INFORMATICA, 2017, Vol. 28, No. 1, 45–78

DOI: http://dx.doi.org/10.15388/Informatica.2017.120

A Model for Multigranular Data and Its Integrity

Stephen J. HEGNER[1]*, M. Andrea RODRÍGUEZ[2]
[1] *Umeå University, Department of Computing Science, SE-901 87 Umeå, Sweden*
[2] *Departamento Ingeniería Informática y Ciencias de la Computación*
Edmundo Larenas 219, Universidad de Concepción, 4070409 Concepción, Chile
e-mail: hegner@cs.umu.se, andrea@udec.cl

Received: December 2016; accepted: February 2017

Abstract. Data involving spatial and/or temporal attributes are often represented at different levels of granularity in different source schemata. In this work, a model of such multigranular data is developed, which supports not only the usual order structure on granules, but also lattice-like join and disjointness operators for relating such granules in much more complex ways. In addition, a model for multigranular *thematic attributes*, to which aggregation operators are applied, is provided. Finally, the notion of a thematic multigranular comparison dependency, generalizing ordinary functional and order dependencies but specifically designed to model the kinds of functional and order dependencies which arise in the multigranular context, and in particular incorporating aggregation into the definition of the constraint, is developed.

Key words: relational, database, multigranular, integrity.

1. Introduction

The most important type of integrity constraint in both the theory and the practice of relational database systems is undoubtedly the functional dependency, or FD for short (Maier, 1983. Ch. 4). As a concrete example, consider the relational schema $R_{\mathsf{sumb}}\langle A_{\mathsf{Plc}}, A_{\mathsf{Tim}}, B_{\mathsf{Bth}}\rangle$, in which A_{Plc} represents places, A_{Tim} represents time, and B_{Bth} represents a number of births. The FD $\{A_{\mathsf{Plc}}, A_{\mathsf{Tim}}\} \to B_{\mathsf{Bth}}$ asserts that place and time determine the number of births; if there are two tuples of the form $\langle p, t, b\rangle$ and $\langle p, t, b'\rangle$, then $b = b'$.

The attributes of this schema have additional, multigranular properties. Places, for example, may be at the level of cities (e.g. the Chilean city $\mathsf{\textit{Concepción}}$) as well as at the level of national regions (e.g. $\mathsf{\textit{Región VIII}}$, in which Concepción lies). This induces an ordering defined by spatial inclusion: $\mathsf{\textit{Concepción}} \sqsubseteq \mathsf{\textit{Región VIII}}$. Similarly, intervals of time have an ordering; if $\mathsf{\textit{Q2Y2014}}$ represents the second quarter of year 2014, while $\mathsf{\textit{Y2014}}$ represents the entire year, then $\mathsf{\textit{Q2Y2014}} \sqsubseteq \mathsf{\textit{Y2014}}$. Thus, if $\langle \mathsf{\textit{Concepción}}, \mathsf{\textit{Q2Y2014}}, n_C\rangle$ and $\langle \mathsf{\textit{Región VIII}}, \mathsf{\textit{Y2014}}, n_B\rangle$ are two tuples in this relation, with n_C and n_B representing the number of births in each case, then it must be the case that $n_C \leqslant n_B$. This is a manifestation of the *order dependency* $\{A_{\mathsf{Plc}}, A_{\mathsf{Tim}}\} \overset{\leqslant}{\to} B_{\mathsf{Bth}}$,

*Corresponding author.

Part 1			Part 2		
A_{Plc}	A_{Tim}	B_{Bth}	A_{Plc}	A_{Tim}	B_{Bth}
Región_I	*Q1Y2014*	n_1	*Chile*	*Q1Y2014*	b_1
Región_II	*Q1Y2014*	n_2	*Chile*	*Q2Y2014*	b_2
...	...	...	*Chile*	*Q3Y2014*	b_3
Región_XV	*Q1Y2014*	n_{15}	*Chile*	*Q4Y2014*	b_4

Fig. 1. Two databases with the same schema

which asserts that for any two tuples $\langle p_1, t_1, b_1\rangle$ and $\langle p_2, t_2, b_2\rangle$, if $p_1 \sqsubseteq p_2$ and $t_1 \sqsubseteq t_2$, then $b_1 \leqslant b_2$ (Ginsburg and Hull, 1983; Ng, 2001; Szlichta *et al.*, 2012).

