

d1g1tal ge0

Project – Advance Databases



PERRY HOUSER

RICHARD SAMPLES

Table of Contents

[1.1 DATABASE SCOPE AND REQUIREMENTS 2](#_Toc437416611)

[Figure 1: The Geologic Time Scale. Colors of rock units on all geologic maps follow the colors based on the 2014 ICC Map. 5](#_Toc437416612)

[Figure 2: USGS Geologic Map and Stratigraphic Column of Mount Cristo Rey. 6](#_Toc437416613)

[Figure 3: Cristo Rey Geologic Map, by Perry Houser. Inset image (upper-right) USGS Geologic Map and Stratigraphic column for Mount Cristo Rey 7](#_Toc437416614)

[1.2 REQUIREMENTS AND ASSUMPTIONS 8](#_Toc437416615)

[1.3 NON FUNCTIONAL REQUIREMENTS 9](#_Toc437416616)

[1.3 SCHEMA MODEL (CONCEPT MAP) 11](#_Toc437416617)

[Figure 4: Concept Map Overview of Data Integration and Visualization Applied to Geologic Field Mapping 11](file:///C:\Users\PeachyJoKiller88\Dropbox\database\2%20-%20FinalProjectReport.docx#_Toc437416618)

[Figure 5: Concept map for Geologic Time Scale and Stratigraphic integration, including data from USGS and NOAA. Dashed lines indicate information inferred by the reasoner and ontology. 13](#_Toc437416619)

[1.4 DATABASE RECORDS 14](#_Toc437416620)

[Figure 6: Quaternary Instance 14](#_Toc437416621)

[Figure 7: Archean Instance. 15](#_Toc437416622)

[Figure 8: Landslide Instance 16](#_Toc437416623)

[Figure 9: Muleros Andesite Instance 17](#_Toc437416624)

[Figure 10: Observation Kd\_PH Instance 18](#_Toc437416625)

[Figure 11 Qr\_PH Instance 19](#_Toc437416626)

[1.5 QUERIES 20](#_Toc437416627)

[1.6 CONCLUSION 25](#_Toc437416628)

[1.7 REFERENCES 26](#_Toc437416629)

# DATABASE SCOPE AND REQUIREMENTS

The purpose of the class project is to provide geologist with a portable “offline” database that structures data specific to whatever particular location the geologist will map. Recording information to the database can allow the user to use the database as a digital field notebook, or knowledge base for geologist. In creating this knowledge base, we will use an Ontology to store geologic field data and create formalized explicit relationships within the data in order to extend beyond a database and be able to consistently form implicit knowledge formation, Noy (2001). This computer aided inferred data can provide a secondary point-of-view, that can be used to augment and assist the geologist situational awareness of the study area. The current system of mapping does not adequately assist the user in data mining during field mapping exercises, since it consists principally of paper field notes which are not easily amenable to change, there is difficultly in overseeing, collecting, and sharing relevant information which would be useful to other geologist.

Team D1g1tal Ge0 aims to provide an alternative which involves: collecting, maintaining, and managing geological information from a consortium of various sources so that users can access and modify information related to their study area, while the system also suggests supplementary information to the user about the field site. Our hypothesis is that an ontology system can be beneficial to geologists whom perform extensive fieldwork, both as training during their education and as their professional tasks after graduating. Assisting the field mapping process and training is of surmount importance since it provides the bases for the professional performance of geologists, it involves a large number of recorded observations and measurements of geologic features, understanding complex relationships and forces involved, and sharing a description of the evolution and current state of the study area, usually visualized as a geologic map, however we believe that an ontology can provide another service for standardizing field data, integrating other data sources, assisted data mining, and sharing and collaboration within a study area. We scoped the project build an ontology that would contain information about a single study area and implement a rule chain to include Steno’s Laws.

The GeoField ontology was an initial effort to provide an improved field training experience by integrating diverse information that can be used for field trips. The long term vision is to provide a comprehensive data model to integrate information about a certain location which will be visited during a field trip, pack such information and have it available for the duration of the trip, without having to depend on online access for updates. Other data considered for integration in future iterations of the ontology are scientific papers concerning the location itself, same as a geo-referenced model of geological characteristics.

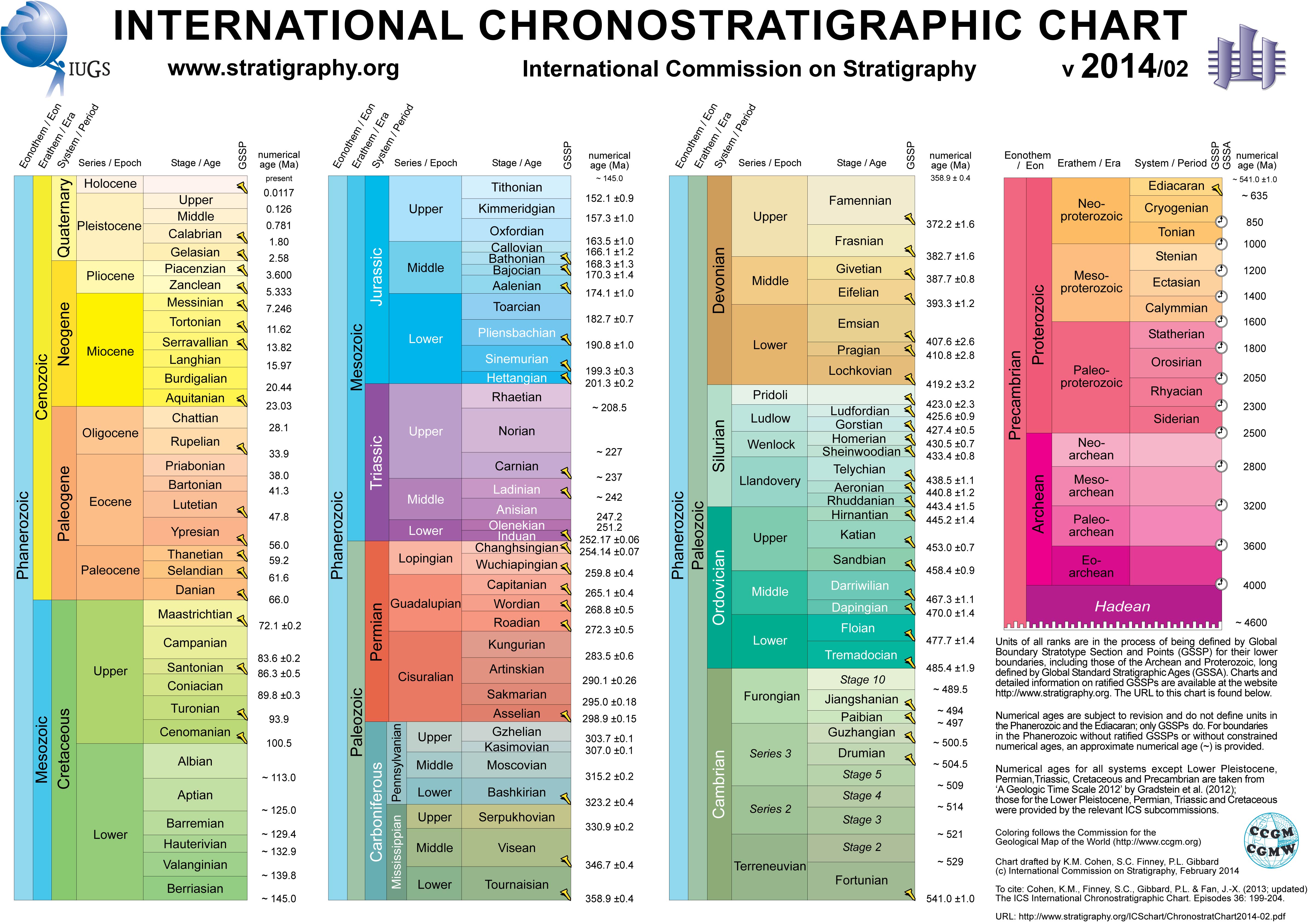
For the class project, the GeoField-TimeScale ontology was created to extend the GeoField ontology to include data from a USGS geologic map and inset stratigraphic column (see Figure 2) was used to create part of the dataset used in the ontology. Elements included from the USGS stratigraphic column are the geologic feature's name, aberration, geologic age, and the "rank" or location within the column derived by setting the “youngest” unit (Qc) as rank 1, continuing down the column to the “oldest” geologic feature in the study area, the Finlay Formation with a rank equaling 15. Figure 2 shows data taken from a 2008 geologic field mapping exercise created by Perry Houser, data points provides specific data regarding geometric measurements, name, and type of rock. As stated, we also will be able to importing the GeoField Ontology (2015), which expands the ontology to include earthquake information from the USGS and weather observations should they exist; as well as an expanded set of classes to record field notebook information.

In order to implement Steno’s Laws, the field data must include information about the local stratigraphy, which studies both ground surface geologic features (e.g. landslides) and subsurface information on the geologic rock units. Stratigraphy comes from the initial concepts established by Steno’s Laws, formulated by Nicolaus Stenos circa 1669 and which explain the behavior of soil and rock strata, this is important concept for students to comprehend and apply towards learning geologic structure and processes, Gray (2011). These rules can be used within an ontology to assist the user in providing a possible list of rock units or features within the rocks, which a user can look for.

