at the base of Jurassic can be retrieved by tracing Jurassic, BaseJurassic, BaseJurassicTime, and BaseJurassicUncertainty in a sequence.

Besides version control for deep time concepts and attributers, we are able to use the vocabulary scheme URIs to mark the version of relationships between concepts. Figs. 5 and 6 show how the changes in the relationships between Quaternary and other concepts are captured in the new knowledge graph. Fig. 6 show a part of the code for Quaternary in the knowledge graph. The specified relationships are based on the W3C Simple Knowledge Organization System (SKOS) (Miles and Bechhofer, 2009). SKOS uses simple relationships such as broader, narrower, broaderTransative, and narrowerTransative to specify the hierarchical structure of the geologic time scale. In the complete code of the knowledge graph, there are also precise specification of the relationships by using the Time Ontology (Cox and Little, 2020). Interested readers can check the source code through the link given in the section Computer Code Availability. Based on this new structure for version control of concepts, attributes and relationships, we are able to represent all versions of ISC charts in a single knowledge graph, and make it easier to track the version history.

2.3. Consistency with existing ontologies and vocabularies

The approach of using vocabulary scheme URIs in the specification of concepts, attributes and topological relationships is simple for implementation. Our work reused the existing ontologies and did not make any changes. All the versions of the ISC charts were represented in a single knowledge graph. The URI of each time concept or attribute remains the same as in the existing vocabularies, such as isc:Jurassic, isc:BaseJurassicTime, and isc:BaseJurassicUncertainty. To have a precise and consistent representation of all the versions of the ISC charts, we designed a new vocabulary scheme URI for each ISC chart (Table 2). Note that those new scheme URIs were used in our knowledge graph result for tracking versions of concepts, attributes and relationships. We did not create a separate vocabulary for each scheme URI. Instead, in our result we used owl:sameAs to set up a mapping back to the vocabulary schemes in the existing knowledge graph (Cox, 2019) (Table 2).

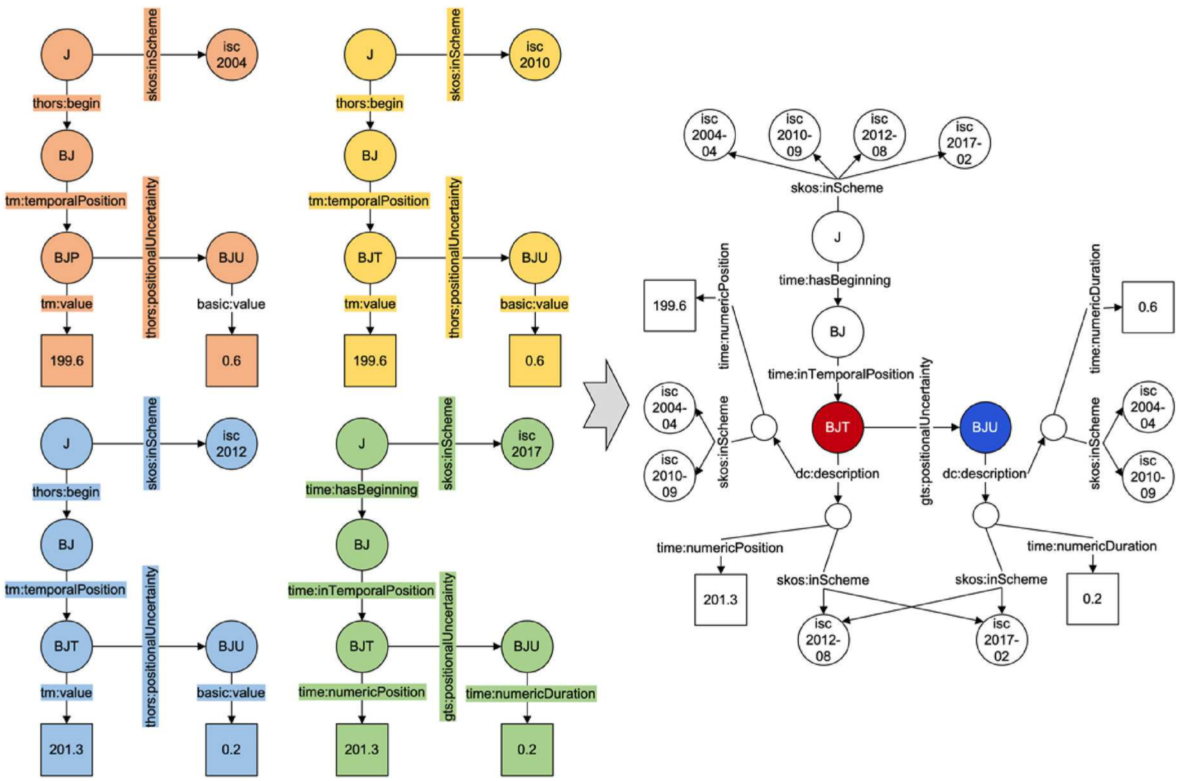

Fig. 3. A new structure for version control of both concepts and attributes. Here Jurassic and the age at the base boundary of Jurassic are used as an example. The four small diagrams on the left are corresponding to the code shown in Fig. 2, following the same color scheme. Abbreviations: isc 2004–04/2010–09/2012–08/2017–02 – vocabulary scheme for the ISC 2004–04/2010–09/2012–08/2017-02 chart, J – Jurassic, BJ – BaseJurassic, BJP/T – BaseJurassicPosition/Time, BJU – BaseJurassicUncertainty. For the diagram of the new version control structure on the right side, BJT and BJU are highlighted because Fig. 4 will show the corresponding code for them, with the same color scheme. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Version control for isc:BaseJurassicTime

```
isc:BaseJurassicTime
    dc:description
      [
        time:numericPosition "201.3"^^xsd:decimal ;
        skos:inScheme ts:isc2018-08, ts:isc2018-07,
        ts:isc2017-02, ts:isc2016-10, ts:isc2016-04,
        ts:isc2015-01, ts:isc2014-10, ts:isc2014-02,
        ts:isc2013-01, ts:isc2012-08
      ] ,

      [
        time:numericPosition "199.6"^^xsd:decimal ;
        skos:inScheme ts:isc2010-09, ts:isc2009-08,
        ts:isc2008-08, ts:isc2006-04, ts:isc2005-12,
        ts:isc2004-04
      ] .
```