For the most part, previous work on multigranular attributes has focused on such subsumption (order) structure (Camossi *et al.*, 2006; Rodríguez and Bravo, 2012; Bravo and Rodríguez, 2014). There are, however, important kinds of constraints in the multigranular context which cannot be represented solely via order dependencies. As a concrete example, for A_{Plc}, it is also possible to assert that Chile is composed of exactly fifteen nonoverlapping regions via a join-like rule of the following form:[1]

$$\mathit{Chile} = \bigsqcup_{\perp}\{\mathit{Región\_R} \mid I \leqslant R \leqslant XV\}. \qquad \text{(r-Chile)}$$

The symbol $\bigsqcup_{\perp}$ means that its arguments join disjointly; that any pair $\{\mathit{Región\_i}, \mathit{Región\_j}\}$ with $i \neq j$ is disjoint; i.e. nonoverlapping spatially.[2]

To illustrate the particular issues which arise in the multigranular framework, consider the two databases shown in Fig. 1. For convenience, it is shown in two parts, with the first part containing tuples of $R_{\mathsf{sumb}}\langle A_{\mathsf{Plc}}, A_{\mathsf{Tim}}, B_{\mathsf{Bth}}\rangle$ with values of A_{Plc} at the granularity of regions, and the second with tuples with values of A_{Plc} at the granularity of countries. The semantics implied by (r-Chile) require that the sum of the number of births over the regions for *Q1Y2014* agree with the value for all of Chile; that is, $b_1 = \sum_{i=1}^{15} n_i$. The main contribution of this paper is to provide a model of data granules which supports rules such as (r-Chile) succinctly, including formulations of not only basic (often spatio-temporal) attribute such as A_{Plc} and A_{Tim}, but also models of *thematic attributes* such as B_{Bth}, including how they may be embellished with aggregation operators. Finally, a means to employ these concepts in the expression of such integrity constraints is provided.

Multigranular data often arise when monogranular data from different sources, at different granularities, are to be combined. For example, Part 1 and Part 2 of the relation of Fig. 1 might have come from distinct, monogranular sources. Thus, the issues considered here might be recast as a restricted form of a data integration problem (Lenzerini, 2002), in which all relations to be integrated are assumed to have the same structure; only

[1] Actually, there is no Región XIII; it is called Región RM; this detail is ignored here.

[2] Work such as Egenhofer and Franzosa (1991) (using topology) and Randell *et al.* (1992) (using a special logic) includes enough structure to be able to formulate the notion of connectedness, so that touching without sharing interior points may be distinguished from sharing interior points. In this work, there is no such topology or logic, and so no way to distinguish these. Therefore, *nonoverlapping* here means simply that the regions do not share points.

the granularities may differ. There are many other important issues surrounding the problem of data integration, but in order to focus upon multigranular constraints, they are not considered.

The work reported here is based upon the conference paper Hegner and Rodríguez (2016). However, much, if not most, of the framework of that earlier paper has been reworked entirely. In particular, an approach to modelling granularity constraints (such as (r-Chile)) which is allied much more to techniques of mathematical logic has been employed. In addition, the way in which aggregation is modelled has been much improved, to allow ordinary aggregation operators rather than requiring ones specialized for the multigranular context.

2. Background Concepts and Notation

In this section, notation and terminology regarding mathematical and database-related topics which are used throughout the paper are collected.

2.1. *Special Notation*

Some notation which is not completely standard or may not be known to all readers is collected here. $X \subseteq_f Y$ means that X is a *finite* subset of Y. $\mathbf{2}^X$ denotes the *powerset* of X; i.e. the set of all subsets of X. The backslash symbol denotes set difference: $S_1 \setminus S_2 = \{x \in S_1 \mid x \notin S_2\}$. For functional composition, $(f \circ g)(x)$ means $f(g(x))$. In other words, in $f \circ g$, application is right to left.

