Volunteered Geographic Information: Level III of a Digital Earth System

Karl Grossner and Alan Glennon
Department of Geography, University of California, Santa Barbara

The term Digital Earth has come to represent a global technological initiative—in a sense, an intellectual movement. In Grossner, Goodchild and Clarke (in press), we propose renewing the process of definition and design for a particular (lower-case) *digital earth system*. The Digital Earth concept as introduced in a 1998 Al Gore speech is inclusive of the next-generation geolibrary, the global digital atlas, and to some extent, geographic information system (GIS) software. A *digital earth system* is then a hybrid of these which does not yet exist, "a distributed digital geolibrary for which the principal user interface is a global atlas, having at least some of the typical functionality of a GIS." Phrased another way, it is "a comprehensive, massively distributed geographic information and knowledge organization system."

In parsing that definition to define terms: it is *comprehensive* in that it must contain complete, "blanket" or "Level I" spatial coverage of the globe for a set of base thematic layers at a uniform scale or set of scales (*Figure 1*). Further, it will contain such additional thematic layers of georeferenced data at any scale, level of detail (LOD) or coverage extent as are made available and accepted for inclusion by expert reviewers (Level II). A third (Level III) tier of content will be un-reviewed material submitted by the global public at large—either explicitly as a candidate for Level II status or simply posted for others to view. This constitutes the *volunteered geographic information* under discussion at this meeting.

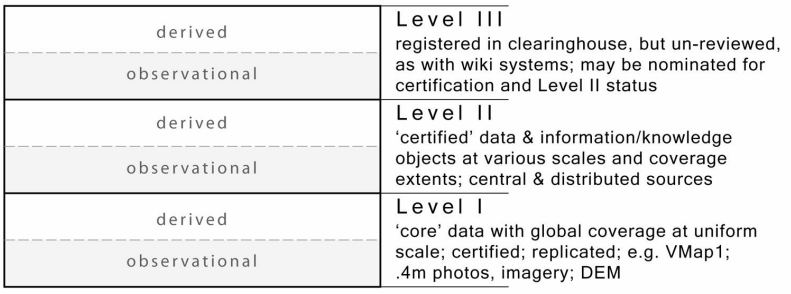


Figure 1: Tiered data source structure. Data models must account for observational data and derived information/knowledge objects being explicitly differentiated at all levels.

This digital earth system is *distributed* because, (1) there are necessarily multiple, geographically dispersed data stores providing content and (2) the processing load of server-based query and analytical processes must be shared for performance reasons (*Figure 2*).