Version control for isc:BaseJurassicUncertainty

```
isc:BaseJurassicUncertainty
    dc:description
      [
        time:numericDuration "0.2"^^xsd:decimal ;
        skos:inScheme ts:isc2018-08, ts:isc2018-07,
        ts:isc2017-02, ts:isc2016-10, ts:isc2016-04,
        ts:isc2015-01, ts:isc2014-10, ts:isc2014-02,
        ts:isc2013-01, ts:isc2012-08
      ] ,

      [
        time:numericDuration "0.6"^^xsd:decimal ;
        skos:inScheme ts:isc2010-09, ts:isc2009-08,
        ts:isc2008-08, ts:isc2006-04, ts:isc2005-12,
        ts:isc2004-04
      ] .
```

Fig. 4. Specification of version control for part of the attributes of Jurassic in the updated deep time knowledge graph. The code is corresponding to the lower part of the diagram at the right side of Fig. 3. Besides the 2010 to 2017 vocabulary schemes, URIs of other vocabulary schemes are also included in the code. The vocabulary scheme URIs in the code include both year and month because for a few years there are more than one released version of the ISC chart.

3. The resulting knowledge graph and experiments with it

This section will first introduce the resulting deep time knowledge graph with the new structure of version control, and then present several SPARQL query examples enabled by this new structure.

3.1. A comprehensive deep time knowledge graph

In the vocabularies developed by Cox (2019), detailed metadata were given to each vocabulary scheme. We retained those metadata in the new knowledge graph to give accreditation to the authors of those existing vocabularies. Then, we implemented the above-mentioned structure for version control, and built a single knowledge graph that includes all the versions of the ISC chart from 2004 to 2018. The source