$\mathbb{Z}$ denotes the set of integers, $\mathbb{N}$ the set of natural numbers (nonnegative integers), and $\mathbb{R}$ denotes the set of real numbers. Unless otherwise stated, intervals always consist of integers. Thus, $[x, y]$ denotes $\{z \in \mathbb{Z} \mid x \leqslant z \leqslant y\}$. The *clopen interval* $[x, y)$ is $\{z \in \mathbb{Z} \mid x \leqslant z < y\}$. If another base set is used, it is shown explicitly; e.g. $[x, y]_{\mathbb{R}}$ denotes $\{z \in \mathbb{R} \mid x \leqslant z \leqslant y\}$.

Given an equivalence relation $\equiv$ on a set P and $p \in P$, $[p]_{\equiv}$ denotes the equivalence class of $\equiv$ containing p. Furthermore, $\mathsf{Blocks}\langle\equiv\rangle$ denotes the set of all equivalence classes (or *blocks*) of $\equiv$. When the context is clear, $[p]_{\equiv}$ may be shortened to just $[p]$.

2.2. *Order Structures*

Familiarity with notions such as *partial order* and *Boolean algebra*, as covered in Davey and Priestly (2002) are assumed; only terminology and special notation are presented here.

A *preorder* is a pair $\mathbf{P} = (P, \leqslant_{\mathbf{P}})$ in which P is a set and $\leqslant_P$ is a relation which is reflexive and transitive, but not necessarily symmetric. An equivalence relation $\equiv_{\mathbf{P}}$ may be defined on P with $x \equiv_{\mathbf{P}} y$ equivalent if $x \leqslant_P y \leqslant_P x$. The pair $[\mathbf{P}] = (\mathsf{Blocks}\langle\equiv_{\mathbf{P}}\rangle, \leqslant_{\mathbf{P}})$ then becomes a partial order with $[x] \leqslant_{[P]} [y]$ if $x \leqslant_P y$.

A poset is *upper bounded* if it has the greatest element, denoted $\top_{\mathbf{P}}$, and *lower bounded* if it has the least element, denoted $\bot_{\mathbf{P}}$. An *upper bounded poset* $\mathbf{P} = (P, \leqslant_P, \top_P)$ and a *bounded poset* $\mathbf{P} = (P, \leqslant_P, \bot_P, \top_P)$ are defined in the obvious way. More generally, within a poset or preorder $\mathbf{P}$, $\mathsf{LUB}_{\mathbf{P}}\langle S\rangle$ (resp. $\mathsf{GLB}_{\mathbf{P}}\langle S\rangle$) denotes the least upper bound (resp. greatest lower bound) of the set S, when it exists.

A preorder has these properties precisely in the case that the associated poset of equivalence classes has. In particular, an LUB or GLB may consist of a set of equivalent elements.

A *Boolean algebra* is denoted $\mathbf{L} = (L, \vee_{\mathbf{L}}, \wedge_{\mathbf{L}}, \complement_{\mathbf{L}}, \bot_{\mathbf{L}}, \top_{\mathbf{L}})$, with, in particular, $\complement_L$ the complement operation. For join $\vee_{\mathbf{L}}$ and meet $\wedge_{\mathbf{L}}$, the larger versions $\bigvee_{\mathbf{L}}$ and $\bigwedge_{\mathbf{L}}$ are used for these operations on sets. Thus, for example, for $S = \{x_1, x_2, \ldots, x_n\} \subseteq L$, $\bigvee_{\mathbf{L}} S = x_1 \vee_{\mathbf{L}} x_2 \vee_{\mathbf{L}} \ldots \vee_{\mathbf{L}} x_n$.

2.3. *Multisets, and Functions*

A *multiset* (or *bag*) is just like a set, except that a multiset may have several occurrences (a finite number) of the same element. To distinguish them from ordinary sets, multisets are written using multiset brackets; for example, $\wr a, a, b, c, c, c \int$ denotes a multiset with two occurrences of a, one of b, and three of c. Given a multiset X, $\mathsf{SetOf}\langle X\rangle$ denotes the underlying set; e.g. $\mathsf{SetOf}\langle \wr a, a, b, c, c, c \int\rangle = \{a, b, c\}$.