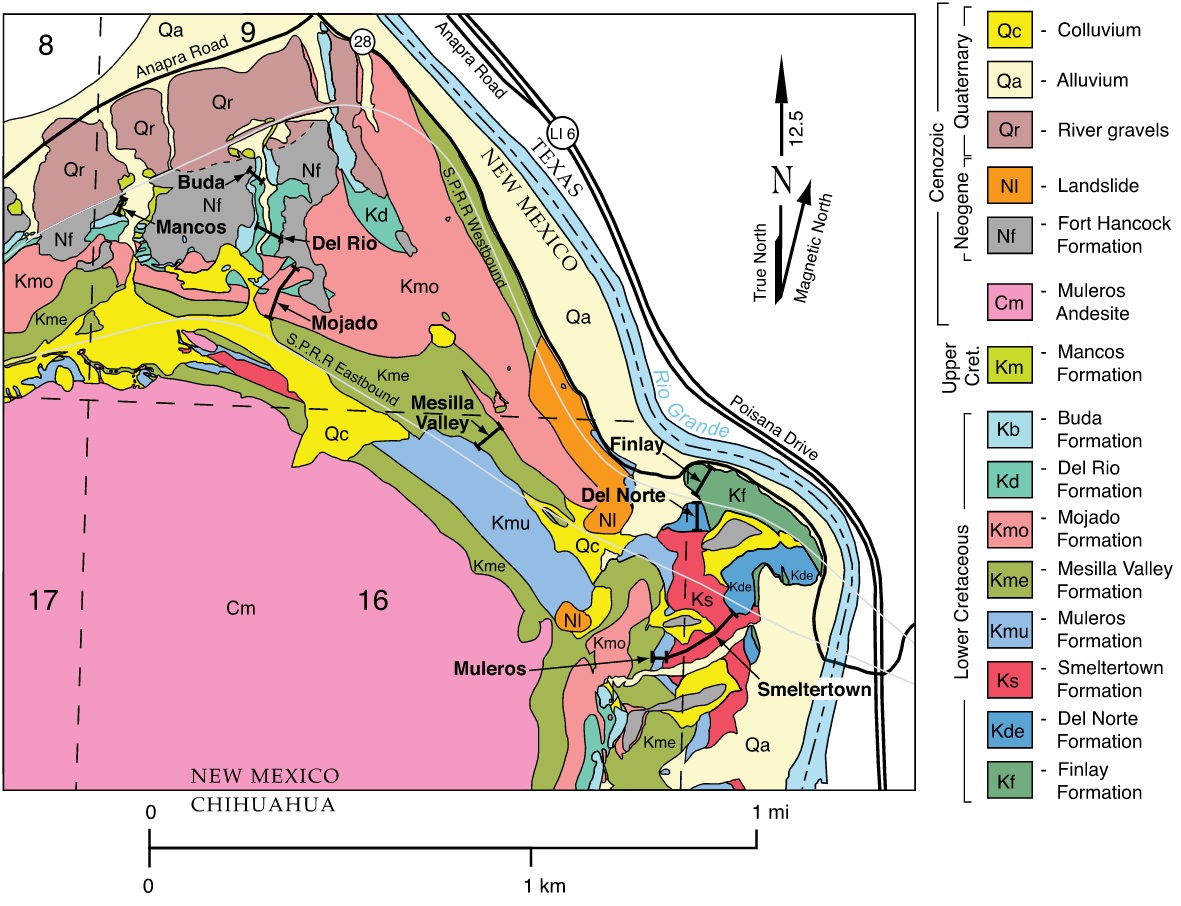
Steno’s laws are:

1. Law of superposition: Younger layers of rock sit above older layers.
2. Law of original horizontal: Layers are originally deposited horizontally, though some tilting may occur later.
3. Law of cross-cutting relationships: When an intrusion occurs, the layers in which the new rock intrudes is older than the intruder.
4. Law of lateral continuity: Layers are continuous until they encounter other solid bodies, which block their deposition or are taken away by carrying agents appearing after deposition.

Integration of the geologic time scale information and a stratigraphic column information may permit predictions of the ground as you walk from one rock type to another, and provide the user with extra information based on the user entry or user query. The stratigraphic column of the local mapping area contains a list all of the rock types listed in the area, the local rock units can be aged and standards for assigning data about color and abbreviations are based off the International Chronostratigraphic Chart (ICC), see Figure 1. Predictions can be suggested by using this data and rules within the ontology. For example, based on the given stratigraphic column data, asking the ontology to predict or aid the user to state the next possible rock units that the user would expect to walk over next.

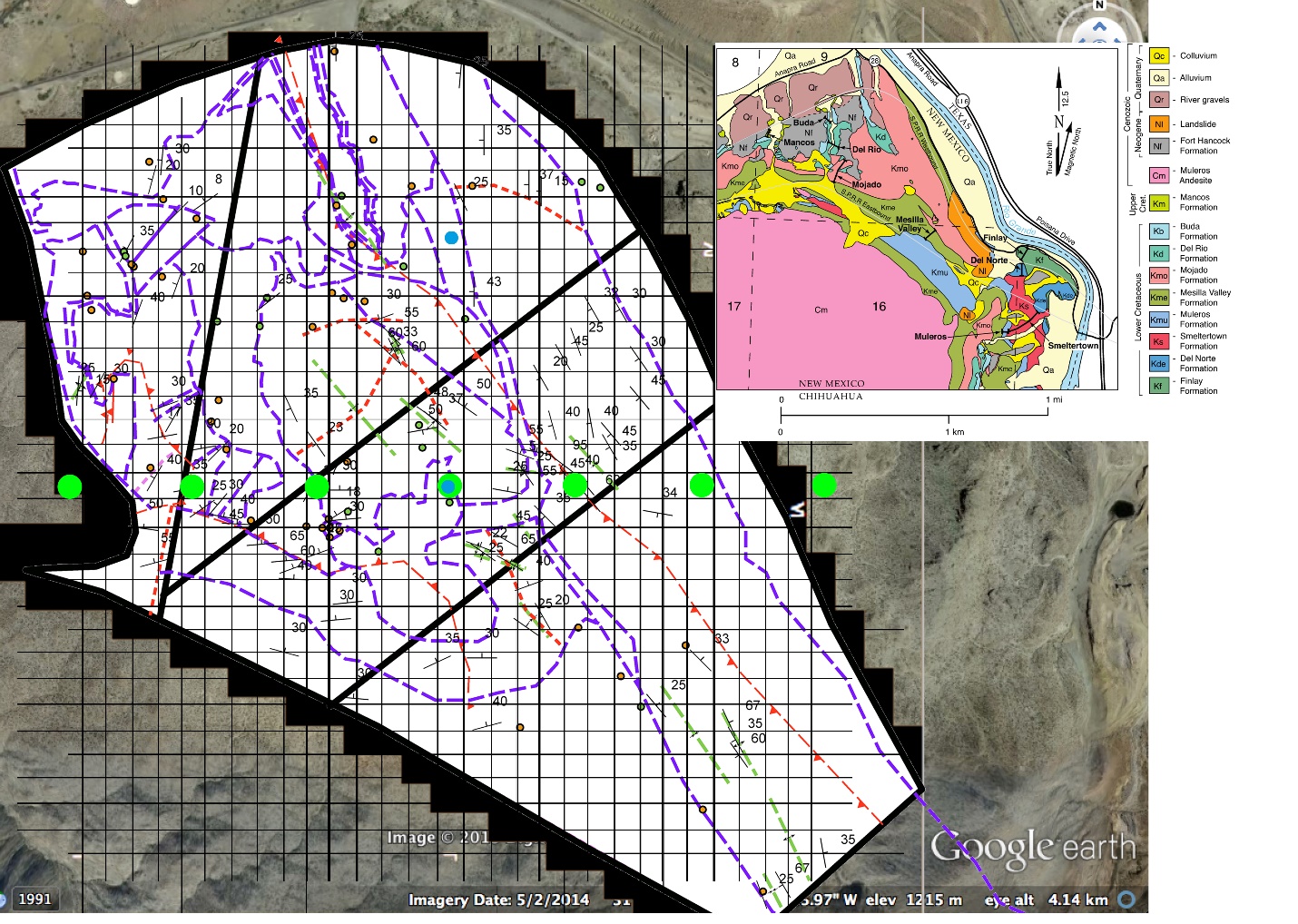


## Figure 1: The Geologic Time Scale. Colors of rock units on all geologic maps follow the colors based on the 2014 ICC Map.



## Figure : USGS Geologic Map and Stratigraphic Column of Mount Cristo Rey.

The study site selected for the project scope is Mount Cristo Rey, New Mexico. Our project domain is this site and information within. The site is located on two international borders and three states and it close proximity to the University of Texas at El Paso (UTEP) makes the area a common field trip for geology students, where mapping exercises are performed in a geologic setting that spans for millions of years to present day, featuring dinosaur tracks, marine fossils, and a large igneous intrusion that comprises of the peak of the mountain.



## Figure 3: Cristo Rey Geologic Map, by Perry Houser. Inset image (upper-right) USGS Geologic Map and Stratigraphic column for Mount Cristo Rey

# 1.2 REQUIREMENTS AND ASSUMPTIONS

Development of the ontology was guided by Noy (2001). The first step of Noy’s describes is to define the domain and the scope of the ontology, which was described above (1.1 Database Scope and Requirements). Next, a set of questions that the ontology should be able to answer using information within the knowledge base is given below.

Competency Questions:

1. The system shall determine which rock unit might the user might find next.
2. The system shall determine which rock unit is older and younger than the current unit the user is on.
3. The system shall return results on all the information about a given rock unit.
4. The system shall determine which geologic period or era a particular rock is.
5. The system shall determine what other rocks are associated with this particular unit.
6. The system shall determine what other rocks are associated with this mapping unit.
7. The system shall determine what is the stratigraphic column at a certain location.
8. The system shall determine what fossils exist in a particular unit or area.
9. The system shall determine which rock units have been mapped by which mapper.

Assumptions of the model functionality should also be considered:

1. Data set is a limited sample size, limited to a single location.
2. Users would have to use the same ontology, Protégé, to process OWL files based on XML/RDF.
3. Geo-referencing data would be limited to positioning related to a stratigraphic column.

# 1.3 NON FUNCTIONAL REQUIREMENTS

1. The system shall use the current release of Protégé and Hermit
2. The system shall be user friendly, simple queries and data entry.
3. The system shall be efficient, quick loading and processing of the reasoner.
4. The System shall be extendable to the GeoField ontology.

Next step Noy suggest is to consider reusing existing ontologies. For this project we did consider using several existing ontologies. We considered the SWEET 2.3 ontology and SEEGRID ontology, however due to the large complex nature we found the integration to be more difficult than creating a generalized ontology from scratch. Noy’s next step is to enumerate important terms in the ontology, show as a concept maps in Figure 4 and Figure 5. Step 4, involves establishing a set class hierarchy and defining properties of concepts are visualized as a process through time and space using the concept maps using a top-down development process, that sets classes via the most general concept of the domain and becoming more specific according to the scope of the project. Once the classes are set, the next step is to define the data and object properties that act as relationships. Step 6, involves adding constraints your data and develop rules that begin to turn implicit information from explicit data input, here we added a rule chain to setup the inference of what geologic unit might be expected to be seen next. A ranking system was used to deal with slot cardinality of ordering the series of geologic units that make up the stratigraphic column. The final step is to create the instances. We manually created all the instances that consists of the knowledge base, however an automated or semi-automated system should be considered in future work, See Appendix for instances and queries. Our personal final step was to test the competency questions using the Protégé query language: DL Query.