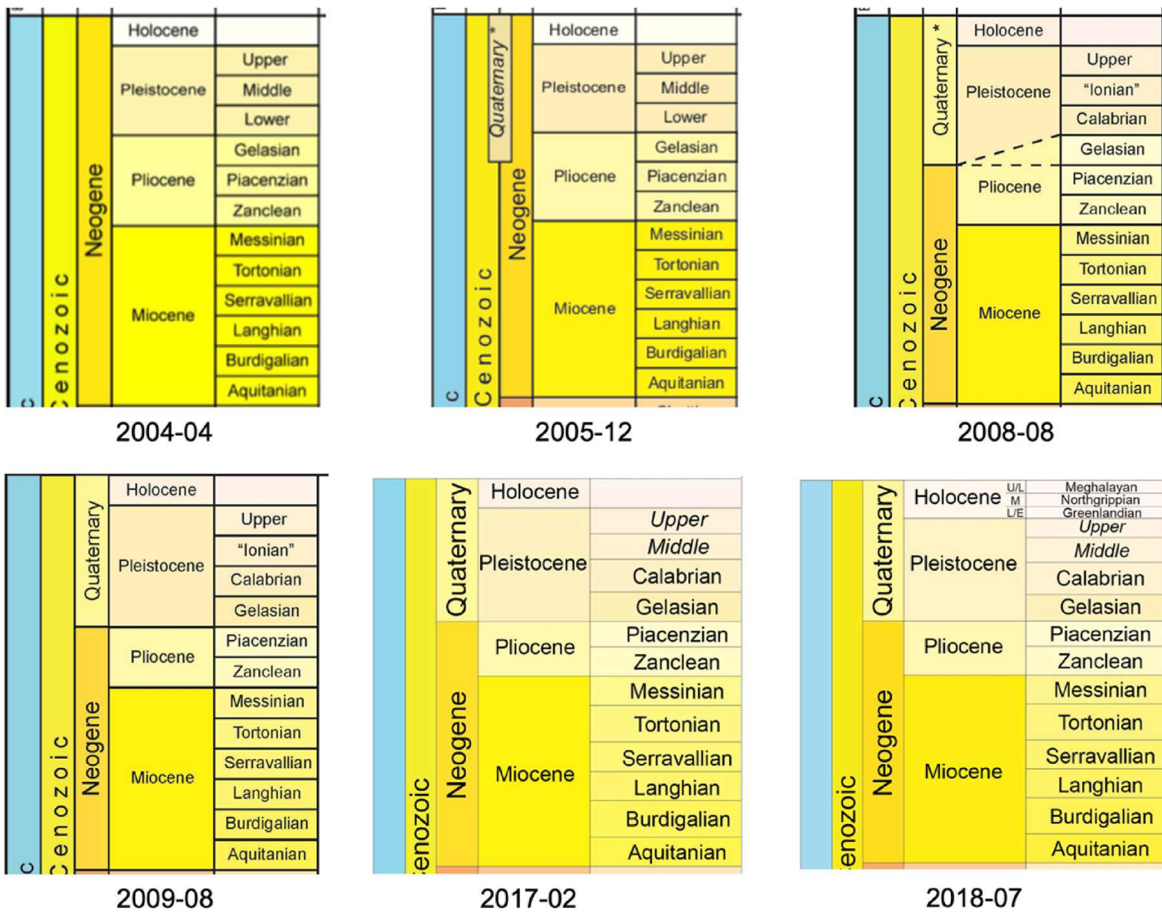

Fig. 5. Changes on Quaternary and its relationships with other concepts in a few selected ISC charts between 2004 and 2018. Several other charts are not shown here because there are no changes on Quaternary in them.

```
isc:Quaternary
   dc:description
     [
        skos:broader isc:Cenozoic ;
        skos:broaderTransitive isc:Phanerozoic ;
        skos:narrower isc:Holocene ;
        skos:narrower isc:Pleistocene ;
        skos:narrowerTransitive isc:Calabrian ;
        skos:narrowerTransitive isc:Gelasian ;
        skos:narrowerTransitive isc:MiddlePleistocene ;
        skos:narrowerTransitive isc:UpperPleistocene ;
        skos:narrowerTransitive isc:Greenlandian ;
        skos:narrowerTransitive isc:Northgrippian ;
        skos:narrowerTransitive isc:Meghalayan ;
        skos:inScheme ts:isc2018-08, ts:isc2018-07
     ] ,

     [
        skos:broader isc:Cenozoic ;
        skos:broaderTransitive isc:Phanerozoic ;
        skos:narrower isc:Holocene ;
        skos:narrower isc:Pleistocene ;
        skos:narrowerTransitive isc:Calabrian ;
        skos:narrowerTransitive isc:Gelasian ;
# code continues in the right column
        skos:narrowerTransitive
        isc:MiddlePleistocene ;
        skos:narrowerTransitive
        isc:UpperPleistocene ;
        skos:inScheme ts:isc2017-02, ts:isc2016-10,
        ts:isc2016-04, ts:isc2015-01, ts:isc2014-10,
        ts:isc2014-02, ts:isc2013-01, ts:isc2012-08,
        ts:isc2010-09, ts:isc2009-08, ts:isc2008-08
     ] ,

     [
        skos:broader isc:Cenozoic ;
        skos:broaderTransitive isc:Phanerozoic ;
        skos:narrower isc:Holocene ;
        skos:narrower isc:Pleistocene ;
        skos:narrowerTransitive isc:Calabrian ;
        skos:narrowerTransitive isc:Gelasian ;
        skos:narrowerTransitive
        isc:MiddlePleistocene ;
        skos:narrowerTransitive
        isc:UpperPleistocene ;
        rdfs:comment "Proposed by ICS"@en ;
        skos:inScheme ts:isc2006-04, ts:isc2005-12
     ] .
```

Fig. 6. A partial list of the code for specifying the changed relationships between Quaternary and other concepts. The changes shown in Fig. 5 are represented in the code here. For the complete code of Quaternary in the knowledge graph please check the link in the section Computer Code Availability at the end of the paper.

Table 2
Mapping between vocabulary scheme URIs by using owl:sameAs.

Scheme URIs in new vocabulary	Scheme URIs of existing vocabularies
ts:isc 2018-08	<http://resource.geosciml.org/classifierscheme/ics/2018/ischart>
ts:isc 2018-07	<http://resource.geosciml.org/classifierscheme/ics/2018/ischart>
ts:isc 2017-02	<http://resource.geosciml.org/classifierscheme/ics/2017/ischart>
ts:isc 2016-10	<http://resource.geosciml.org/classifierscheme/ics/2016/ischart>
ts:isc 2016-04	<http://resource.geosciml.org/classifierscheme/ics/2016/ischart>
ts:isc 2015-01	<http://resource.geosciml.org/classifierscheme/ics/2015/ischart>
ts:isc 2014-10	<http://resource.geosciml.org/classifierscheme/ics/2014/ischart>
ts:isc 2014-02	<http://resource.geosciml.org/classifierscheme/ics/2014/ischart>
ts:isc 2013-01	<http://resource.geosciml.org/classifierscheme/ics/2013/ischart>
ts:isc 2012-08	<http://resource.geosciml.org/classifierscheme/ics/2012/ischart>
ts:isc 2010-09	<http://resource.geosciml.org/classifierscheme/ics/2010/ischart>
ts:isc 2009-08	<http://resource.geosciml.org/classifierscheme/ics/2009/ischart>
ts:isc 2008-08	<http://resource.geosciml.org/classifierscheme/ics/2008/ischart>
ts:isc 2006-04	<http://resource.geosciml.org/classifierscheme/ics/2006/ischart>
ts:isc 2005-12	<http://resource.geosciml.org/classifierscheme/ics/2005/ischart>
ts:isc 2004-04	<http://resource.geosciml.org/classifierscheme/ics/2004/ischart>

code of the resulting vocabulary was serialized in Turtle format and was made open on GitHub. We also set up an open SPARQL endpoint to test the functionality of the vocabulary for version tracking. The section 'Computer Code Availability' at the end of this paper lists the web addresses to access the GitHub repository and the SPARQL endpoint. The next sections will demonstrate a few example studies we conducted with the established endpoint.

3.2. *Experiments of querying versioned information*

We have implemented several use cases to query the knowledge graph built in this study. Those use cases follow the intuition of a geoscientist to query different types of information. The first use case is to find the age and uncertainty at the base boundary of a given interval concept (e.g., Jurassic) in a given version of the ISC chart (e.g., isc 2012-08). By using the pattern of the version control structure, we implemented the code in Fig. 7 and obtained the information we wanted to know about Jurassic. Three variables were used: ?tconcept for the time concept Jurassic, ?basetimevalue for the age value at the base of Jurassic, and ?btuncertvalue for the uncertainty of the age. The code can be easily adapted to query other concepts in different versions of the ISC chart. Only the label of the concept and the URI of the vocabulary scheme need to be updated in a new query.

