*CEE 6110 Assignment #3 Data MOdel design*

*Karun Joseph*

*A02240287*

Data model design for streamflow and water quality measurements

Introduction

Data modelling is important for loading and storage steps in the data life cycle. It provides a degree of structure, and sets various rules and conventions which would make it easier for data querying, manipulation and analysis. This report identifies and describes important entities and relationships to model streamflow and water quality measurement data in the form of a logical data model. Such a model when physically implemented would allow researchers to easily and automatically retrieve, store and manage data, and facilitate scalable software development.

Methods

Sample data and metadata files of streamflow and water quality measurements were obtained from an organization that operates a network of aquatic monitoring sites. These comma separated value files contained watershed, site, source, time series, and each time series with a set of observations for various measured variables. First a conceptual data model was developed that defines the primary entities and relationships amongst them. These were identified to be watershed, site, source, time series, variables and observations. Relational type database such as Observation Data Model (ODM) (Horsburgh et al. 2008) was the chosen logical data model type in terms of their mappings due to their software implementation capabilities and ease in capturing data semantics. Next, attributes were identified for each entity. For example, a ‘site’ entity had attributes like ‘site\_code’, site\_name’ and others. The entities and their attributes were named in such a way that they were unique, short and avoiding any reserved or special characters. A primary key was assigned to every entity that was unique and persistent. Relationships between entities were determined along with their cardinality and participation, which indicate maximum and minimum number of times an entity instance is related to the other respectively. Foreign key is inserted into a child entity to show how it is related to a parent entity. An Entity Relationship (ER) diagram is presented in Fig. 1 which represents the developed logical data model. The ER diagram was developed in MySQL workbench, a popular open source relational database management system. Entities and relationships were established with normalization in mind which minimizes redundancy and dependency. They would enable easier additions, deletions, and modifications of an entry. For example, it would be very easy to access or delete data associated with one site and its corresponding time series with a single query.

Results

The ER diagram shows a parent entity ‘watershed’ with a child entity ‘site’ related by zero to many relationship. Each ‘site’ entity has a ‘source’ parent and ‘timeseries’ child with one to one and zero to many relationships respectively. Each ‘timeseries’ child then has a ‘variable’ parent and an ‘observation’ child with one to one and one to many relationships respectively.

Conclusions

The logical model presented here provides a rigorous framework to streamflow and water quality measurement data, and would support robust data handling and agile software development. However, the data model could make future development opinionated and difficult to change if project specifications change.

References

Horsburgh, Jeffery S., David G. Tarboton, David R. Maidment, and Ilya Zaslavsky. 2008. “A Relational Model for Environmental and Water Resources Data.” *Water Resources Research* 44 (5): W05406. doi:10.1029/2007WR006392.

Appendix A: Figures

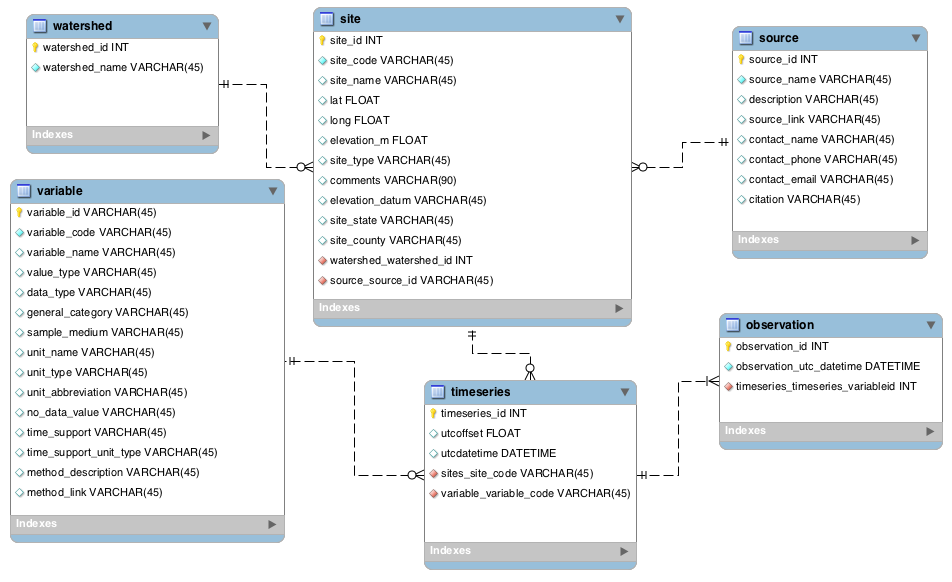


Fig 1: Entity relationship diagram