# 1.3 SCHEMA MODEL (CONCEPT MAP)

## Figure : Concept Map Overview of Data Integration and Visualization Applied to Geologic Field Mapping

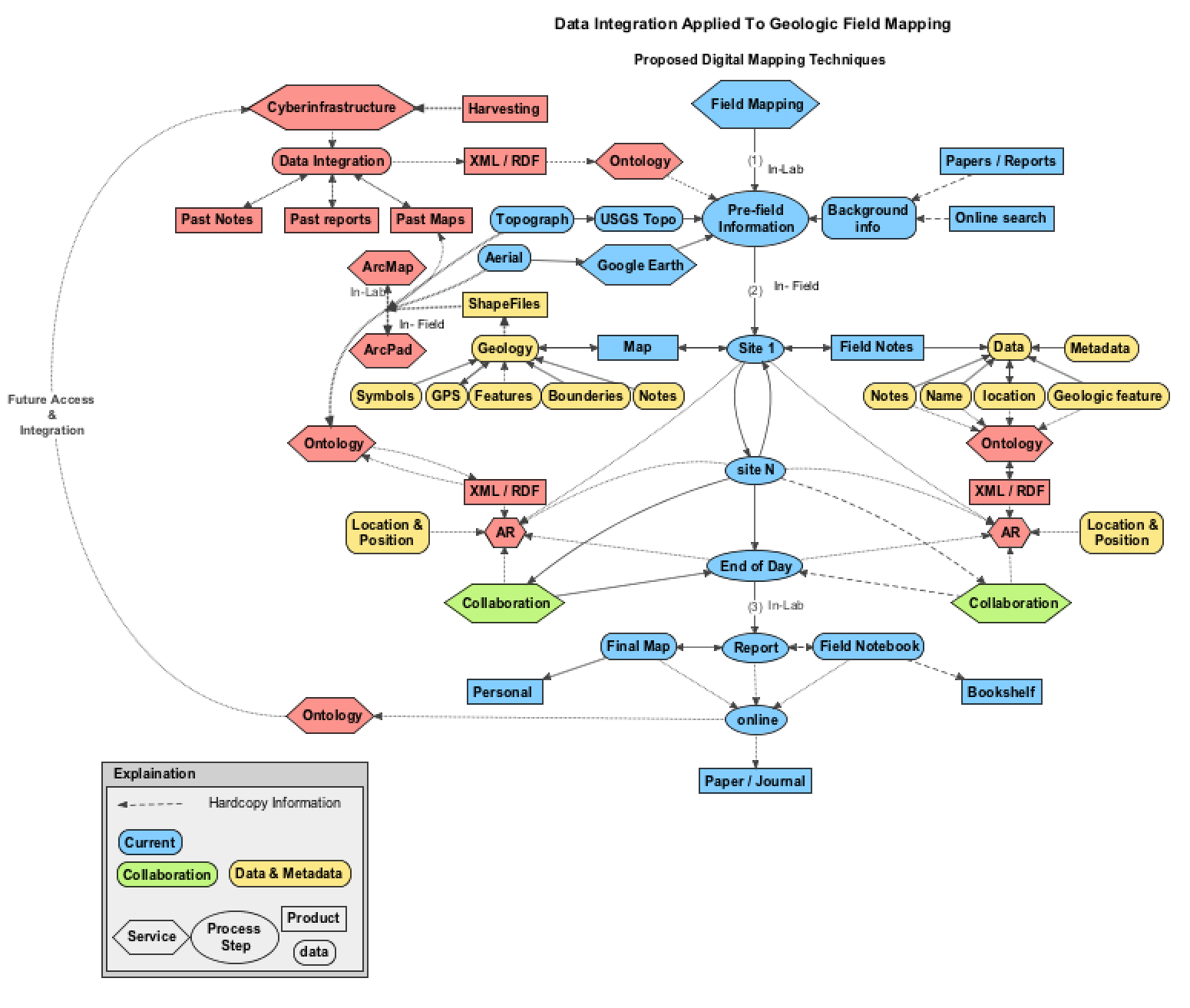
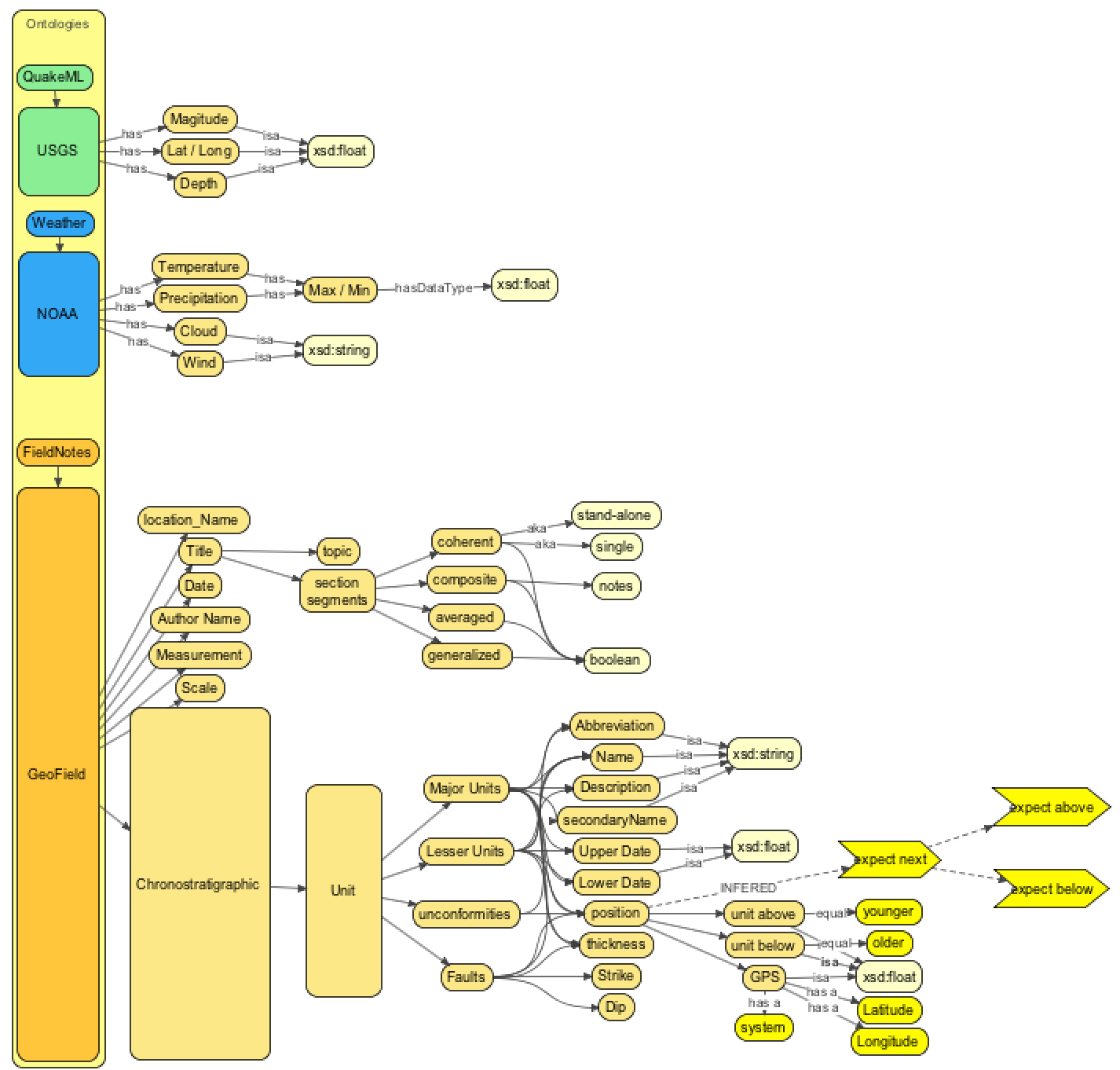
Figure 4 shows a high-level overview of cyber-infrastructure applications to geologic field mapping. Blue colored nodes represent current mapping techniques commonly used today; red notes are the proposed extension to the current mapping process, which uses elements of cyber-infrastructure such as semantics, databases, visualization, and collaboration opportunities. Green nodes indicate possible opportunities to collaborate and share (digitally copy) data. Yellow nodes are data requirements such as information stored as strings or integers. Starting at the top-center node labeled, "Field Mapping" the field mapping process normally begins before entering the field site in order to gather information about the study site. Geologist typically have limited resources for field mapping, such as time and money therefore it is very important to acquire a field map and notes so that an in-depth analysis and report or final map can be made later, away from the field site. This information may consist of gathering publications, imagery (e.g. topographic maps, aerial photography), this information, while useful to people, may consist of both digital and printed hardcopies that have inadequate relationships for which a computer can find useful, which can be viewed as a limitation with the knowledge resource when in the field. When visiting the actual study area, a map is developed, typically digitally using ERSI ArcPad, and field notes are taken, typically hand written or meta-data within the map. Map information such as geologic points and lines can be vector data and imagery is typically raster formats. Global Positioning Systems (GPS) are used to geo-reference the data to the map. Field notes may contain expanded metadata detailing an observation or series of measures. If there are multiple geologists mapping the same area, information may be shared upon meeting, which might be limited to verbal communications. Then once done, the geologist will leave the field site to create the final map and/or report.

Figure 4, demonstrates the process to augment the current field mapping process with the use of cyber-infrastructure such as implementation of semantics, databases, visualization and collaboration. Visualization of data using Augmented Reality (AR), is the next step of the future project expansion and is outside of our current project scope, however this project establishes a base for AR use. The blue nodes describe altering the current process, by including a search and download of XML/RDF based geologic data of the study site and recording data into an ontology while mapping and upon completion of the study that may occur beyond the field mapping exercise. Figure 5, shows how information can be inferred based of data input from the user during mapping. This can be extremely useful for data mining and represents another source of publication. Ontologies can assist in fact checking or sharing large complex sets of information while still in the field. Testing your data against the real world field study environment in real-time is an important process of ground-truthing your data, both verification and validation of your model can be assisted with an ontology, as well as sharing data with others. Development of the system as a general setup for any field mapping situation, can have broad impacts that can improve field work for other disciplines, e.g. biologist, ecologist, etc.

# 1.6 CONCLUSION

The final product was a successful build of a base ontology useful in entering personal user field notes, adding notes from outside, non-ontology, sources such as geologic thickness for units and other elements from Lucas (2010), which shows that users could enter information from sources such as academic publications into relationships built in the ontology for data mining and presentation to user as they needed upon mapping a field area. The database provided by D1g1tal Ge0 aims to ameliorate various issues of dealing with the traditional method of geologist's field notes which are encumbered by various issues of: the size of the data, issues of updating, reducing redundancy, allowing this data to grow and most importantly all the while still being both manageable and available to other geologist are of pivotal importance, and just the problems suitable for a database provided by D1g1tal Ge0 to solve.



## Figure 5: Concept map for Geologic Time Scale and Stratigraphic integration, including data from USGS and NOAA. Dashed lines indicate information inferred by the reasoner and ontology.