Our second use case was an extension to the first use case to make it robust to tackle some unique situations. When we began to query other concepts, we found a few unique patterns. For example, the ages at the base of some concepts have no uncertainty, such as Holocene. Another interesting concept is Archean. In some versions of the ISC chart, the base of Archean has no numerical age, but just a nominal description. Accordingly, in the SPARQL query we used an OPTIONAL clause to tackle the issue caused by nominal description of base age, and a UNION clause to tackle the issue of a blank uncertainty value. Fig. 8 shows the resulting SPARQL code. The label "Archean" in the code can be replaced by "Holocene" to see the example of blank uncertainty.

In the third use case we aimed to retrieve the whole version history of age and uncertainty at the base of a concept across all the vocabulary schemes. Interestingly, we found that only by doing some minor changes to the SPARQL code in the second use case we could achieve the goal. Fig. 9 takes Triassic as an example. Compared to the code in Fig. 8, the major change was using a variable ?schemeid to replace the fixed vocabulary scheme ts:isc 2004-04. At the end of the code we used an ORDER BY DESC (?schemeid) clause to list the results in a reverse chronological order (i.e. from the latest to the earliest).

We have run several other use cases in Jupyter Notebook (with R kernel). One is about retrieving the attributes of a concept from the latest vocabulary scheme, as many users just want to see records in the latest version. This was realized by conducting two queries. The first is to retrieve the URI for the latest vocabulary scheme. Once that URI is known, the second query is to find the attributes of a given concept in that specific vocabulary scheme. The link to the code of this use case is accessible at Ma (2020a). We also achieved the same result with a nested SPARQL query, but it took much longer time comparing with the two-query approach. Another use case is about finding all sub-/-super-concepts of a given concept. This was achieved by using the skos:narrower, skos:narrower Transitive, skos:broader, and skos:broader Transitive statements in the knowledge graph, as well as properties from the Time Ontology. The link this use case is accessible at Ma (2020b). All the use cases developed in this research are shared on GitHub and a link to the list is given in the section Computer Code Availability at the end of this paper.

4. Discussion

The ISC chart released by the International Commission on Stratigraphy is a widely used reference for global geologic time scale and concepts. The chart is under continuous maintenance and update, so there have been many versions of it since the early 2000s. Documenting the detailed version history of the ISC chart in a knowledge graph can make it a comprehensive reference for many geoscience researchers. However, version control is a less-discussed topic in the existing research of deep time knowledge graph. This study designs and a new structure for the version control of deep time concepts, attributes and relationships, and implements it in a knowledge graph to represent and track the version history of the ISC chart between 2004 and 2018. In the work we made no changes to the existing ontologies of geologic time (Cox and Richard, 2015), and we reused the existing vocabulary schemes (Cox, 2019). Those operations made our result fully consistent with the existing ontologies and vocabularies. A service for the knowledge graph has been set up. Our experimental results with the service website show that it is convenient for tracking version information of deep time concepts, attributes and relationships. The version control structure established in the deep time knowledge graph and the service have established a reliable source to retrieve the precise meaning of a deep time concept in a certain context. The disambiguation is a big advantage for multi-source deep time data integration, where heterogeneous deep time concepts are used in the data records.

The ontologies and vocabularies presented in Cox and Richard (2015) and Cox (2019) demonstrate the latest progress on a machine-readable deep time knowledge graph. In their current work, a separate vocabulary was built for each version of the ISC chart, together with a SPARQL endpoint for it. In the vocabulary service based on SISSVoc, the query to a time concept will automatically redirect to the information of the concept in the latest version of the vocabulary. This is a novel architecture for vocabulary service. Yet, we can see that it is a tedious job to build a separate vocabulary and endpoint for each version of the ISC chart. Recently, the Time Ontology and GeoSPARQL was incorporated into the ontologies for the deep time knowledge graph (Cox, 2016, 2019). A few vocabularies for the latest ISC charts applied the new ontologies, but many vocabularies created earlier were still

SPARQL query code

```
prefix dc: <http://purl.org/dc/elements/1.1/>
prefix gts: <http://resource.geosciml.org/ontology/timescale/gts#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix time: <http://www.w3.org/2006/time#>
prefix ts: <http://resource.geosciml.org/vocabulary/timescale/>

SELECT ?tconcept ?basetimevalue ?btuncertvalue
WHERE
{
 GRAPH <http://deeptimekb.org/iscallnew>
  {
      ?tconcept skos:prefLabel "Jurassic"@en ;
                dc:description
                  [
                    time:hasBeginning ?baseboundary ;
                    skos:inScheme ts:isc2012-08
                  ] .
      ?baseboundary time:inTemporalPosition ?basetime .
      ?basetime dc:description
             [
               time:numericPosition ?basetimevalue ;
               skos:inScheme ts:isc2012-08
             ] ;
      gts:positionalUncertainty ?basetimeuncert .
      ?basetimeuncert   dc:description
             [
               time:numericDuration ?btuncertvalue ;
               skos:inScheme ts:isc2012-08
             ] .
  }
}
```

Result

tconcept	basetimevalue	btuncertvalue
http://resource.geosciml.org/classifier/ics/ischart/Jurassic	201.3	0.2

Fig. 7. SPARQL query to find the age and uncertainty at the base of Jurassic in the ISC chart 2012–08.

based on the original ontologies (Cox, 2019). These different ontology patterns cause challenges for queries across the vocabulary schemes. For example, if we want to design a federated SPARQL query to search the version history of Jurassic across the endpoints of those separate vocabularies, we should arrange the SPARQL code according to the ontology patterns in each vocabulary, which is a hard job even for an expert of SPARQL.

