CO-459

EMPLOYING GEODATABASES FOR GIS IN PLANETARY MAPPING

VAN GASSELT S.(1), NASS A.(2)

(1) Freie Universitaet Berlin, BERLIN, GERMANY ; (2) German Aerospace Center (DLR), BERLIN, GERMANY

IINTRODUCTION AND BACKGROUND

The use of Geographic Information Systems (GIS) in the field of planetary sciences, and in particular for planetary geosciences mapping has been steadily increasing after a number of new missions have been launched in the late 1990s which delivered a wealth of new data. The information basis for planetary-surface studies is broad in terms of (a) pixel scale, (b) regional extent and (c) data types and sources. Data are downloaded from individual experiment sites or central archives and are subsequently processed, calibrated, corrected and integrated into a common analysis system. Regarding the general procedure, planetary data processing does not differ significantly from earth-oriented data processing and the existence of a common data processing platform for most of the available remote-sensing data and a central data archive for localizing and retrieving data from essentially all planetary bodies simplify processing tasks considerably. Apart from a number of different and specific analysis methods, the general technical GIS-based workflow, i.e. data import, spatial referencing, creating and analyzing raster and vector data in particular in the field of surface mapping, remains identical to earth-oriented data-analysis workflows. In contrast to processing and technical workflow demands, geoscientific mapping in terms of delineation and characterization of surface units is considerably different.

GIS-based surface mapping is today usually carried out in a dissociated way by integrating raster data (topography, image, and multispectral data) from a file-based hierarchy into a GIS-based project file. Unit delineation in the course of mapping takes place on the basis of isolated vector files that are hierarchically integrated into the overall file-system structure and carried around, copied and modified for further use. For isolated research this procedure is fast and usually sufficient though it does not make use of the full potential of desktop GIS and in particular of GIS connected to a geodatabase (GDB) backend.

In the course of an ongoing effort to streamline and simplify geologic mapping for planetary bodies, we have been working on a design for a potential overarching GDB that can be easily transported, modified and integrated into a state-of-the-art GIS environment. With this model it shall be possible to merge isolated mapping efforts that have already been conducted on a desktop level into an existing GDB context. The general definition and setup of a planetary GDB and GIS-integration is envisaged to be conducted in several phases: (a) define low- and high level requirements for planetary geoscientific mapping, (b) conceptually design an overall entity relationship model in response to the requirements, (c) modify and physically implement the model to fit into a file-based GDB (FGDB) infrastructure as currently only provided by Environmental Systems and Research Institute’s (ESRI) ArcGIS environment, (d) physically implement the model into a full-scale geodatabase management system (GDBMS) backend and modify elements in order to take full advantage of specific tools, or access and query features provided by one or the other GIS system.

The overall GDB concept is complex as it has to deal with a number of different aspects: while planetary mapping and assignment of units and ancillary data in terms of relations are the major objectives, the GDB design is expected to support workflows concerned with sensor-data search and query and search of previous mapping efforts that are depicted by footprint representations and associated metadata. A number of datamodels for earth-oriented geological mapping have been defined since the late 1990s, however, they are either both too specific and granular or they do not cover specific requirements for planetary mapping making them hard to adapt and implement for planetary mapping purposes. On the other hand, we here introduce concepts and solutions for implementation that are also applicable for earth-oriented geoscientific mapping and provide additional value for conducting such mapping more efficiently. We are mainly interested in handling different surface unit types in terms of general classes, materials and their specific stratigraphic context which will be described in more detail below. Thus, reviewing terrestrial datamodels was an essential aspect in developing own strategies and adapt features already communicated to users and partially tested.

A geologic data model forms the framework for integrating spatial geologic data, measurements as well as metadata (cf. Richard, 1998). The data model establishes a link between spatial geologic units and boundaries and quantitative data as for example age measurements of rocks and surfaces. The efficient handling of any systematics (surface types, ages, morphometry) under consideration of different planetary objects is of utmost importance to access data by queries. Geologic datamodels were primarily designed and implemented for use in ESRI's commercial ArcGIS environment although some of them were intended to be created on a conceptual level in order to provide a technology-independent solution (e.g. NADM).