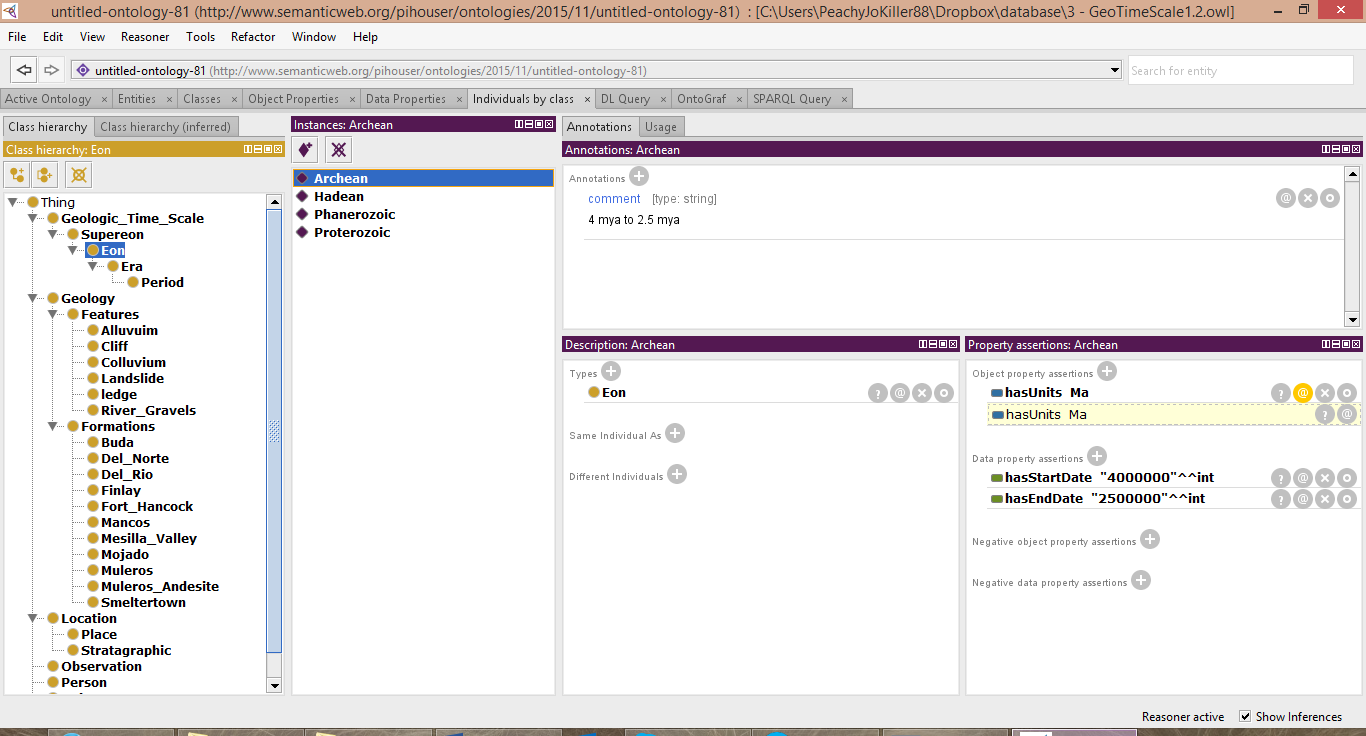
# Appendix

# 1.4 DATABASE RECORDS



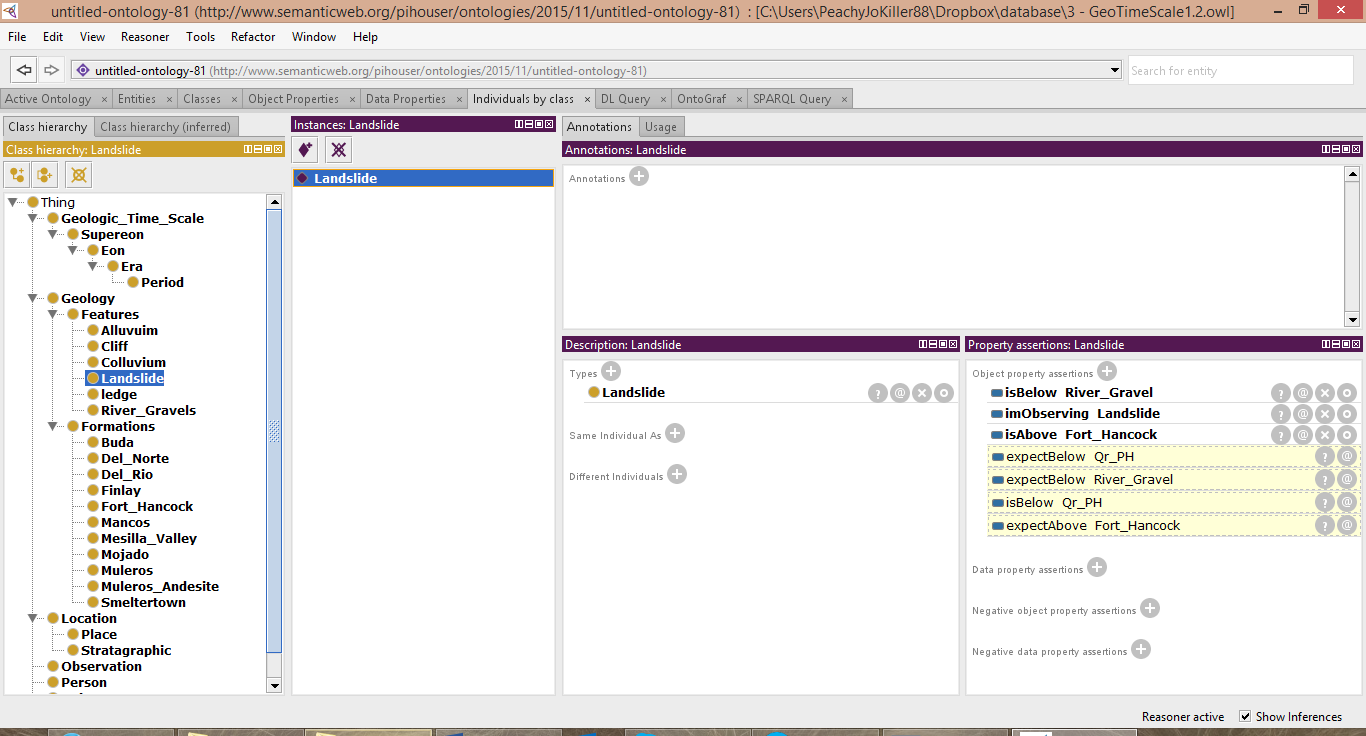
## Figure 6: Quaternary Instance

The instance "Quaternary" has object properties establishing that it has a geologic era in the Cenozoic and has an object property that the geologic time period is measured in millions of years (Ma). There are data properties for the start and end dates which use integers or floats



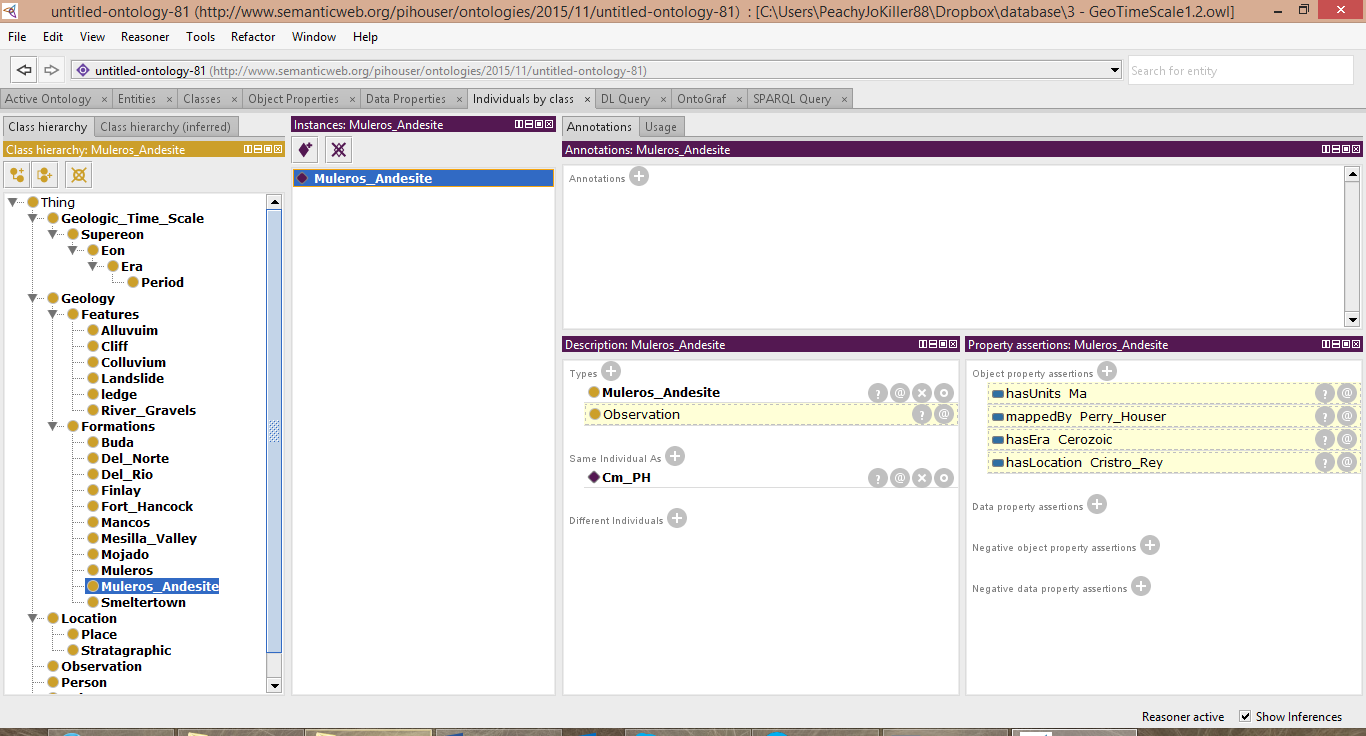
## Figure 7: Archean Instance.

Eon example instance for the Archean geologic time with comment tag annotation indicating that the data entry user can relate extended notes about a feature in a natural human language format.