A multiset is represented as a set by tagging the elements with positive integers; for example, $\wr a, a, b, c, c, c \int$ is represented by $\{\langle a, 1\rangle, \langle a, 2\rangle, \langle b, 1\rangle, \langle c, 1\rangle, \langle c, 2\rangle, \langle c, 3\rangle\}$. When defining functions between multisets, it is generally necessary to use such tagged versions, and properties are those of the underlying sets. Thus, to say that $f : \wr a, a, b \int \to \wr c, c, c \int$ is injective means that each of $\langle a, 1\rangle$, $\langle a, 2\rangle$, and $\langle b, 1\rangle$ map to distinct element of $\{\langle c, 1\rangle, \langle c, 2\rangle, \langle c, 3\rangle\}$. Note that there can be an injective function $f : X \to Y$ without any injective function $f' : \mathsf{SetOf}\langle X\rangle \to \mathsf{SetOf}\langle Y\rangle$.

For a more comprehensive treatment of multisets and function, see Girish and John (2009).

2.4. *Fields of Sets as Boolean Algebras*

Call a set $\mathbf{T}$ a *set of subsets* if $\mathbf{T} \subseteq \mathbf{2}^S$ for some set S with $S \in \mathbf{T}$. In this case, it is called a set of subsets *over* S. A set $\mathbf{T}$ of subsets (over S) is a *field of sets* (over S) if $S \in \mathbf{T}$, $\emptyset \in \mathbf{T}$, and it is closed under union, intersection, and complement. In other words, for $S_1, S_2 \in \mathbf{T}$, $S_1 \cup S_2,\ S_1 \cap S_2,\ S \setminus S_1 \in \mathbf{T}$ as well.

If $\mathbf{T}$ is any set of subsets of S, it may be completed to a field of sets by adding a largest set and $\emptyset$, and then closing it up under union, intersection, and complement. More precisely, define $\mathsf{FClosure}\langle\mathbf{T}\rangle$ to be the smallest set of subsets of S with $\mathbf{T} \subseteq \mathsf{FClosure}\langle\mathbf{T}\rangle$, $S,\ \emptyset \in \mathsf{FClosure}\langle\mathbf{T}\rangle$, and for any $S_1, S_2 \in \mathsf{FClosure}\langle\mathbf{T}\rangle$, $S_1 \cup S_2, S_1 \cap S_2, S \setminus S_1 \in \mathsf{FClosure}\langle\mathbf{T}\rangle$ as well.

Given a set $\mathbf{T}$ of subsets over S, define $\mathsf{FoSLat}\langle\mathbf{T}\rangle = (\mathsf{FClosure}\langle\mathbf{T}\rangle, \cup, \cap, \complement_{\mathbf{T}}, \emptyset, \bigcup\mathbf{T})$, with $\complement_{\mathbf{T}}$ the operator defined by $X \mapsto (\bigcup\mathbf{T}) \setminus X$. It is immediate that $\mathsf{FoSLat}\langle\mathbf{T}\rangle$ is a Boolean algebra. What is remarkable is that the converse also holds.

2.5. *Representation of Boolean Algebras*

Every Boolean algebra $\mathbf{L}$ is isomorphic to $\mathsf{FoSLat}\langle \mathbf{T} \rangle$ for some set $\mathbf{T}$ of subsets. Furthermore, if $\mathbf{L}$ is finite, then $\mathbf{T}$ may be chosen so that $\mathbf{T} = 2^S$ for some finite set S.

Proof. Follows from Stone's representation theorem for Boolean algebras (Davey and Priestly, 2002, 11.4). □

3. The Formalism for Multigranular Attributes

One of the key features of the formalism for multigranular attributes presented in Hegner and Rodríguez (2016), which sets it apart from earlier work such as Rodríguez and Bravo (2012) and Bravo and Rodríguez (2014), is that it supports the representation of (disjoint) join constraints, such as that illustrated in (r-Chile) of Section 1, in addition to subsumption constraints. The approach taken in Hegner and Rodríguez (2016) is to work directly and from the start with rules, such as (r-Chile). A structure which satisfies such constraints, called an SBBP, is then introduced to model and support certain aspects of these rules. While correct, this approach is in reverse of more conventional approaches in mathematical logic, in which structures are introduced first and then the semantics of constraints (rules here) are defined via satisfaction of structures. It has the further disadvantage of making comparison to related approaches, such as the partition model, more difficult.