Our work addresses the insufficiency of version control in the current deep time knowledge graph. We used identifiers of the vocabulary schemes in the specification of concepts, attributes and topological relationships. In other words, we treated each vocabulary scheme as a node in the deep time knowledge graph rather than a separate vocabulary. Each vocabulary scheme identifier represents a version of the ISC chart. This structure is simple but functional for version control and tracking. As demonstrated in Section 2, this new structure of version control is realized by using skos:inScheme to specify the vocabulary scheme in which a certain concept, attribute or topological relationship appears. In the vocabularies of Cox (2019), such a structure already existed in the specification of time concepts, but not for the attributes and topological relationships. Through the new work of this research, we were able to represent the version history across many ISC charts (2004–2018) in a single knowledge graph. The ontologies we used were the latest from Cox (2019), which made the work consistent with existing community standards. We also retained the metadata of vocabulary schemes to give full accreditations to the original authors. The results of version tracking experiments prove the functionality of the structure developed in this study.

The designed version control structure can be reused for digital records of many other objects. We have identified three categories of changes in the ISC charts: concept, attribute of concept, and relationship between concept. The changes at those three levels may also exist in the digital records of datasets, samples, software programs, and more. Taking dataset as an example, the issue of version control for dataset is written as a section in the recent W3C Data Catalog Vocabulary (DCAT) (Albertoni et al., 2020), and even the versioning of relationships between datasets is mentioned. Nevertheless, the technical approach is not specified in detail in that section. The real-world scientific practices can bring many complicated and interesting situations for the research of version control. For example, originally a dataset was derived from a sample, and now a new test (i.e. new attributes) has been taken with the original sample and a few new samples, and a new researcher has joined the team for the data collection. At least, the version control structure proposed in this paper will be able to provide an initial thought and a start point to tackle the need of version control in that example.

We are now extending the knowledge graph to incorporate two new versions of the ISC chart released in 2019 and 2020. For a new version of the ISC chart, now there is no need to build a separate vocabulary. Instead, we just need to create a vocabulary scheme identifier for it, and then use the identifier to renew the version information of concepts, attributes and topological relationships inside the knowledge graph. Once the work is complete, we will update the knowledge graph code and metadata description on our GitHub repository (see the section Computer Code Availability at the end of the paper). We will observe future versions of the ISC chart to keep the knowledge graph and the SPARQL endpoint up-to-date. Another work we are developing now is a visualization of the knowledge graph, which will incorporate the SPARQL code we have developed in the use cases. The visualization will help set up graphic user interface to query the knowledge graph, where no SPARQL coding is needed for the end user. This will make it easy for geoscientists to access the knowledge graph, especially the version

SPARQL query code

```
prefix dc: <http://purl.org/dc/elements/1.1/>
prefix gts: <http://resource.geosciml.org/ontology/timescale/gts#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix time: <http://www.w3.org/2006/time#>
prefix ts: <http://resource.geosciml.org/vocabulary/timescale/>

SELECT ?tconcept ?basetimevalue ?btuncertvalue
WHERE
{
 GRAPH <http://deeptimekb.org/iscallnew>
  {
      ?tconcept skos:prefLabel "Archean"@en ;
               dc:description
                 [
                   time:hasBeginning ?baseboundary ;
                   skos:inScheme ts:isc2004-04
                 ] .
      ?baseboundary time:inTemporalPosition ?basetime .
      {
         ?basetime dc:description
                [
                  time:numericPosition ?basetimevalue ;
                  skos:inScheme ts:isc2004-04
                ] .
      }
      UNION
      {
         ?basetime dc:description
                [
                  time:nominalPosition ?basetimevalue ;
                  skos:inScheme ts:isc2004-04
                ] .
      }

      OPTIONAL
      {
        ?basetime gts:positionalUncertainty ?basetimeuncert .
        ?basetimeuncert   dc:description
                [
                  time:numericDuration ?btuncertvalue ;
                  skos:inScheme ts:isc2004-04
                ] .
      }
  }
}
```

Result

tconcept	basetimevalue	btuncertvalue
http://resource.geosciml.org/classifier/ics/ischart/Archean	"Not defined"^^<http://www.w3.org/2001/XMLSchema#string>	

Result if the label in the SPARQL code is changed from "Archean" to "Holocene"

tconcept	basetimevalue	btuncertvalue
http://resource.geosciml.org/classifier/ics/ischart/Holocene	0.0115	

Fig. 8. SPARQL query to find the age and uncertainty at the base of Archean and Holocene in the ISC chart 2004–04.

information (cf. Ma et al., 2012; Wang et al., 2018).

We also plan several works for the near future. The first is to build vocabularies for regional and local standards of geologic time concepts, such as those listed in Haq (2007). We will reuse ontologies in the existing deep time knowledge graph to make the result consistent within the current framework. The second is to use the topology of deep time concepts and semantic reasoning in deep time researches. The incorporation of Time Ontology in the deep time knowledge graph make the topology of deep time concepts explicit. We can use the hierarchical ordinal relationships to address many challenges in data-intensive deep time research, such as concept mapping among heterogeneous databases, precise deep time data retrieval, and cross-disciplinary deep time data integration and synthesis. A few initial ideas have been explored in one of our previous studies, such as using knowledge graph for geologic map generalization (Ma, 2017). The third future work is to refine the service of the deep time knowledge graph. Besides the visualization and graphic user interface, we will also develop an R package to allow users query the knowledge graph from a workflow environment, such as R MarkDown or Jupyter Notebook.