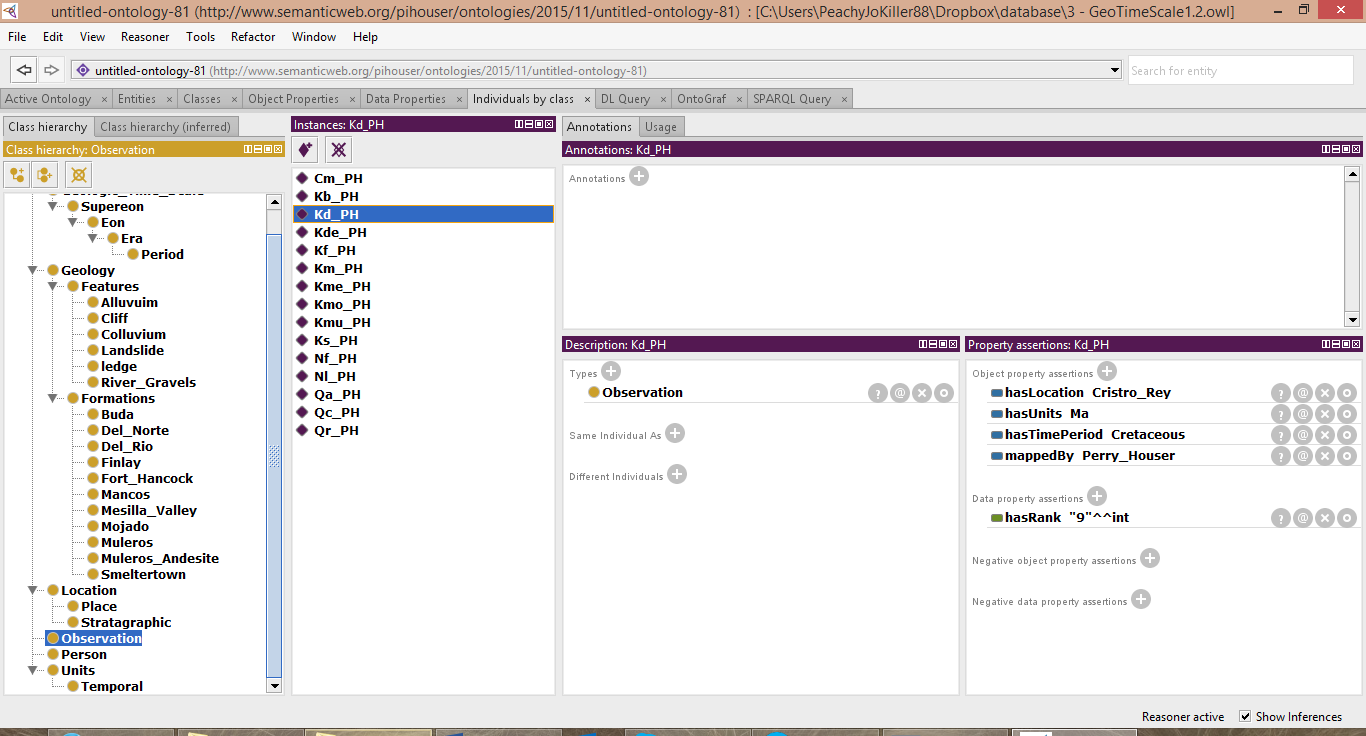
## Figure 8: Landslide Instance

The Yellow-Highlighted properties, then this means that the Hermit reasoner that Protégé uses has inferred the property from the rules and structure of the ontology. Hermit is able to correctly suggest that the next possible rock units might be river gravels or the Fort Hancock Formation, as well as noting that there is an observation recorded by Perry Houser that is related to the landslide.



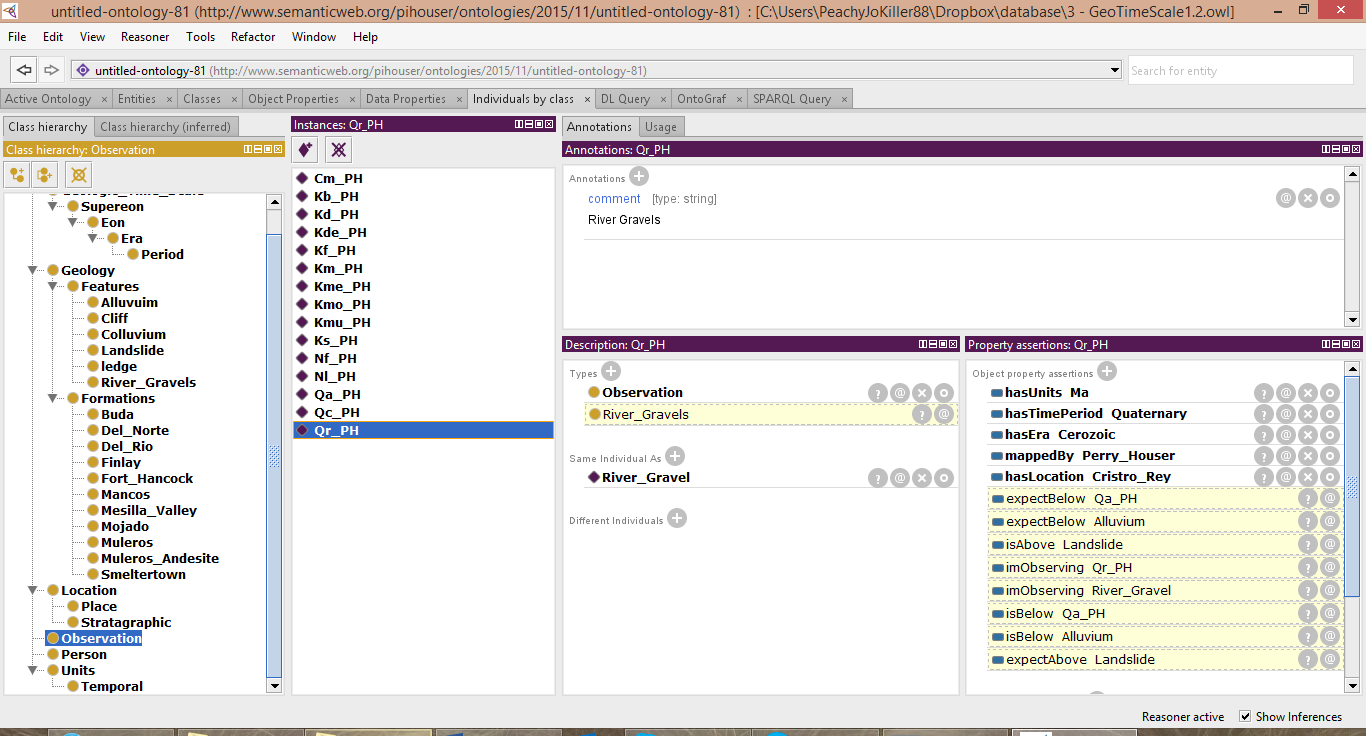
## Figure 9: Muleros Andesite Instance

Example showing the inferences provided about one of the rock formations by adding a “Same individual as” data relation to an Observation entry. Here the Andesite formation is linked to an observation.



## Figure 10: Observation Kd\_PH Instance

Screenshot of the list of test observations by Perry Houser, these observations have a location, the name of the mapper, a Time period and its related units, and contained a user-based ranking system that places the unit being mapped within a stratigraphic column.