In this paper, a method of realizing what is basically the same semantics as presented in Hegner and Rodríguez (2016), but with the more conventional approach of defining structures first, and then the semantics of rules via satisfaction of such models, is presented. A comparison to the partition model (Spyratos, 1987; Cosmadakis *et al.*, 1986; Molnár, 2007) is also provided.

3.1. *Granularity Schemata*

In the classical relational model, the columns are labelled with *attributes*, with each attribute A assigned a set of *domain elements* from which the values for A are taken. In the granulated approach, each attribute carries the further structure of a partially ordered set of *granularities*. The domain elements, called *granules*, also have a natural order structure which is tied to that of the granularities. Since distinct attributes may nevertheless have the same granules and granularities, it is convenient to encapsulate this information. This is done in two steps. In Section 3.2, the necessary definition for the granularity structure is made, while in Section 3.3, the structures for granules are introduced.

Formally, a *granularity schema* is an ordered pair $\mathfrak{S} = (\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle)$ in which $\mathbf{Glty}\langle\mathfrak{S}\rangle = (\mathsf{Glty}\langle\mathfrak{S}\rangle, \leqslant_{\mathbf{Glty}\langle\mathfrak{S}\rangle}, \top_{\mathbf{Glty}\langle\mathfrak{S}\rangle})$ is a granularity poset (to be defined in Section 3.2) and $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle = (\mathbf{Gnle}\langle\mathfrak{S}\rangle, \Pi_{\mathsf{Gnle}}\langle\mathfrak{S}\rangle)$ is a granule assignment for $\mathfrak{S}$ relative to $\mathbf{Glty}\langle\mathfrak{S}\rangle$ (to be defined in Section 3.3).

Associated with each *granulated attribute* A is a granularity schema name, denoted $\mathfrak{G}_A$.

3.2. *Granularity Posets*

A *granularity poset* is an upper-bounded poset $\mathbf{P} = (P, \leqslant_P, \top_P)$. The elements in P are called the *granularity identifiers* or, less formally, just the *granularities*. As noted in Section 3.1 above, with each granularity schema name $\mathfrak{S}$ is associated a granularity poset, denoted $\mathbf{Glty}\langle\mathfrak{S}\rangle = (\mathsf{Glty}\langle\mathfrak{S}\rangle, \leqslant_{\mathbf{Glty}\langle\mathfrak{S}\rangle}, \top_{\mathbf{Glty}\langle\mathfrak{S}\rangle})$.

The relation scheme $R_{\mathsf{sumb}}\langle A_{\mathsf{Plc}}, A_{\mathsf{Tim}}, B_{\mathsf{Bth}}\rangle$ of Section 1 provides a context for examples. First of all, each of the three attributes has a coarsest granularity, which recaptures no information about the domain value: $\top_{\mathbf{Glty}\langle\mathfrak{S}_{A_{\mathsf{Plc}}}\rangle}$ corresponds to all of Chile, $\top_{\mathbf{Glty}\langle\mathfrak{S}_{A_{\mathsf{Tim}}}\rangle}$ lumps all time values into one, and $\top_{\mathbf{Glty}\langle\mathfrak{S}_{B_{\mathsf{Bth}}}\rangle}$ lumps all numbers into one. The spatial attribute schema $\mathfrak{S}_{A_{\mathsf{Plc}}}$ might have, in addition to $\top_{\mathbf{Glty}\langle\mathfrak{S}_{A_{\mathsf{Plc}}}\rangle}$, Region, City, and $\mathsf{NatRegion}$ (identifying natural, as opposed to political, regions) as granularities, with $\mathsf{City} \leqslant \mathsf{Region} \leqslant \top_{\mathbf{Glty}\langle\mathfrak{S}_{A_{\mathsf{Plc}}}\rangle}$ and $\mathsf{NatRegion} \leqslant \top_{\mathbf{Glty}\langle\mathfrak{S}_{A_{\mathsf{Plc}}}\rangle}$. It has no least granularity; no granularity which is finer than both cities and natural regions is modelled.