One of the most evolved datamodel designs is the one presented as the North American Geologic Map Data Model (NADM-C1, NADM Steering Committee, 2004) which is a descendant of the NADM 4.3 developed by the United States Geological Survey (USGS), the Association of American State Geologists (AASG) and the Geological Survey of Canada (GSC) working group (Johnson et al., 1999). It represents a highly integrated data model designed to cover all sorts of conceptual aspects related to geologic mapping in connection with in-situ data and it forms an essential part of the overarching USGS National Geologic Map Database (NGMDB, Richard et al., 2004). Later models, such as CordLink, or the data model designed by the Arizona Geologic Survey or the Central Kentucky Prototype (Richard, 1998; Richard and Orr, 2001; Soller et al., 2002; Brodaric et al., 1999), often build upon the NADM approach and used it as a conceptual fundament. However, it has never been fully physically implemented and distributed to the community.

The Geologic Mapping Template (GMT) developed by the ESRI Cartography Team (ESRI Cartography Team, 2009) focuses in particular on the implementation of the cartographic symbolization guidelines by the Federal Geographic Data Committee (FGDC, Federal Geographic Data Committee, 2008; Geologic Data Subcommittee, 2006) and builds upon the concepts developed in the framework of the NCGMP. As stratigraphic systematics and ages are also incorporated through the FGDC guidelines, the model copes with stratigraphic systematics via coded-value domains and which are hard-coded and directly related to stratigraphic systems and series for the Earth.

Both, geologic surface type assignments in connection with symbol assignments via FGDC as well as stratigraphic age relationships are not covered satisfactorily for planetary use for two major reasons: first, assigning surface types by looking up reference numbers in the FGDC standard as advised in the data model documentation are far from being intuitive and such a process easily leads to (logical) inconsistencies in the database. Secondly, hard-coded systematics and age assignments for stratigraphy are feasible if there is a common (global) system and a certain agreement on age boundaries. However, for planetary use across different planetary objects, chronologies and nomenclature differ significantly which leads to half a dozen of different age boundaries for, e.g. Mars or the Moon (e.g., Tanaka and Hartmann, 2008). Improving such an awkward treatment is one goal of this work and therefore a datamodel design taking into account planetary boundary conditions has been undertaken.

APPROACH AND CONCEPTS

The requirement to specifically define a targeted FGDB design for implementation is caused by general workflows communicated by users in the course of mapping work conducted in various environments. First of all, an ArcGIS FGDB schema definition and contents are transportable via the XMI interchange format and can thus be easily integrated into an existing mapping environment elsewhere. The concept does not allow raster data to be included within the FGDB environment for reasons of data integrity and size issues; consequently, the XMI-based schema is small in size and can be distributed and loaded easily. Secondly, the setup of a full-scale DBMS backend mostly requires interaction with system administration, thus updating the GDB requires some third-party interaction. Apart from this, a DBMS-backend is in particular feasibly for projects with multi-user access. Though envisaged on a mid-term scale (see d), we first want to focus on a solution that is easily applicable in terms of costs and time. However, ArcGIS FGDB solutions are relatively limited when it comes to querying data by attributes and topology simultaneously. Also projecting a relation on specific attributes, i.e. extracting 'columns' from a 'table', remains a tedious task as this is only manageable indirectly via reporting. Finally, the use of primary and foreign attribute keys especially when working with imported data is highly limited as important data type options are simply missing (e.g. explicit auto increments needed for primary keys) which limits the design process. Consequently, a redesign of the entity-relationships for FGDB usage has to be made to account for certain limitations and such a datamodel is not necessarily free of redundancy.