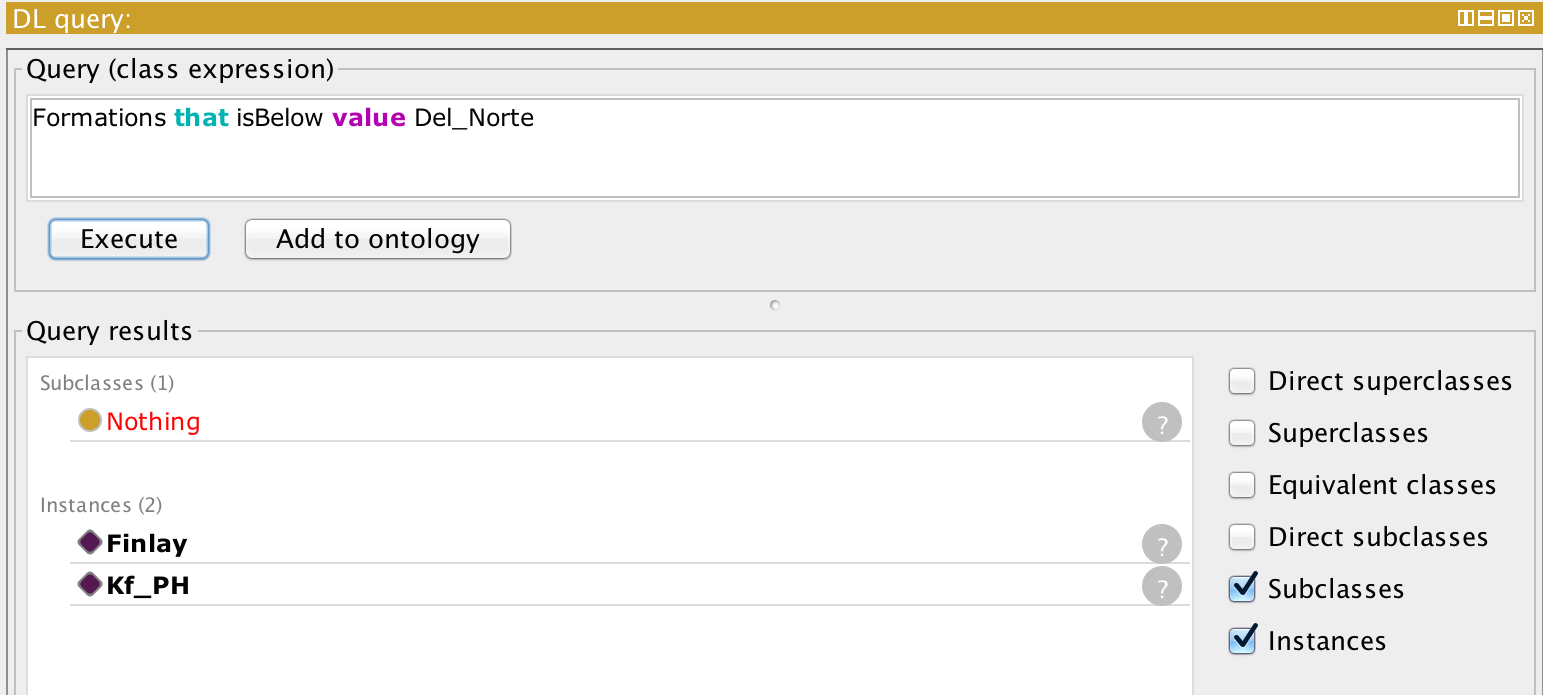
## Figure 11 Qr\_PH Instance

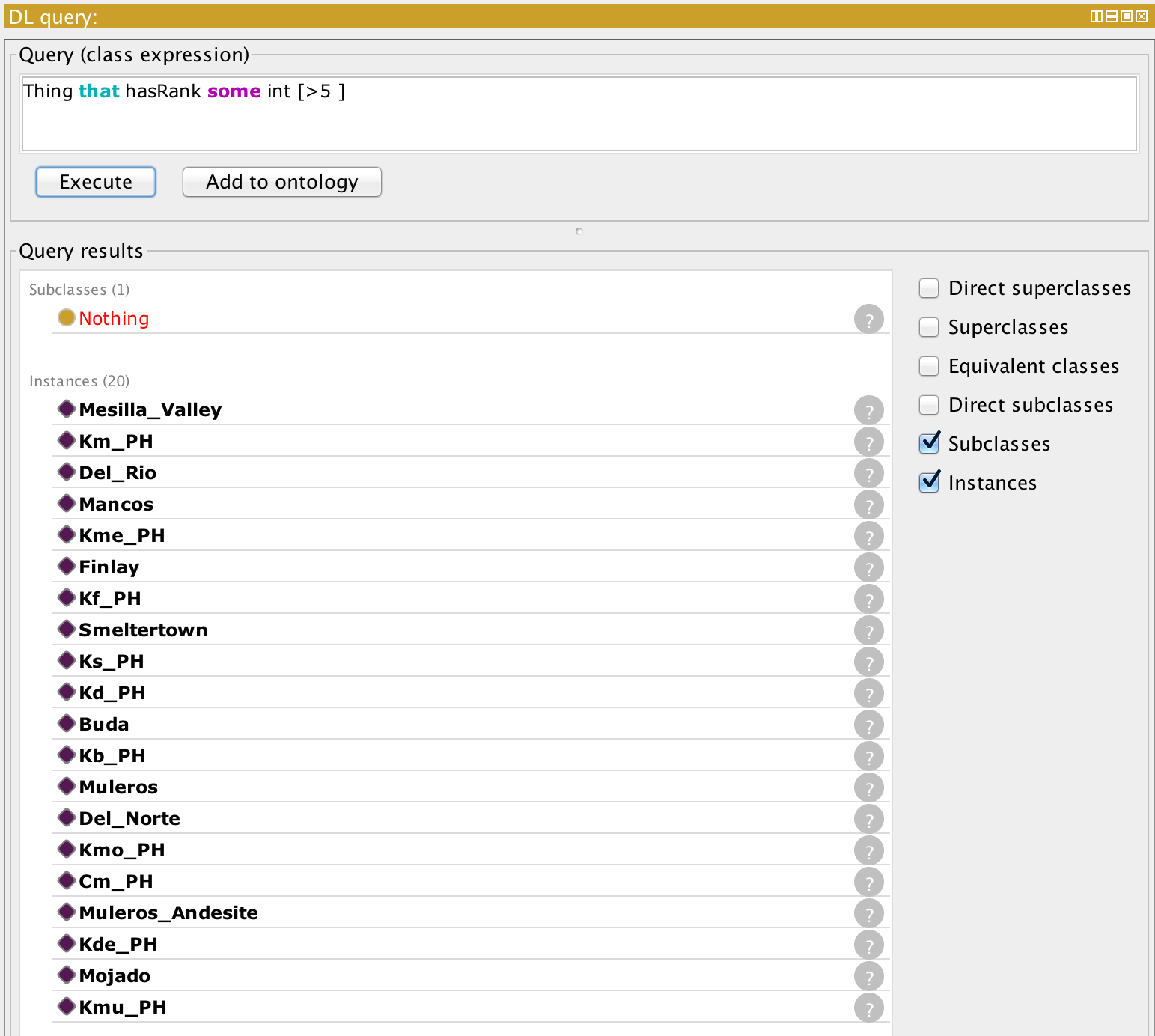
Screenshot showing an example of the reasoner active and inferring data about the observation taken. The annotation fields are used and the instance is linked to the River Gravel class.

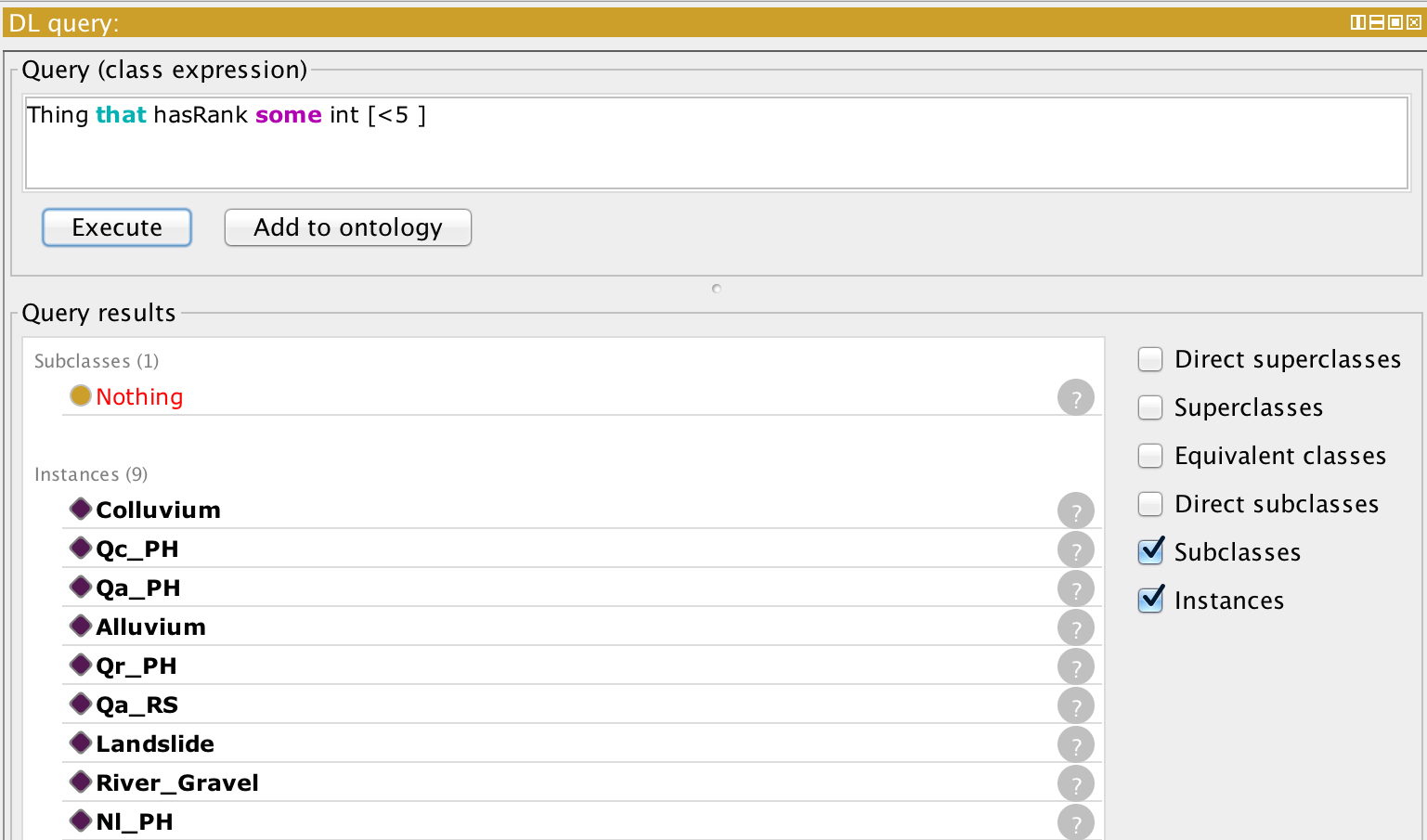
# 1.5 QUERIES

Some competency questions/queries and results:

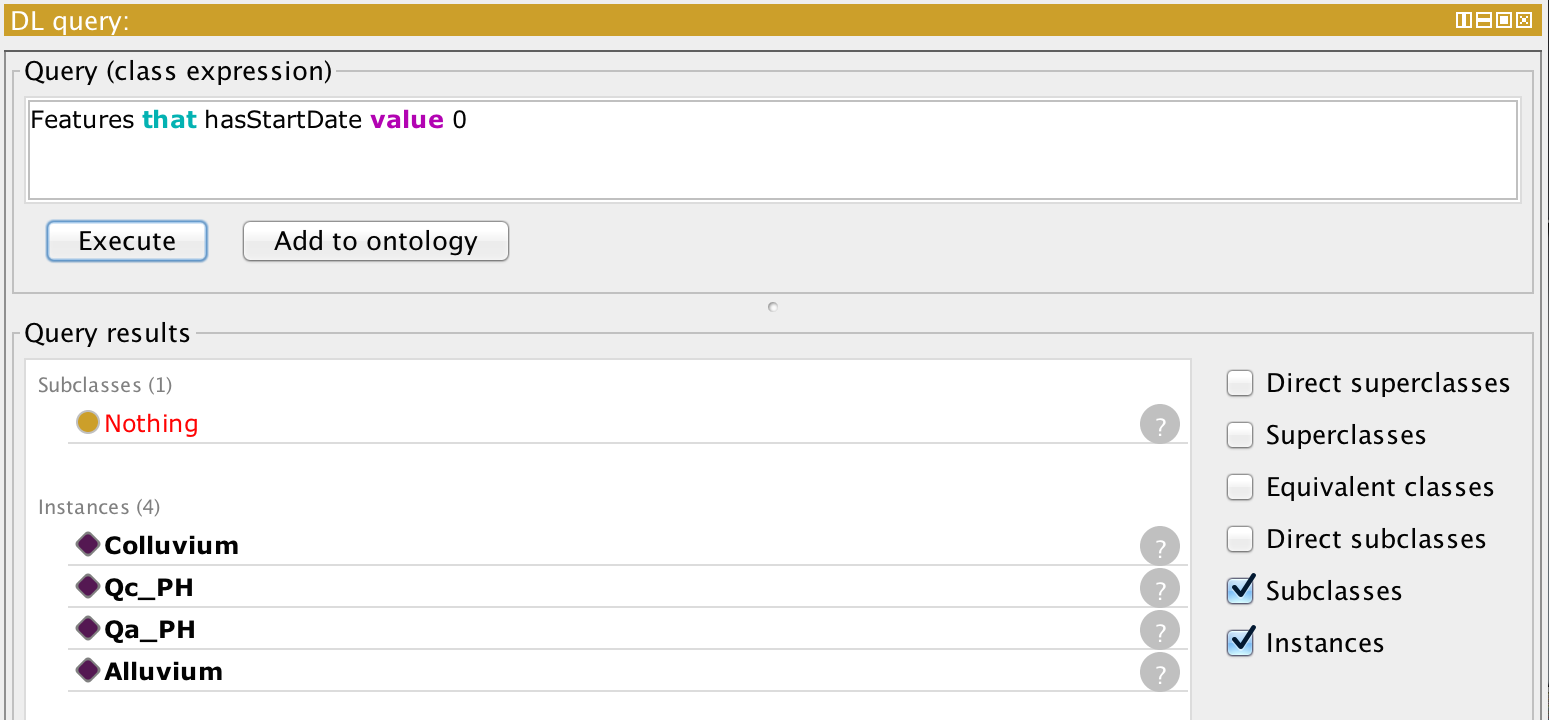
1. I am on Del Norte Rock unit, what rock unit might I find next?