The temporal attribute schema $\mathfrak{S}_{A_{\mathsf{Tim}}}$ might have, in addition to $\top_{\mathbf{Glty}\langle\mathfrak{S}_{A_{\mathsf{Tim}}}\rangle}$, $\mathsf{QuarterYr}$, $\mathsf{MonthYr}$, and WeekYr as granularities, with $\mathsf{MonthYr} \leqslant \mathsf{QuarterYr}$ and $\mathsf{WeekYr} \leqslant \top_{\mathbf{Glty}\langle\mathfrak{S}_{A_{\mathsf{Tim}}}\rangle}$. Here $\mathsf{QuarterYr}$ represents a quarter of a given year; similarly for $\mathsf{MonthYr}$ and WeekYr. $\top_{\mathbf{Glty}\langle\mathfrak{S}_{A_{\mathsf{Tim}}}\rangle}$ lumps together all of time. Note that neither $\mathsf{WeekYr} \leqslant \mathsf{MonthYr}$ nor $\mathsf{WeekYr} \leqslant \mathsf{QuarterYr}$ holds, since a single week may span two months or two quarters. It has no least granularity since the overlap of a week and a month need not correspond to any granularity.

Finally, for the attribute schema $\mathfrak{S}_{B_{\mathsf{Bth}}}$, fix $\mathsf{maxr} \in \mathbb{N}^+$. For $i \in [0, \mathsf{maxr}]$, the granularity round_i identifies rounding to the nearest 10^i. In particular, round_0 identifies no rounding at all, and is thus the least element of $\mathbf{Glty}\langle\mathfrak{S}_{B_{\mathsf{Bth}}}\rangle$; i.e. $\mathsf{round}_0 = \bot_{\mathbf{Glty}\langle\mathfrak{S}_{B_{\mathsf{Bth}}}\rangle}$. Thus $\bot_{\mathbf{Glty}\langle\mathfrak{S}_{B_{\mathsf{Bth}}}\rangle} = \mathsf{round}_0 \leqslant \mathsf{round}_i \leqslant \mathsf{round}_j \leqslant \top_{\mathbf{Glty}\langle\mathfrak{S}_{B_{\mathsf{Bth}}}\rangle}$ for $j < i$.

To elaborate these examples, it is necessary to have a representation for granules as well. This issue is substantially more complex, and is examined next.

3.3. *Granule Assignment*

A granule assignment for a granularity schema $\mathfrak{S}$ extends the idea of a domain assignment for an ordinary relational attribute. It provides a basic order on the granules, and, in addition, it assigns a set of granules to each granularity in a way which respects this preorder. The preorder structure induces equivalence relation on the granules, so that two which will have the same underlying semantics (as defined by a granule structure — see Section 3.5), may nevertheless have different names.

As noted in Section 3.1 above, with each granularity schema $\mathfrak{S}$ is associated a granule assignment $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$, relative to its granularity poset $\mathbf{Glty}\langle\mathfrak{S}\rangle$. Formally, this *granule assignment* is an ordered pair $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle = (\mathbf{Gnle}\langle\mathfrak{S}\rangle, \Pi_{\mathsf{Gnle}}\langle\mathfrak{S}\rangle)$ in which the following three conditions hold.

(grasgn-i) $\mathbf{Gnle}\langle\mathfrak{S}\rangle = (\mathsf{Granules}\langle\mathfrak{S}\rangle, \overline{\sqsubseteq}_{\mathfrak{S}}, \top_{\mathfrak{S}}, \bot_{\mathfrak{S}})$ is a bounded preorder, called the *granule preorder*. $\mathsf{Granules}\langle\mathfrak{S}\rangle$ is called the set of *granule identifiers* or just *granules*. $\top_{\mathfrak{S}}$ is called the *universal granule* and $\bot_{\mathfrak{S}}$ is called the *empty granule*. This preorder is required to have the further property that $[\bot_{\mathfrak{S}}]_{\mathsf{Gnle}\langle\mathfrak{S}\rangle} = \{\bot_{\mathfrak{S}}\}$; i.e. the bottom granule $\bot_{\mathfrak{S}}$ is only equivalent to itself.