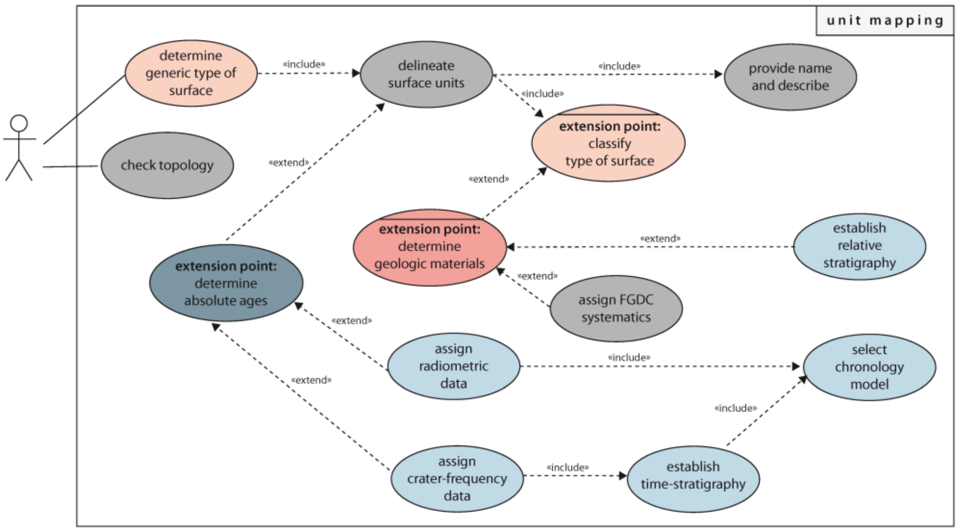

Figure 1: Use-cases for working with stratigraphic and surface material assignments in planetary geologic mapping as further detailed in figure 3 and 4.

The basic use-case scenario follows the outline highlighted in figure 1. This (actual) mapping use-case depicts a central part of the overall datamodel design and is separated into a number of subcomponents as represented by additional tasks. These use cases implicitly define requirements that need to be respected in the course of the design approach and which affect the external as well as the conceptual layer.

In order to follow the requirements discussed below, some of the basic vocabulary needs to be established and described in the light of the importance for planetary mapping. This vocabulary is related to so-called stratigraphic aspects in general which are the fundamental basis for geologic mapping. A mapped surface unit usually owns a number of attributes that need to be provided on either a detailed (high-) or outline (low-) level by the datamodel relations and attribute values.

(1) A unit can either be a geologic unit, a geomorphologic unit or it can be both, geologic and geomorphologic. While a geologic unit is described by materials deposited or emplaced within a distinct time span by a distinct process, a geomorphologic unit is generally described by the shape and surface expression of landforms. (2) Geologic units emplaced within a distinct time-span have a certain so-called stratigraphic relationship to underlying and overlying units. Materials deposited on top of a unit are younger (deposited later), thus a relative stratigraphic timing can be established. If this stratigraphy is put into an absolute context, either by the method of impact crater size frequency statistics from orbital data or by radiogenic isotope measurements of samples returned from the Moon, a unit can be described chronostratigraphically (based on a planetary chronology model). Each method's characteristic attributes must be included in the datamodel design. (3) Any chronostratigraphic systematics relies on a chronology which defines time boundaries that are established as reference. For terrestrial planetary bodies, chronologies exist in a number of flavors and model-dependent time boundaries. Such time boundaries and hierarchically described systematics must be properly depicted in the data model. In short, each unit, represented as polygon feature class, can have multiple assignments of surface types: (1) general unit type: geologic or geomorphologic; (2) unit material: e.g. sandstone, limestone, volcanic rocks which are linked to a distinct process; (3) chronostratigraphic context for each planetary body and with associated age boundaries and chronology models: e.g., Eratosthenian system and Upper Eratosthenian series for the Moon or the Cretaceous system with the Upper Cretaceous time-stratigraphic series for the Earth (figure 2).