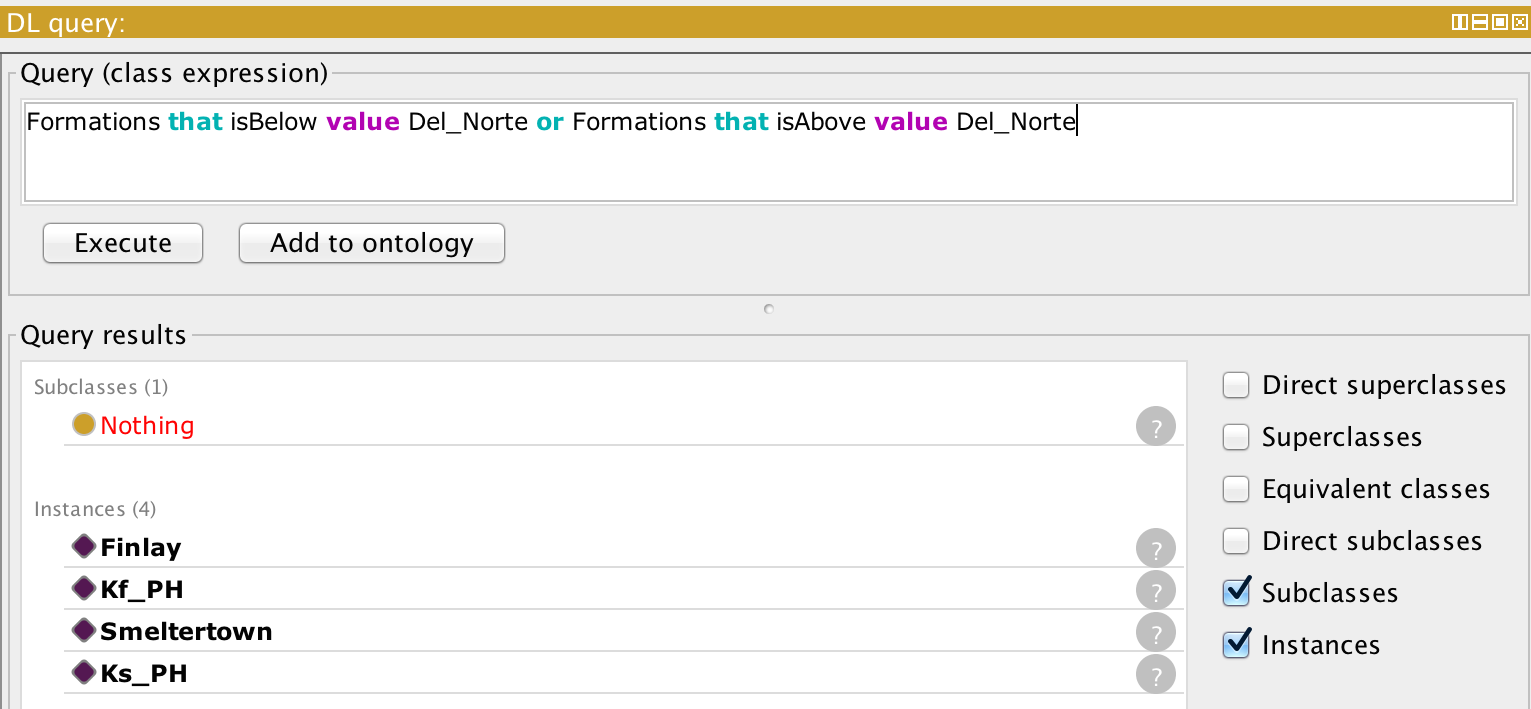
1. What rock unit is older than the current unit I'm on?  
   
2. What rock unit is younger than the current unit I'm on?

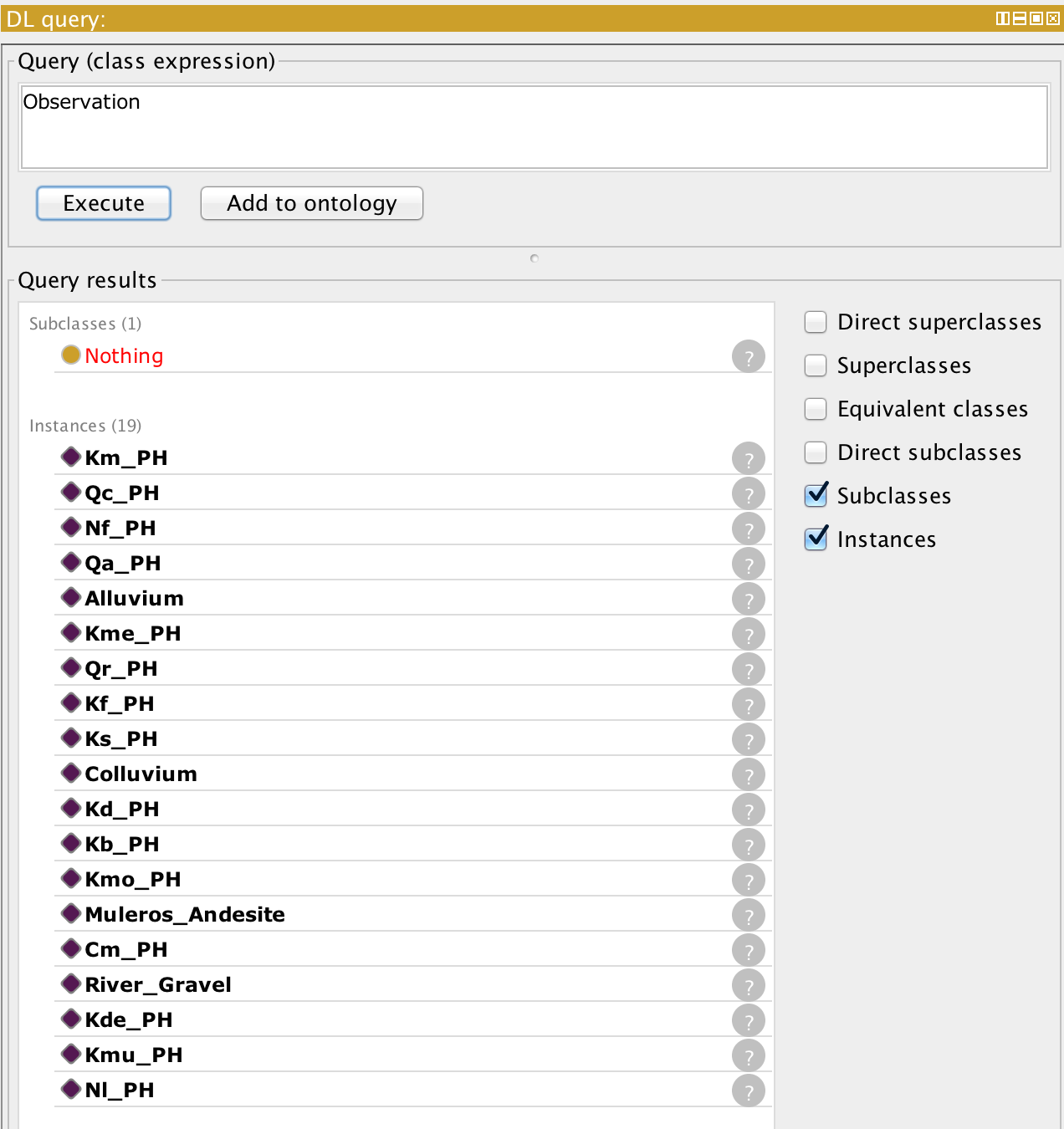


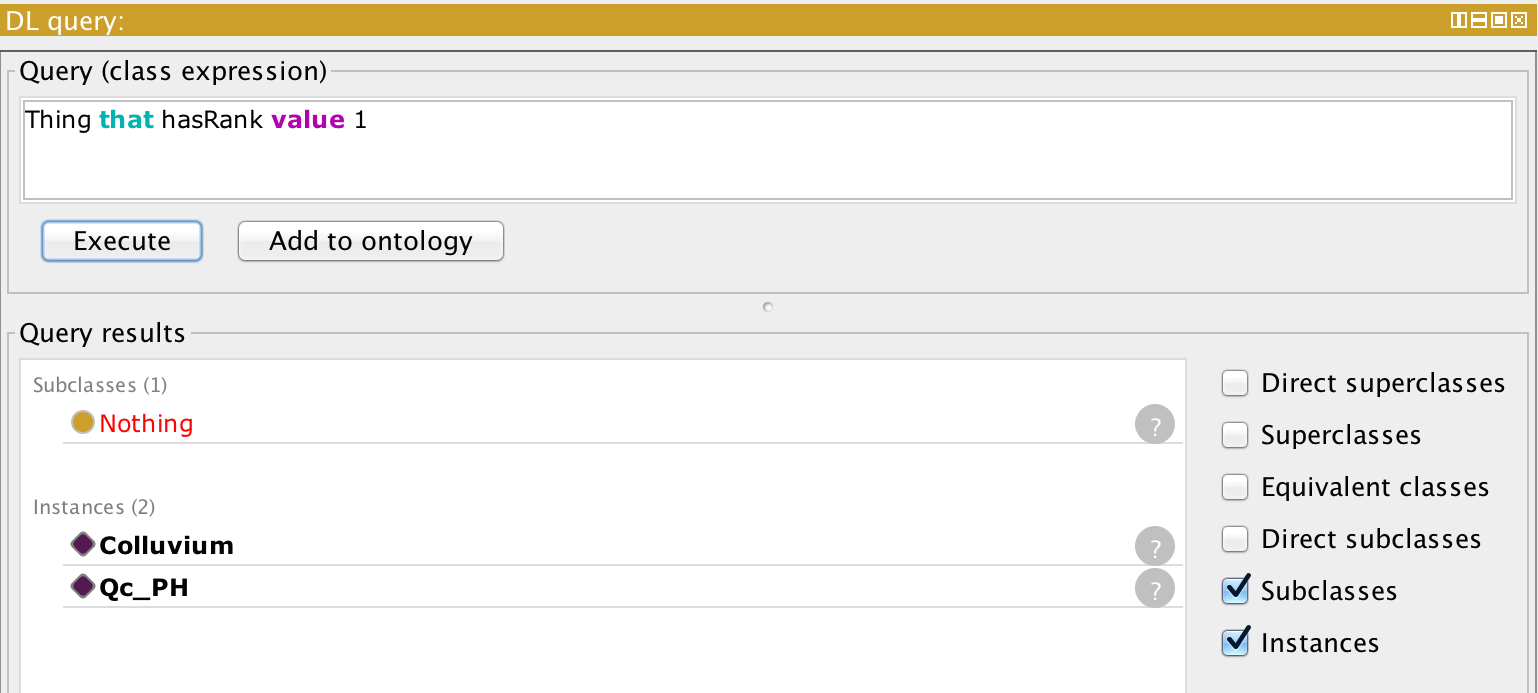
1. What is the strike and dip for Finlay Formation rock unit?
2. What geologic features are the youngest/newest?



