BedformsML0, a preliminary metadata language for a large, engineered and freely accessible bed form database

Ronald R. Gutierrez, Frank E. Escusa, Alice Lefebvre, Carlo Gualtieri, Francisco Nunez-Gonzalez, and Marc Roche
Pontifical Catholic University of Peru, Civil Engineering, Lima, Peru (rgutierrezl@pucp.pe)

Open and data-driven paradigms have allowed to answer fundamental scientific questions in different disciplines such as astronomy, ecology and fluid mechanics, among others. Recently, the need to collaboratively build a large, engineered and freely accessible bed form database has been highlighted as a necessary step to adopt these paradigms in bed form dynamics research.

Most large database architectures have followed the principles of relational databases model solutions (RDBMS). Recently, non-relational (NoSQL) architectures (e.g., key-value store, graph databases, document-oriented, etc.) have been proposed to improve the capabilities and flexibility of RDBMS. Both RDBMS and NoSQL architectures require designing an engineered metadata structure to define the data taxonomy and structure, which are subsequently used to develop a metadata language for data querying. Past research suggests that the development of a metadata language needs a collaborative and iterative approach.

Defining the data taxonomy and structure for bed form data may be challenging because: [1] there is not a standardized protocol for conducting field and laboratory measurements; [2] it is expected that existing bed form data have a wide spectrum of data characteristics (e.g. length, format, resolution, structured or non-structured, etc.); and [3] bedforms are studied by scientists and engineers from different disciplines (e.g., geologists, ecologists, civil and water engineers, etc.).

In recent years, several data repositories have been built to manage large datasets related to the Earth System. One of these repositories is the Earth Science Information Partners, which has proposed standards to promote and improve the preservation, availability and overall quality of Earth System related data. These standards map the roles of participants (e.g., creators, intermediaries and end users) and delivers protocols to ensure proper data distribution and quality control.

This contribution presents the first iteration of a metadata language for subaqueous bed form data, named BedformsML0, which adopts the standards of the Earth Science Information Partners. BedformsML0 may serve as a prototype to describe bed form observations from field and laboratory measurements, model outputs, technical reports, scientific papers, post processed data, etc. Biogeoenvironmental observations associated to bed form dynamics (e.g., hydrodynamics, turbulence, river and coastal morphology, biota density, habitat metrics, sediment
transport, sediment properties, land use dynamics, etc.) may also be represented in BedformsML. It could subsequently be improved in future iterations via the collaboration of professionals from different Earth science fields to also describe subaerial, and extraterrestrial bed form data. Likewise, BedformsML can be used as machine search query selection for massive data processing and visualization of bed form observations.