5. Conclusions

Machine-readable ontologies, vocabularies and knowledge graphs are a useful method to promote data interoperability. In geoscience, the deep time knowledge graph has received a lot of discussion and developments in the past decades. This paper focuses on a small topic in

SPARQL query code

```
prefix dc: <http://purl.org/dc/elements/1.1/>
prefix gts: <http://resource.geosciml.org/ontology/timescale/gts#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
prefix time: <http://www.w3.org/2006/time#>
prefix ts: <http://resource.geosciml.org/vocabulary/timescale/>

SELECT ?tconcept ?basetimevalue ?btuncertvalue ?schemeid
WHERE
{
 GRAPH <http://deeptimekb.org/iscallnew>
  {
      ?tconcept skos:prefLabel "Triassic"@en ;
              dc:description
                [
                  time:hasBeginning ?baseboundary ;
                ] .
      ?baseboundary time:inTemporalPosition ?basetime .
      {
          ?basetime dc:description
                [
                  time:numericPosition ?basetimevalue ;
                  skos:inScheme ?schemeid
                ] .
      }
      UNION
      {
          ?basetime dc:description
                [
                  time:nominalPosition ?basetimevalue ;
                  skos:inScheme ?schemeid
                ] .
      }

      OPTIONAL
      {
        ?basetime gts:positionalUncertainty ?basetimeuncert .
        ?basetimeuncert   dc:description
                [
                  time:numericDuration ?btuncertvalue ;
                  skos:inScheme ?schemeid
                ] .
      }
  }
}
```

Fig. 9. SPARQL query to find the whole version history for the age and uncertainty at the base of Triassic.

the deep time knowledge graph: how to realize version control for concepts, attributes and topological relationships. By using the identifiers of vocabulary schemes, we designed a new structure to represent the version history across many versions of the ISC chart. Our aim is not to redesign the wheel of the deep time knowledge graph. Instead, we have built our structure on the top of existing ontologies and vocabularies to make the result fully consistent with the current framework. We have communicated the work with the Semantic Technologies Committee in the Earth Science Information Partners, and will promote collaboration on the deep time knowledge graph. The focus of this study is the version control and tracking of deep time knowledge graph, but the underlying methods are applicable to a variety of knowledge graphs in geoscience and other disciplines, such as SWEET (Raskin and Pan, 2005) and the Open Knowledge Network (Baru, 2018; Guha and Moore, 2016).

6. Computer Code Availability

The source code of the deep time knowledge graph developed in this work is made open and accessible on GitHub. 1) Name: A comprehensive knowledge graph for all versions of the international chronostratigraphic chart; 2) Version: V1.0; 3) License: All code is made open source under the MIT license; 4) Link to the knowledge graph code: https://github.com/xgmachina/geotimeversion. A SPARQL endpoint has been set up for the knowledge graph at http://virtuoso.nkn.uidaho.edu:8890/sparql/using the graph name http://deeptimekb.org/iscallnew. 5) Link to use cases of Jupyter Notebooks (with R kernel) querying the SPARQL endpoint: https://github.com/xgmachina/DeepTimeKB/tree/master/Notebooks.

Author contribution statement

Author Contribution Statement: X.M. designed the method. X.M. led the coding of the knowledge graph with assistance from C.M. and C.W. All authors contributed to the manuscript writing.

```
ORDER BY DESC (?schemeid)
```

Result

tconcept	basetimevalue	btuncertvalue	schemeid
http://resource.geosciml.org/classifier/ics/ischart/Triassic	251.902	0.024	http://resource.geosciml.org/vocabulary/timescale/isc2018-08
http://resource.geosciml.org/classifier/ics/ischart/Triassic	251.902	0.024	http://resource.geosciml.org/vocabulary/timescale/isc2018-07
http://resource.geosciml.org/classifier/ics/ischart/Triassic	251.902	0.024	http://resource.geosciml.org/vocabulary/timescale/isc2017-02
http://resource.geosciml.org/classifier/ics/ischart/Triassic	252.17	0.06	http://resource.geosciml.org/vocabulary/timescale/isc2016-10
http://resource.geosciml.org/classifier/ics/ischart/Triassic	252.17	0.06	http://resource.geosciml.org/vocabulary/timescale/isc2016-04
http://resource.geosciml.org/classifier/ics/ischart/Triassic	252.17	0.06	http://resource.geosciml.org/vocabulary/timescale/isc2015-01
http://resource.geosciml.org/classifier/ics/ischart/Triassic	252.17	0.06	http://resource.geosciml.org/vocabulary/timescale/isc2014-10
http://resource.geosciml.org/classifier/ics/ischart/Triassic	252.17	0.06	http://resource.geosciml.org/vocabulary/timescale/isc2014-02
http://resource.geosciml.org/classifier/ics/ischart/Triassic	252.17	0.06	http://resource.geosciml.org/vocabulary/timescale/isc2013-01
http://resource.geosciml.org/classifier/ics/ischart/Triassic	252.2	0.5	http://resource.geosciml.org/vocabulary/timescale/isc2012-08
http://resource.geosciml.org/classifier/ics/ischart/Triassic	251	0.4	http://resource.geosciml.org/vocabulary/timescale/isc2010-09
http://resource.geosciml.org/classifier/ics/ischart/Triassic	251	0.4	http://resource.geosciml.org/vocabulary/timescale/isc2009-08
http://resource.geosciml.org/classifier/ics/ischart/Triassic	251	0.4	http://resource.geosciml.org/vocabulary/timescale/isc2008-08
http://resource.geosciml.org/classifier/ics/ischart/Triassic	251	0.4	http://resource.geosciml.org/vocabulary/timescale/isc2006-04
http://resource.geosciml.org/classifier/ics/ischart/Triassic	251	0.4	http://resource.geosciml.org/vocabulary/timescale/isc2005-12
http://resource.geosciml.org/classifier/ics/ischart/Triassic	251	0.4	http://resource.geosciml.org/vocabulary/timescale/isc2004-04

Fig. 9. (*continued*).