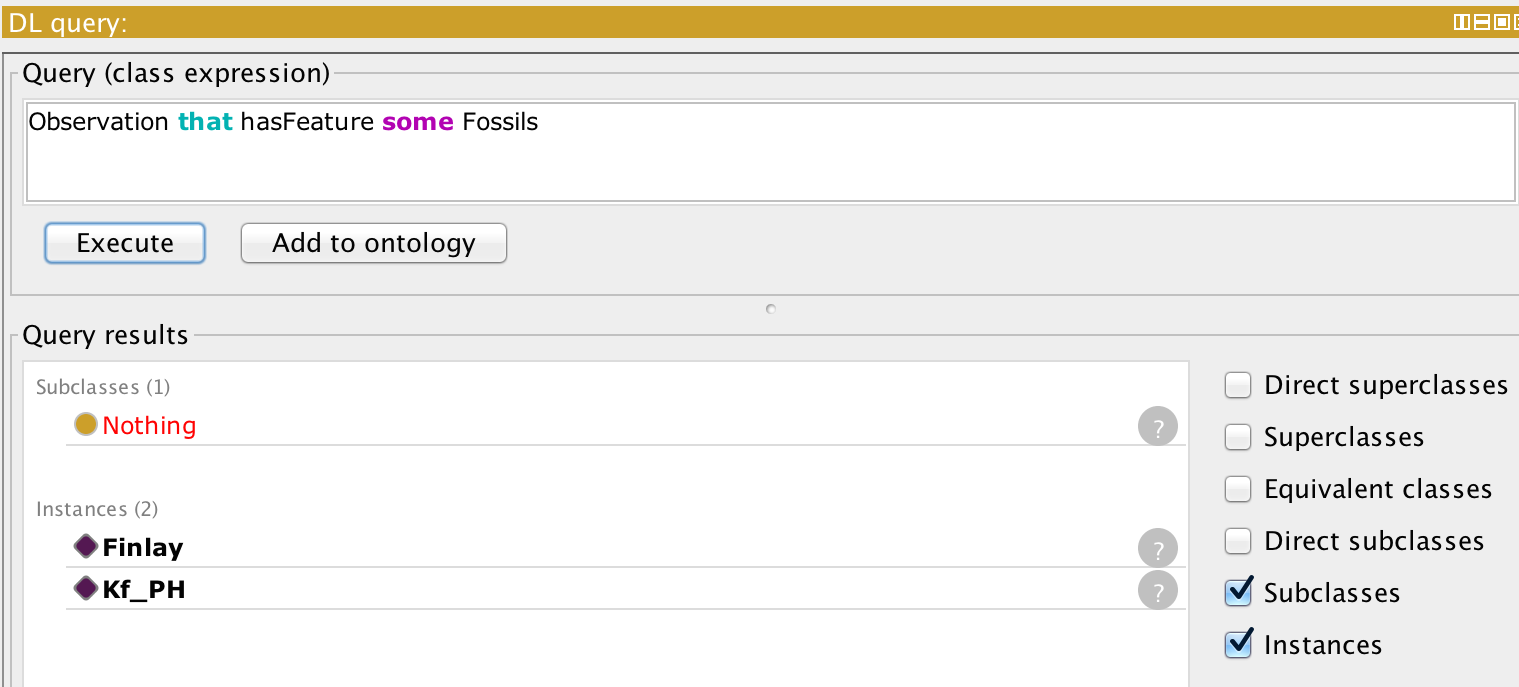
1. What other rock units are associated with this unit?



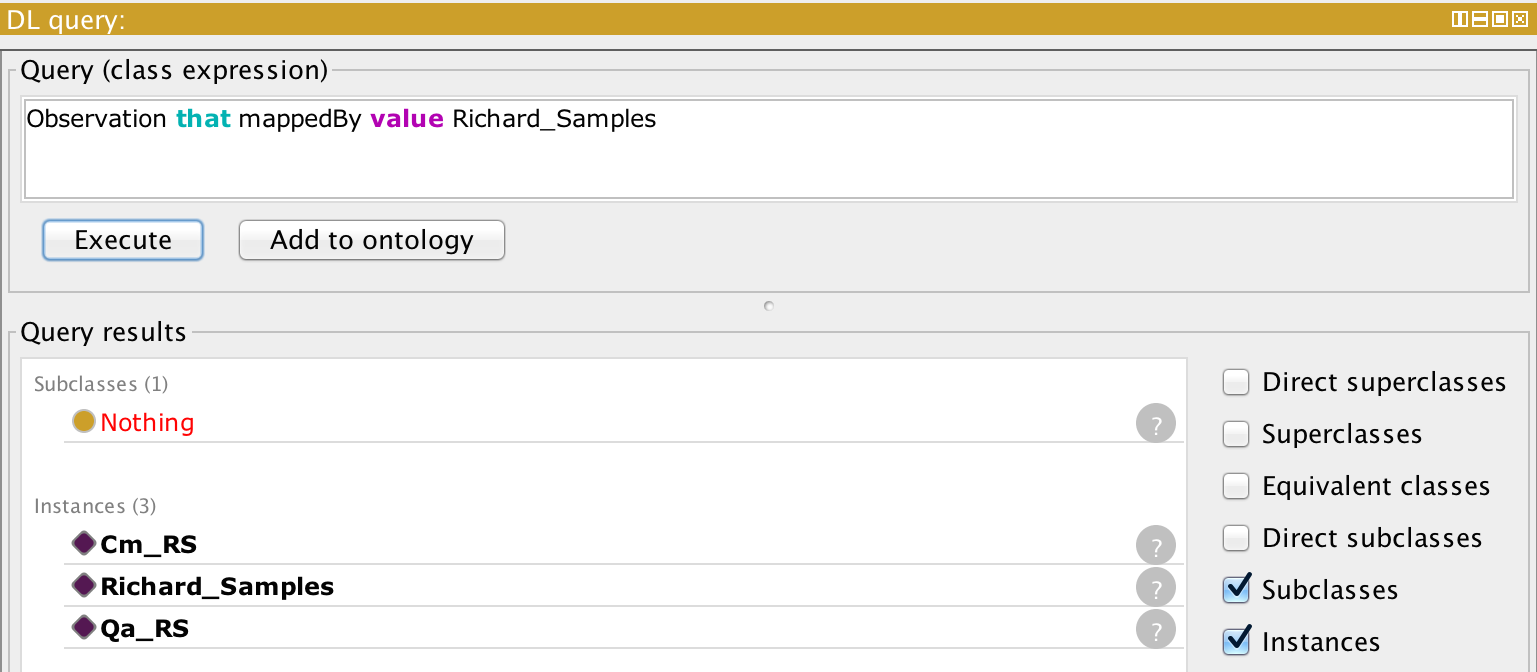
1. What other rock units are associated with this mapping unit? 
2. What is the first item on the stratigraphic column?



1. What fossils exist in this unit, or in the area?



1. What rock units have been mapped by Richard Samples?



**Team Participation**

While each team member participated in the overall creation, we did have some focused material assigned to each member.

Perry Houser

Project Concept, Database development and queries

Miguel Royo-Leon

Project Concept, Database development and integrations

Georgia Almodovar

Queries and logic build

Richard Samples

Queries and neo4j integration

# 1.7 REFERENCES

Babaie, H. (2006). Designing a modular architecture for the structural geology ontology, 269 – 282.

Baker, T. R., Battersby, S., Bednarz, S. W., Bodzin, A. M., Kolvoord, B., Sinton, D., … Uttal, D. (2014). A Research Agenda for Geospatial Technologies and Learning A Research Agenda for Geospatial Technologies and Learning. Journal of Geography, (November), 37–41. doi:10.1080/00221341.2014.950684

Berners-lee, T. I. M., Hendler, J., & Lassila, O. R. A. (2001). The Semantic Web, (May), 1–4.

Bishr, Y. (1998). Overcoming the semantic and other barriers to GIS interoperability. International Journal of Geographical Information Science, 12(4), 299–314. doi:10.1080/136588198241806

Calcagno, P., Courrioux, G., Guillen, A., Fitzgerald, D., & McInerney, P. (2006). How 3D implicit geometric modelling helps to understand geology: the 3DGeoModeller methodology. Proceedings Internationl Association for Mathematical Geology XIth International Congress, Universit, 233, 2004–2007. Retrieved from:<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:How+3D+implicit+Geometric+Modelling+Helps+To+Understand+Geology+:+The+3DGeoModeller+Methodology#0>

Feigenbaum, L. (2011). SPARQL By Example A Tutorial Why SPARQL ? Structure of a SPARQL Query.

Gomes, J., Montenegro, N., Urbano, P., & Duarte, J. (2012). A land use identification and visualization tool driven by OWL ontologies. CEUR Workshop Proceedings, 849, 1–8.

Gray, K., & Owens, K. (2011). An Exploratory Study Using Hands-On Physical Models in a Large Introductory, Earth Science Classroom: Student Attitudes and Lessons Learned. *… Journal of Science …*, *12*(2). Retrieved from http://ejse.southwestern.edu/article/download/7391/6904

Gruber, T. R. (1993). Technical Report KSL 92-71 Revised April 1993 A Translation Approach to Portable Ontology Specifications by A Translation Approach to Portable Ontology Specifications. Knowledge Creation Diffusion Utilization, 5(April), 199–220. doi:http://dx.doi.org/10.1006/knac.1993.1008

Liadey, D. M. (2012). Spatial Ontology for the Production Domain of Petroleum Geology, 116.

Lucas, S. G., Krainer, K., & Spielmann, J. A. (2010). Cretaceous stratigraphy, paleontology, petrography, depositional environments, and cycle stratigraphy at Cerro de Cristo Rey, Dona Ana County, New Mexico. *New Mexico Geology*, *32*(4), 103–130.

Movva, S., Ramachandran, R., Li, X., Khaire, S., Keiser, K., Conover, H., & Graves, S. (2005). Syntactic and semantic metadata integration for science data use. Computers and Geosciences, 31(9), 1126–1134. doi:10.1016/j.cageo.2004.12.011

Noy, N., & McGuinness, D. (2001). Ontology development 101: A guide to creating your first ontology. Development, 32, 1–25. doi:10.1016/j.artmed.2004.01.014

Sinha, a. K., Malik, Z., Rezgui, A., Barnes, C. G., Lin, K., Heiken, G., … Zimmerman, H. (2010). Geoinformatics: Transforming data to knowledge for geosciences. GSA Today, 20(12), 4–10. doi:10.1130/GSATG85A.1

Tomai, E., & Spanaki, M. (2004). From ontology design to ontology implementation : A web tool for building geographic ontologies. Language.

Raskin, R. G., & Pan, M. J. (2005). Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). Computers and Geosciences, 31(9), 1119–1125. doi:10.1016/j.cageo.2004.12.004

Rushing, J., & Conover, H. (2006). Bird ’ s-eye view of data mining in geosciences. Geological Society of America Special Papers, 397, 235–247. doi:10.1130/2006.2397

Wiegand, N., & García, C. (2007). A task-based ontology approach to automate geospatial data retrieval. Transactions in GIS, 11(3), 355–376. doi:10.1111/j.1467-9671.2007.01050.x

Wu, Q., & Xu, H. (2003). An approach to computer modeling and visualization of geological faults in 3D. Computers and Geosciences, 29(4), 503–509. doi:10.1016/S0098-3004(03)00018-9

Wu, Q., Xu, H., & Zou, X. (2005). An effective method for 3D geological modeling with multi-source data integration. Computers & Geosciences, 31(1), 35–43. doi:10.1016/j.cageo.2004.09.005