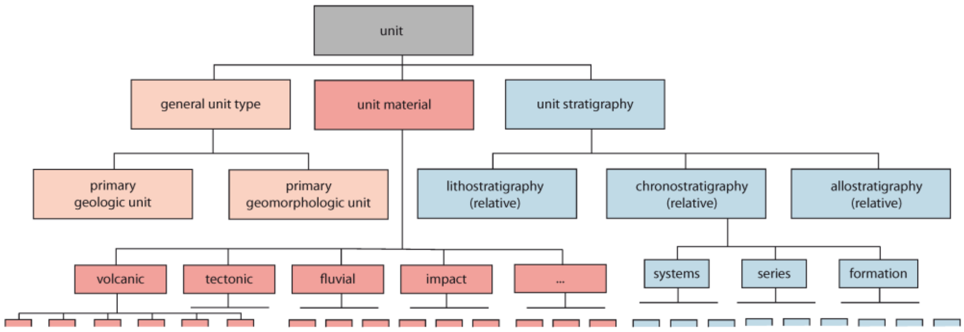

Figure 2: Object-class diagram depicting surface-unit types (general, material and stratigraphy) as highlighted in figure 3.

The mapping component consists of subcomponents that deal with the following low-level objectives:

(a) The datamodel must cope with different planetary bodies to be mapped geoscientifically. It must provide (relational) interfaces to adapt mapping contents and body-specific map-unit vocabulary (in particular stratigraphic nomenclature and associated systematics in terms of time-stratigraphic units and rock-stratigraphic units).

(b) Apart from any stratigraphic systematics, mapped units are defined by their geoscientific type, i.e. it has to be defined in how far a mapped unit is a geologic or a geomorphologic unit. This classification enables the proper use of topology constraints. It is, for instance, possible to put a geomorphologic unit on top of a geologic one but it must not be possible to stack two geological units. The proper use of geologic vs. geomorphologic mapping is an issue that has been widely discussed since the early years of planetary mapping and remains still unsolved today.

(c) The datamodel must be expandable with respect to definitions of planet-specific attributes and attribute values. This comprises not only stratigraphic systematics (see 1), but in particular the geodetic characterization of reference bodies and specific units and feature types.

(d) Earlier and model-independent (termed dissociated) mapping efforts must be integrable into the existing data model by directly importing attribute values into existing attributes or by adding relations and attributes for ancillary data such as user-defined measurements that are not modeled a-priori.

(e) It must not be possible to enter attribute values being thematically dependent on each other in an independent way, i.e. without proper domain control and without establishing a relationship control.

The high-level objectives comprise a large number of items of which the following are considered the most important ones.

(f) For delineation of map units, the mapper can choose between different thematic map components for representing specific contents, e.g. mappers can choose to focus mapping on geomorphology and/or geology.

(g) For delineation of map units, it must be possible for each mapper to select and apply unit definitions described in the standard document of the Federal Geographic Data Committee (FGDC) on cartographic representations. As an additional requirement the assignment should by based upon a hierarchical structure defined in the datamodel which enables the mapper to select surface unit types by their names rather than by their FGDC numbers. FGDC-based systematics must be planet-specific in cases where such definitions can be established and they must be expandable.

(h) For entering attribute values with respect to the planetary chronology and chronostratigraphy, the variety of different planetary chronology models as well as radiometric ages of in-situ samples must be properly stored and accessible.

An additional requirement is that each mapping component must be integrable into the overarching datamodel design via appropriate and expandable interfaces. This ensures that an update of subcomponents can easily be integrated into an existing datamodel without having to update all sub components. For testing purposes the component must work in a stand-alone way which also means that a certain redundancy in terms of relations on the FGDB-level is needed. For DBMS-backend solutions such redundancy must be avoided in order to achieve at least the first normal form (1NF) and especially avoid nullable attributes.