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.

Acknowledgements

The work presented in this paper was supported by the National Science Foundation, United States (No. 1835717). Additional support was provided by the IUGS Deep-time Digital Earth (DDE) Big Science Program, the Deep Carbon Observatory, the Alfred P. Sloan Foundation, and the Carnegie Institution for Science for communicating the research progress at several workshops and meetings. We thank the Computers & Geoscience editor and four anonymous reviewers for their constructive comments on an earlier version of the manuscript.

References

4D Initiative Team, 2018. White paper of the 4D initiative: deep-time data driven discovery. https://4d.carnegiescience.edu/sites/default/files/4D_materials/4D_WhitePaper.pdf. (Accessed 4 March 2020).

Albertoni, R., Browning, D., Cox, S., Beltran, A.G., Perego, A., Winstanley, P., Maali, F., Erickson, J., 2020. Data Catalog Vocabulary (DCAT) -Version 2. https://www.w3.org/TR/vocab-dcat-2/. (Accessed 26 August 2020).

Baru, C., 2018. Perspectives on Open Knowledge Networks. The First U.S. Semantic Technologies Symposium (US2TS), Dayton, OH, USA. Oral Presentation.

Berners-Lee, T., 2000. Semantic web on XML. In: Presentation at XML 2000 Conference, Washington DC. http://www.w3.org/2000/Talks/1206-xml2k-tbl. (Accessed 12 July 2017).

Bishr, Y., 1998. Overcoming the semantic and other barriers to GIS interoperability. Int. J. Geogr. Inf. Sci. 12 (4), 299–314.

Brodaric, B., 2007. Geo-pragmatics for the geospatial semantic web. Trans. GIS 11 (3), 453–477.

Brodaric, B., 2018. Interoperability of representations. In: Richardson, D., Castree, N., Goodchild, M.F., Kobayashi, A., Liu, W., Marston, R.A. (Eds.), The International Encyclopedia of Geography. John Wiley & Sons. https://doi.org/10.1002/9781118786352.wbieg0894.pub2.

Cohen, K.M., Finney, S.C., Gibbard, P.L., Fan, J.-X., 2013. The ICS international chronostratigraphic chart. Episodes 36, 199–204.

Cox, S.J., 2016. Time ontology extended for non-Gregorian calendar applications. Semantic Web 7 (2), 201–209.

Cox, S.J., 2019. Vocabulary files for the international stratigraphic chart. http://resource.geosciml.org/vocabulary/timescale/. (Accessed 21 February 2020).

Cox, S., Little, C., 2020. Time ontology in OWL. https://www.w3.org/TR/owl-time/. (Accessed 30 July 2020).

Cox, S.J., Richard, S.M., 2005. A formal model for the geologic time scale and global stratotype section and point, compatible with geospatial information transfer standards. Geosphere 1 (3), 119–137.

Cox, S.J., Richard, S.M., 2015. A geologic timescale ontology and service. Earth Sci. India 8 (1), 5–19.

Cox, S.J., Yu, J., Rankine, T., 2016. SISSVoc: a Linked Data API for access to SKOS vocabularies. Semantic Web 7 (1), 9–24.

Cutcher-Gershenfeld, J., Baker, K.S., Berente, N., Flint, C., Gershenfeld, G., Grant, B., Haberman, M., King, J.L., Kickpatrick, C., Lawrence, B., Lewis, S., Lenhardt, W.C., Mayernik, M., McElroy, C., Mittleman, B., Shin, N., Stall, S., Winter, S., Zaslavsky, I., 2017. Five ways consortia can catalyse open science. Nature 543 (7647), 615–617.

Fox, P., Hendler, J., 2014. The science of data science. Big Data 2 (2), 68–70.

Gil, Y., Pierce, S.A., Babaie, H., Banerjee, A., Borne, K., Bust, G., Cheatham, M., Ebert-Uphoff, I., Gomes, C., Hill, M., Horel, J., 2019. Intelligent systems for geosciences: an essential research agenda. Commun. ACM 62 (1), 76–84.

Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum. Comput. Stud. 43, 907–928, 1995.

Guha, R.V., Moore, A., 2016. The OKN white paper: open knowledge network: creating the semantic information infrastructure for the future, 7. http://ichs.ucsf.edu/wp-content/uploads/2017/08/OKN-White-Paper.docx. (Accessed 12 April 2018).

Haq, B.U. (Ed.), 2007. The Geological Time Table, sixth ed. Elsevier, Amsterdam, p. 1.

Hazen, R.M., Downs, R.T., Eleish, A., Fox, P., Gagne, O., Golden, J.J., Grew, E.S., Hummer, D.R., Hystad, G., Krivovichev, S.V., Li, C., Liu, C., Ma, X., Morrison, S.M., Pan, F., Pires, A.J., Prabhu, A., Ralph, J., Rumyon, S.E., Zhong, H., 2019. Data-driven