Because it is used frequently, the set $\mathsf{Granules}\langle\mathfrak{S}\rangle \setminus \{\bot_{\mathfrak{S}}\}$ has the special notation $\mathsf{Granules}_{\not\bot}\langle\mathfrak{S}\rangle$. Also, since no confusion can result, for any $g \in \mathsf{Granules}\langle\mathfrak{S}\rangle$, $[g]_{\mathsf{Gnle}\langle\mathfrak{S}\rangle}$ will be shortened to just $[g]_{\mathfrak{S}}$.

(grasgn-ii) $\Pi_{\mathsf{Gnle}}\langle\mathfrak{S}\rangle = \{\mathsf{Granules}\langle\mathfrak{S}|G\rangle \mid G \in \mathsf{Glty}\langle\mathfrak{S}\rangle\}$ is a partition of $\mathsf{Granules}_{\not\bot}\langle\mathfrak{S}\rangle$ with the following two properties.

(grtognle-i) $\mathsf{Granules}\langle\mathfrak{S}|\top_{\mathsf{Glty}\langle\mathfrak{S}\rangle}\rangle = [\top_{\mathfrak{S}}]_{\mathfrak{S}}$.

(grtognle-ii) $(\forall G \in \mathsf{Glty}\langle\mathfrak{S}\rangle \setminus \{\top_{\mathsf{Glty}\langle\mathfrak{S}\rangle}\})(\forall g_1, g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G\rangle)$
$$((g_1 \neq g_2) \Rightarrow ([g_1]_{\mathfrak{S}} \neq [g_1]_{\mathfrak{S}})).$$

(grtognle-iii) $(\forall G \in \mathsf{Glty}\langle\mathfrak{S}\rangle \setminus \{\top_{\mathsf{Glty}\langle\mathfrak{S}\rangle}\})(\forall g_1, g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G\rangle)$
$$((g_1 \neq g_2) \Rightarrow (\mathsf{GLB}_{\mathsf{Gnle}\langle\mathfrak{S}\rangle}\langle\{g_1, g_2\}\rangle = \bot_{\mathfrak{S}}).$$

$\mathsf{Granules}\langle\mathfrak{S}|G\rangle$ is called the set of *granules of granularity* G.

(grasgn-iii) $(\forall G_1, G_2 \in \mathsf{Glty}\langle\mathfrak{S}\rangle)((G_1 \leqslant G_2) \Leftrightarrow$
$$(\forall g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle)(\exists g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle)(g_1 \mathrel{\bar{\sqsubseteq}_{\mathfrak{S}}} g_2)).$$

Equivalent granules in this preorder will map to the same underlying set, as defined in Section 3.5, and so they are *aliases* of one another, of sorts.

(grtognle-i) of (grasgn-iii) stipulates that only $\top_{\mathfrak{S}}$, together with any other granule which is equivalent to it, are associated with the universal granule $\top_{\mathsf{Glty}\langle\mathfrak{S}\rangle}$. This is the only case in which two distinct, equivalent names may be associated with the same granularity.

(grtognle-ii) mandates that, with the exception of $\top_{\mathsf{Glty}\langle\mathfrak{S}\rangle}$, equivalent granules may not be associated with the same granularity.

(grtotnle-iii) ensures that any two distinct granules of the same granularity may be defined so as not to overlap. However, by itself, it does not guarantee that they do not overlap. Overlap is defined first with the concept of a structure in Section 3.5.

(grasgn-ii) implies in particular that each granule belongs to only one granularity. It is convenient to have a function which identifies this association. To this end, define $\mathsf{Glty}_{\mathfrak{S}} : \mathsf{Granules}_{\not\bot}\langle\mathfrak{S}\rangle \to \mathsf{Glty}\langle\mathfrak{S}\rangle$ to be the function which sends $g \in \mathsf{Granules}_{\not\bot}\langle\mathfrak{S}\rangle$ to the unique $G \in \mathsf{Glty}\langle\mathfrak{S}\rangle$ for which $g \in \mathsf{Granules}\langle\mathfrak{S}|G\rangle$.