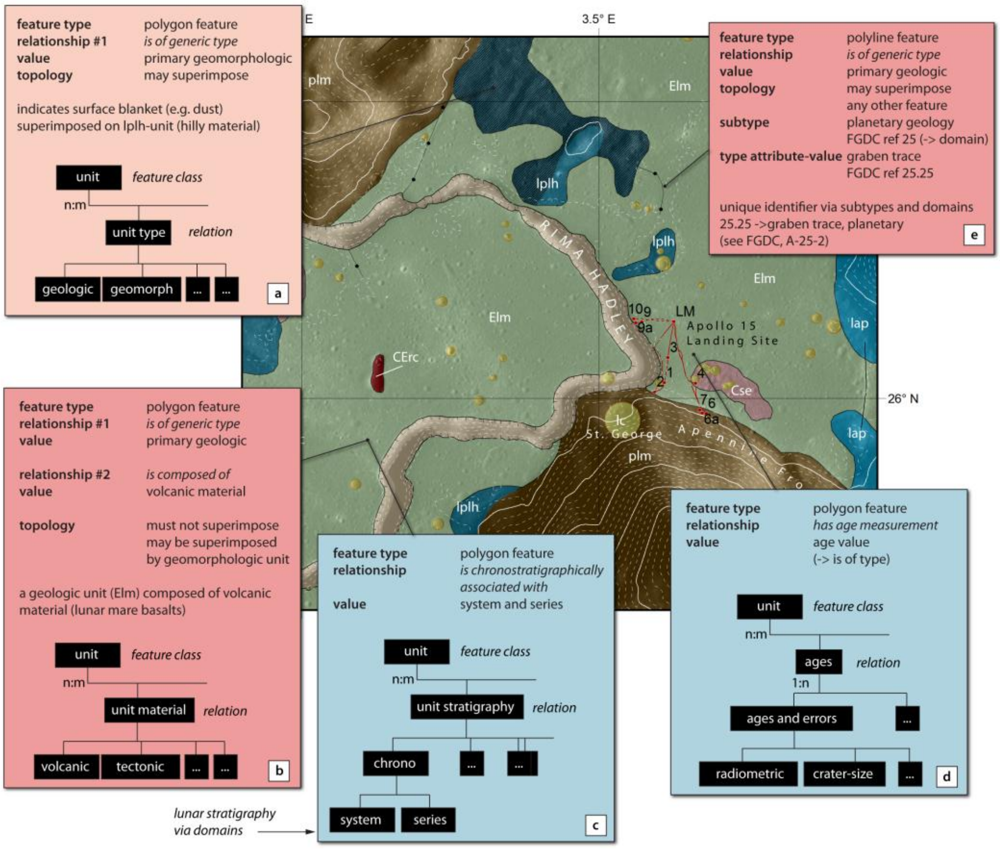

Figure 3: Geologic map and associated model components for the lunar Apollo 15 landing site at Hadley rille. Color-coding is the same as for the object-class diagram in figure 2 to improve legibility.

The scenario (figure 3) shows the conceptual approach in response to the requirements stated in the use-case scenario (figure 1) and by integrating the three main unit types (1) general type, (2) material, (3) stratigraphy as outlined above and in figure 2. Each of these subclasses (though not modeled as such due to limitations in the relational base concept) is modeled with its associated ER branch. This leaves room for modifying and updating individual ER branches without having to touch respective ER interfaces. The solution to (a) maintaining topologic integrity, (b) allowing to intuitively assign FGDC-based geologic classes and associated symbols and (c) provide an expandable stratigraphic framework while limiting erroneous data entries is by generating two surface type classes treated separately within the topology rule set, and by establishing a detailed level of domains controlled by subtypes. A test case with explanation on the FGDB implementation is discussed hereafter.

IMPLEMENTATION AND TEST CASE

The ER model re-designed for FGDB-use has been implemented in ESRI's ArcGIS environment and has been tested for efficiency in terms of storing and accessing attribute values. Since spatial combined with attribute querying is rather limited in the case of FGDBs, an FGDB-adapted design of relationship classes should do the work to link entities so that they become directly visible upon feature selection. As highlighted in figure 2, a surface unit represented as simple feature class (either polygon or polyline) can be modeled as either geologic or geomorphologic unit which opens up two different branches to be followed within the datamodel. While symbology assignments affect geologic as well as geomorphologic units, stratigraphic systematics affect geologic objects only.