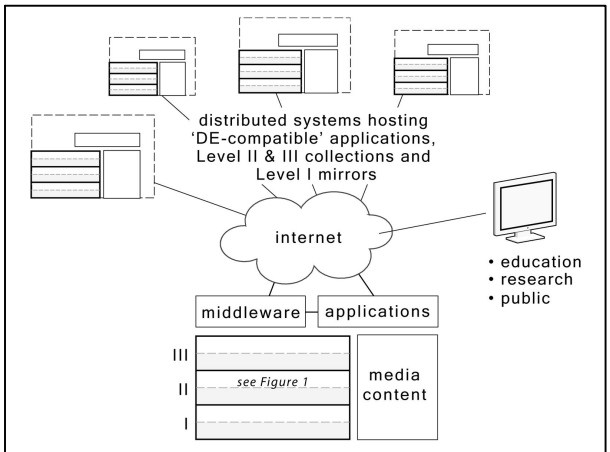
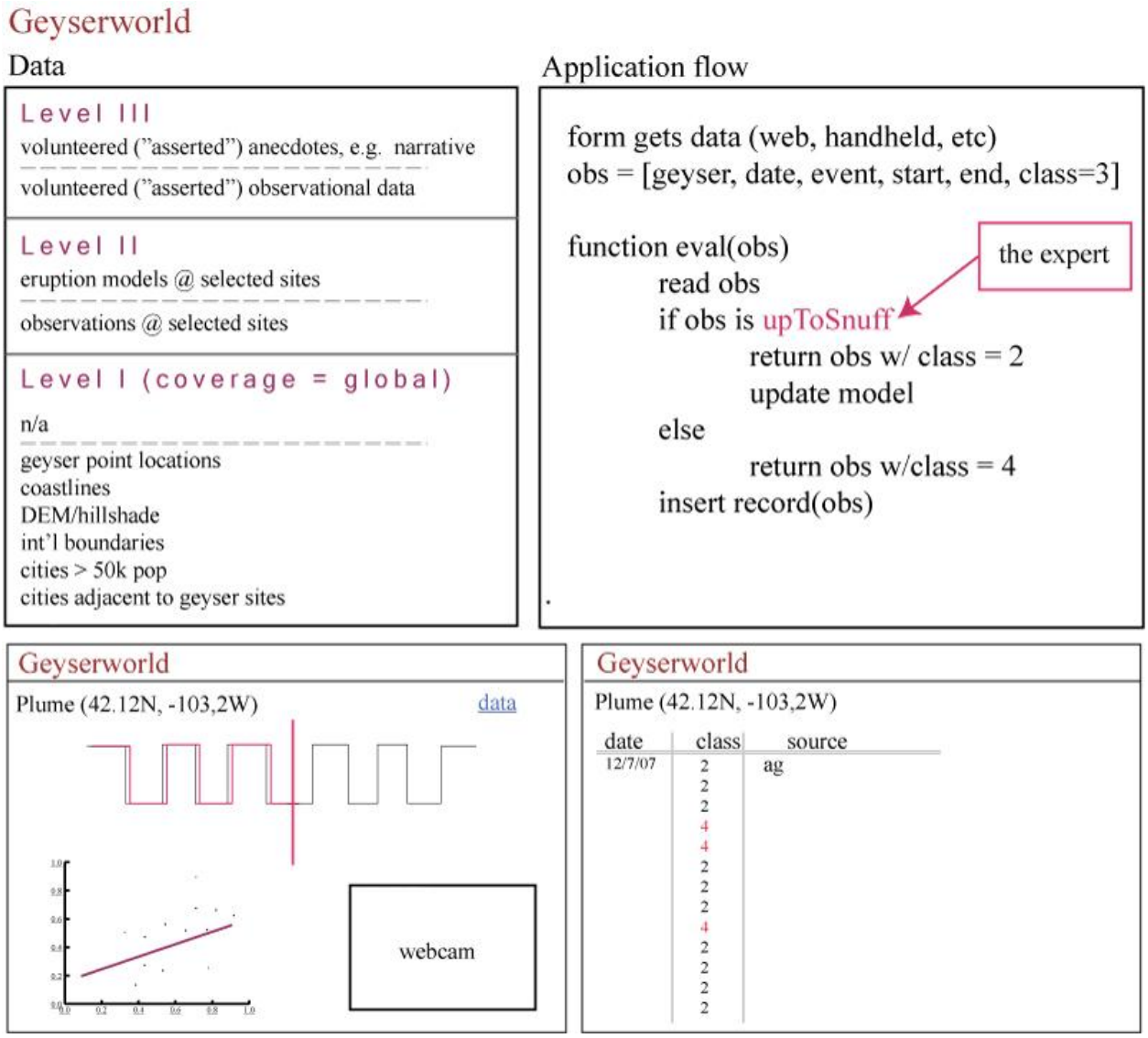


Figure 2: High-level schematic of the distributed, tiered data sources illustrated in Figure 1.

We are developing a simple instantiation of this 3-tier model using volunteered geyser observations. Field observations submitted via hand-held devices by amateurs and specialists alike are filtered by an automated "expert agent" that maintains a mathematical model of eruptive behavior for given geysers, evaluates the

volunteered observations, and classifies the values against predicted expectations. This geyser case illustrates a useful aspect of volunteered geographic information: VGI carries a temporal signature. The signature can exist as both a property of geographic phenomena and a property of data reporting itself. Analyzing VGI temporal characteristics offers an array of classification and validation mechanisms—particularly for assessing erroneous or missing data—that often may not exist for traditional geographic information. Most of the six classes of data in the three tiers are accounted for in this exemplar, as illustrated below.



References

Grossner K, Goodchild M and Clarke K (in press). Defining a Digital Earth System. *Transactions in GIS*.