discovery in mineralogy: recent advances in data resources, analysis, and visualization. Engineering. Press 5 (3), 397–405.
Hogan, A., 2020. The semantic web: two decades on. Semantic Web 11 (1), 169–185.
Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., Gayo, J.E. L., Kirrane, S., Neumaier, S., Polleres, A., Navigli, R., 2020. Knowledge Graphs. arXiv Preprint arXiv:2003.02320.
ISO/TC-211, 2002. ISO 19108:2002 Geographic Information – Temporal Schema. International Organization for Standardization, Geneva, 48pp.
ISO/TC-211, 2011. ISO 19156:2011 Geographic Information – Observations and Measurements. International Organization for Standardization, Geneva, 46pp.
Kendall, E.F., McGuinness, D.L., 2019. Ontology Engineering. Morgan & Claypool, Williston, VT, p. 102.
Kitchin, R., 2014. The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences. Sage, London, p. 222.
Ludäscher, B., Lin, K., Brodaric, B., Baru, C., 2003. GEON: toward a cyber infrastructure for the geosciences—a prototype for geological map interoperability via domain ontologies. In: Soller, D.R. (Ed.), Digital Mapping Techniques '03—Workshop Proceedings, Millersville, PA, USA, pp. 223–229.
Ma, X., 2017. Linked Geoscience Data in practice: where W3C standards meet domain knowledge, data visualization and OGC standards. Earth Sci. India 10 (4), 429–441.
Ma, X., 2020a. Input label - return begin, end and duration of the concept in the latest global vocabulary scheme. https://tinyurl.com/y38e2fu8. (Accessed 26 August 2020).
Ma, X., 2020b. 3-layered concept hierarchy. https://tinyurl.com/yxdajem2. (Accessed 26 August 2020).
Ma, X., Fox, P., 2013. Recent progress on geologic time ontologies and considerations for future works. Earth Sci. India 6 (1), 31–46.
Ma, X., Asch, K., Laxton, J.L., Richard, S.M., Asato, C.G., Carranza, E.J.M., van der Meer, F.D., Wu, C., Duclaux, G., Wakita, K., 2011a. Data exchange facilitated. Nat. Geosci. 4 (12), 814.
Ma, X., Carranza, E.J.M., Wu, C., Van Der Meer, F.D., Liu, G., 2011b. A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps. Comput. Geosci. 37 (10), 1602–1615.
Ma, X., Carranza, E.J.M., Wu, C., van der Meer, F.D., 2012. Ontology-aided annotation, visualization, and generalization of geological time-scale information from online geological map services. Comput. Geosci. 40, 107–119.
Ma, X., Fox, P., Rozell, E., West, P., Zednik, S., 2014. Ontology dynamics in a data life cycle: challenges and recommendations from a Geoscience Perspective. J. Earth Sci. 25 (2), 407–412.
Mascarelli, A.L., 2009. Quaternary geologists win timescale vote. Nature 459, 624.
Miles, A., Bechhofer, S., 2009. SKOS simple knowledge organization system reference. https://www.w3.org/TR/skos-reference/. (Accessed 27 August 2020).
Morrison, M.M., Eleish, A., Prabhu, A., Narkar, S., Pan, F., Huang, F., Fox, P., Hystad, G., Liu, C., Buongiomo, J., Zhang, S., Ma, X., Ralph, J., Krivovichev, S.V., Giovannelli, D., Runyon, S.E., Hummer, D., Golden, J.J., Downs, R.T., Hazen, R.M., 2019. Exploring Carbon Mineralogy and Mineral Evolution through Deep Time with Advanced Analytics and Visualization. AGU Fall Meeting 2019, San Francisco, CA, USA.
Normile, D., 2019. Earth scientists plan a 'geological Google'. Science 363 (6430), 917.
Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E.L., Simonsohn, U., Soderberg, C., Spellman, B.A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E.J., Wilson, R., Yarkoni, T., 2015. Promoting an open research culture. Science 348 (6242), 1422–1425.
NRC (National Research Council), 2008. Origin and Evolution of Earth: Research Questions for A Changing Planet. National Academies Press, Washington, DC, p. 150.
OGC, 2012. OGC GeoSPARQL - A Geographic Query Language for RDF Data. Open Geospatial Consortium, Wayland, MA, p. 57pp.
Paulheim, H., 2017. Knowledge graph refinement: a survey of approaches and evaluation methods. Semantic Web 8 (3), 489–508.
Perrin, M., Mastella, L.S., Morel, O., Lorenzatti, A., 2011. Geological time formalization: an improved formal model for describing time successions and their correlation. Earth Sci. India 4 (2), 81–96.
Raskin, R.G., Pan, M.J., 2005. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). Comput. Geosci. 31 (9), 1119–1125.
Sheth, A.P., 1999. Changing focus on interoperability in information systems: from system, syntax, structure to semantics. In: Goodchild, M., Egenhofer, M., Fegeas, R., Kottman, C. (Eds.), Interoperating Geographic Information Systems. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 5–29.
Sheth, A., Padhee, S., Gyrard, A., 2019. Knowledge graphs and knowledge networks: the story in brief. IEEE Internet Computing 23 (4), 67–75.
Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E., Wyborn, L., 2019. Make scientific data FAIR. Nature 570, 27–29.
Wang, C., Ma, X., Chen, J., 2018. Ontology-driven data integration and visualization for exploring regional geologic time and paleontological information. Comput. Geosci. 115, 12–19.
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., Bouwman, J., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, 160018.
Wood, J., Andersson, T., Bachem, A., Best, C., Genova, F., Lopez, D.R., Los, W., Marinucci, M., Romary, L., Van de Sompel, H., Vigen, J., 2010. Riding the Wave: How Europe Can Gain from the Rising Tide of Scientific Data. Final Report of the High Level Expert Group on Scientific Data–A Submission to the European Commission. European Union, p. 36.
Yang, C., Huang, Q., Li, Z., Liu, K., Hu, F., 2017. Big Data and cloud computing: innovation opportunities and challenges. Int. J. Digital Earth. 10 (1), 13–53.
Zaslavsky, I., Valentine, D., Richard, S.M., Gupta, A., Meier, O., Peucker-Ehrenbrink, B., Hudman, G., Stocks, K.I., Hsu, L., Whitenack, T., Grethe, J.S., 2017. EarthCube Data Discovery Hub: Enhancing, Curating and Finding Data across Multiple Geoscience Data Sources. AGU Fall Meeting, New Orleans, LA, USA.