For a demonstrator test case, we used the lunar Apollo 15 landing site scenario at Hadley Rille as at this site (a) crater-size frequency data for age determinations from remotely sensed orbital observations as well as radiometric ages from lunar surface samples are available (along with associated sampling locations), (b) the landing site area can be well characterized geologically through earlier mapping and a detailed knowledge by in-situ observations and recent re-mapping, and (c) Hadley Rille is a geomorphologic unit and represents a landform that is cut into lava units, i.e. is it mapped as simple feature class polygon topologically superimposed on a geologic unit.

Some of the essential components, in particular those answering the aforementioned requirements are discussed hereafter; the complete datamodel can be obtained from the authors.

- Differentiating between Unit Types: surface units mapped as simple feature classes in a direct or indirect way via feature class conversions, do always belong to a geoscientific type unit. The selection of a type is obligatory and the proper assignment is required for a working topology framework. Geological units cannot superimpose each other, while geomorphological units can superimpose themselves or geological units (e.g. an aeolian dust blanket can be superimposed on a geologically defined lava plain material, see figure 3a-b). The differentiation is not always unambiguously possible so that an mixed surface type is introduced. Rather than working with subtypes or domains, a type unit is more efficiently modeled as an additional relation with an (1,*):(1,*) relationship class "unit is of generic type" to allow later expansion and multiple assignments (see figure 3).

- Working with Genetic Classes and Standard Symbols: once a generic type has been assigned, a number of more specific types as defined by, e.g. the FGDC or any other employed standard can be assigned. The FGDC standards provide a systematic framework for assigning specific surface types by hierarchical methods, i.e. a volcanic surface type is separated into different other volcanic subtypes. While most subdivisions are defined on a genetic level (volcanic, tectonic, aeolian features), planetary symbologies are treated on a topical level including all different sorts of planetary symbols from a variety of planetary objects. The FGDC standard does not provide separate systematics for geologic and geomorphologic units. Although there is a specific planetary symbol set (appendix 25 of the FGDC) which covers specific geologic and geomorphological landforms found on certain planetary objects only (such as coronae structures on Venus), other more general landforms have to be derived from other appendix chapters that are derived from terrestrial geosciences. For this implementation, each type of landform is separated into additional subtypes (see figure 2). Rather than searching through the print-out FGDC for finding a correct landform type and its subtypes represented as numeric codes, landforms are modeled as subclasses with associated coded values representing the FGDC code. Once a type is selected individual domains control the possible selection of entries on the subtype level. These, again are associated with the FGDC-specific code that can be worked with on the symbology level (see figure 3e). The full FGDC code associated with a specific symbol is built from the subtype code and a coded-value domain. As an example (see figure 3): if a landform is mapped that has been identified as volcanic feature, the mapper selects volcanic landforms from the surface-type subclass. The subtype code 18 corresponds to the respective FGDC's appendix and is automatically assigned. Once this is done, the associated domain will allow to only selecting volcanic landforms as summarized in the FGDC standard. When the mapper selects, e.g. a rootless vent area from the surface sub-type menu, the coded value domain returns the value 60. In the symbology menu, subtype code 18 and domain code 60 provide the proper FGDC code "18.60" under which the associated symbol can be found and which can be directly parsed to the GIS via this artificial aggregate key. This approach allows a more natural behavior during mapping conduct as symbols do not have to be defined based on the FGDC lookup code but by selecting and sub-selecting landforms that are actually seen. The technical work and the automatic assignments are dealt with on the physical level. Symbols can then be stored in a relation with primary keys defined by aggregate FGDC codes.

- Assigning Stratigraphic Systematics: the handling of stratigraphic and age information is relatively complex and the FGDB design does not allow to design a high level of nested relations without incorporating problems with respect to information integrity. The aim is to avoid as much redundancy as possible but in order to allow a broad treatment and expandable base-level design, some information needs to be duplicated in the model. For DBMS-driven GIS usage, such redundancy can be eliminated.

Beside surface type information each unit is related to (a) stratigraphic systematics on a relative and absolute scale (see figure 2 and 3c-d). In order to provide proper naming and association of chronologic systems and series, an external relation consisting of attributes for planetary stratigraphic systems needs to be incorporated (see figure 4). Geologic units, such as the mare basalt plains of the Apollo 15 landing site (figure 3) can be further characterized by ages measured at certain locations. These measurements are radiometric (relation termed radiometric in figure 4) or based upon crater-size frequency data (relation termed craterfreq in figure 4). Further relationship cascades allow defining exact locations, age uncertainties and other details. The chronology, including chronostratigraphy is modeled as a different branch that is directly related to a surface unit. More specific, although age determinations might be available, the chronostratigraphic assignment is modeled independently. Such a putative drawback needs to be implemented as chronostratigraphic systematics are highly model-dependent and in contrast to terrestrial geosciences, planetary geology is often characterized by many different chronology models with different age boundaries for specific systems and series. An absolute age determined via radiometric or crater-size frequency analysis therefore does not necessarily lead to unambiguous chronostratigraphic systematics. It has, however, to be kept in mind, that without proper knowledge of applied chronology models, the database might become inconsistent as cross-checking mechanisms are not possible within an FGDC framework.

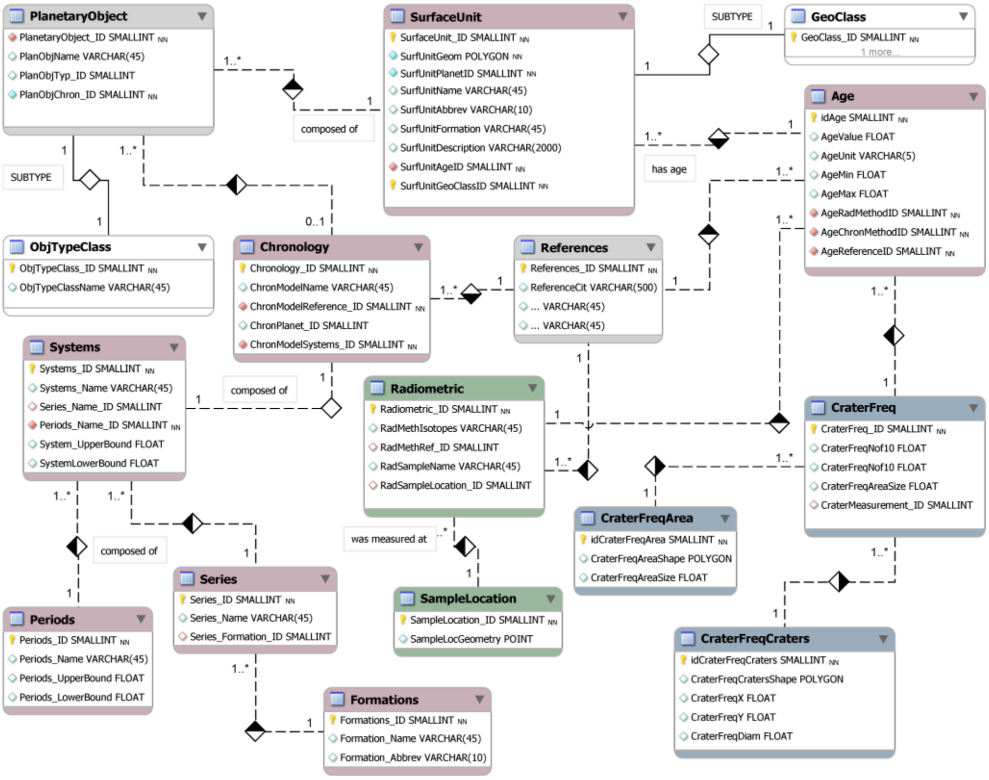

Figure 4: Entity relationship diagram depicting relation and relationship classes dealing with chronostratigraphic assignments on an absolute chronostratigraphic scale via radiometric and impact-crater-statistical analyses as well as on a relative scale as defined by superposition of surface units.

CONCLUSIONS AND OUTLOOK

The data model presented in this work has been tested for a geologic mapping framework which consists of all items encountered during geologic and geomorphologic mapping of planetary objects. The model was evaluated by treating and testing individual requirement branches as formulated in the use-case scenarios. The currently realized approach is designed as a modular approach but suffers from limitations of the general FGDB framework. Transport to other ArcGIS desktop environments via XMI is possible without any major drawbacks; however, portability to other GI systems suffers from different approaches in object-relational modeling capabilities.

The model has been designed to work independently of the mapped planetary object on an abstract level, but in order to be applicable to a broad range of different mapping projects, the detailed FGDC standard as well as stratigraphic and chronology systematics need to be incorporated. Along with this implementation, a solution needs to be found to efficiently treat historic chronologies and naming conventions as well as earlier maps and unit assignments that were superseded in the course of ongoing planetary mapping programs. Although in principle easy to realize the legibility of the data model might suffer from the high degree of granularity so that an intermediate solution must be found.

An integrated model version 1 is going to be released at the end of the year along with a manual and technical documentation regarding expansion capabilities. The FGDB implementation is a targeted design with a high degree of transportability at the cost of some data model and query possibilities. A DBMS-driven framework is currently developed in parallel to assess its usability. Beside the FGDB-based model, a DBMS-framework will be released for public testing as well.

REFERENCES

• Richard, S. M., 1998. Digital geologic database model. Technical Report, ”Draft 1.0”, Arizona Geological Survey.

• NADM Steering Committee, 2004. North American Geologic Map Data Model (NADM) Conceptual Model 1.0: a conceptual model for geologic map information. Technical Report U.S. Geological Survey Open-File Report 2004-1334, U.S. Geological Survey.

• Johnson, B. R., Brodaric, B., Raines, G. L., Hastings, J. T. and Wahl, R., 1999. Digital geologic map data model – version 4.3. Technical report, United States Geological Survey (USGS).

• Richard, S., Craigue, J. and Soller, D., 2004. Implementing NADM C1 for the national geologic map database. In: Digital Mapping Techniques ’04 – Workshop Proceedings, pp. 107–140.

• Richard, S. M. and Orr, T. R., 2001. Data structure for the Arizona Geological Survey geologic information system: Basic geologic map data. In: Digital Mapping Techniques ’01 – Workshop Proceedings, U.S. Geological Survey Open-File Report.

• Soller, D. R., Brodaric, B., Hastings, J. T., Wahl, R. and Weisenfluh, G. A., 2002. The Central Kentucky Prototype: An object-oriented geologic map data model for the national geologic map database. U.S. Geological Survey Open-file Report 02-202, United States Geological Survey (USGS).

• Brodaric, B., Journeay, J. and Talwar, S., 1999. Using the proposed U. S. National Digital Geologic Map Data Model as the basis for a web-based geoscience library prototype. In: Digital Mapping Techniques ’99 – Workshop Proceedings, U.S. Geological Survey Open-File Report.

• ESRI Cartography Team, 2009. Geologic mapping template. Technical report, ESRI.

• Federal Geographic Data Committee, 2008. Geographic information framework data content standard - part 0: Base document.

• Tanaka, K. L. and Hartmann, W. K., 2008. Planetary time scale. In: The Concise Geologic Time Scale, Cambridge University Press, Cambridge, pp. 13–22.

• Geologic Data Subcommittee. 2006. FGDC Digital Cartographic Standard for Geologic Map Symbolization, Federal Geographic Data Committee (FGDC), FGDC-STD-013-2006.