Continuing with the example context of 3.2, $\{\mathit{Región\_R} \mid \mathit{I} \leqslant R \leqslant \mathit{XV}\} \subseteq \mathsf{Granules}\langle\mathfrak{G}_{A_{\mathsf{Plc}}}|\mathsf{Region}\rangle$ and $\{\mathit{Concepción}, \mathit{Santiago}, \mathit{Los\_Ángeles}\} \subseteq \mathsf{Granules}\langle\mathfrak{G}_{A_{\mathsf{Plc}}}|\mathsf{City}\rangle$. Furthermore, *Concepción* $\bar{\sqsubseteq}_{\mathfrak{G}_{A_{\mathsf{Plc}}}}$ *Región VIII*, *Los_Ángeles* $\bar{\sqsubseteq}_{\mathfrak{G}_{A_{\mathsf{Plc}}}}$ *Región VIII*, and *Santiago* $\bar{\sqsubseteq}_{\mathfrak{G}_{A_{\mathsf{Plc}}}}$ *Región XIII*. If $\{\mathit{Concepción}, \mathit{Santiago}, \mathit{Los\_Ángeles}\}$ are the only granules of granularity City, then condition (grasgn-iii) is substantiated for $\mathsf{City} \leqslant_{\mathsf{Glty}\langle\mathsf{A}_{\mathsf{Plc}}\rangle} \mathsf{Region}$ by the fact that each of these three cities lies in one of the fifteen regions.

If Chile is the entire modelling space, then $\{\top_{\mathfrak{G}_{A_{\mathsf{Plc}}}}, \mathit{Chile}\} = \mathsf{Granules}\langle\mathfrak{G}_{A_{\mathsf{Plc}}}|\top_{\mathsf{Glty}\langle\mathfrak{G}_{\mathsf{A}_{\mathsf{Plc}}}\rangle}\rangle$, with, of course, $[\top_{\mathfrak{G}_{A_{\mathsf{Plc}}}}]_{\mathfrak{G}_{A_{\mathsf{Plc}}}} = [\mathit{Chile}]_{\mathfrak{G}_{A_{\mathsf{Plc}}}}$. For another example of equivalence, suppose that there is an additional granularity County, and that every city lies within a county; i.e. $\mathsf{City} \leqslant_{\mathsf{Glty}\langle\mathsf{A}_{\mathsf{Plc}}\rangle} \mathsf{County}$. In some cases, the geographic region of the city and the county may coincide; this is the case with the city of Concepción and the county of Concepción in Chile. Thus, $[\mathit{Concepción}]_{\mathfrak{G}_{A_{\mathsf{Plc}}}} = [\mathit{Concepción\_county}]_{\mathfrak{G}_{A_{\mathsf{Plc}}}}$ with *Concepción_condado* denoting the county and *Concepción* the city. Of course, these are not identical entities, because they may have, for example, different administrative offices, but for the scope of the modelling developed here, they have identical properties.

The set of granules need not be finite, even in practical examples. See Section 4.2 for a detailed description of how the granules of an attribute schema such as $\mathfrak{G}_{B_{\mathsf{Bth}}}$ would be modelled using an infinite set.

While a granule assignment assigns granules to granularities, and provides basic order structure on the granules, it does not convey any other lattice-like information about the granules, such as that embodied in (r-Chile). This task is addressed via an additional construction, the *granule structure*, which is defined next.

3.4. *Notational Convention*

Throughout the rest of this section, unless stated specifically to the contrary, take $\mathfrak{S} = (\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle)$ to be a granulated attribute schema with $\mathbf{Glty}\langle\mathfrak{S}\rangle = (\mathsf{Glty}\langle\mathfrak{S}\rangle, \leqslant_{\mathbf{Glty}\langle\mathfrak{S}\rangle}, \top_{\mathbf{Glty}\langle\mathfrak{S}\rangle})$ and $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle = (\mathbf{Gnle}\langle\mathfrak{S}\rangle, \Pi_{\mathsf{Gnle}}\langle\mathfrak{S}\rangle)$.

3.5. *Granule Structure*

A granule structure starts with a universe, and then assigns a subset of that universe to each granule. It is thus similar to the approach of Bravo and Rodríguez (2014). However, because it is desired to recapture other constraints as well, particularly join rules such as (r-Chile), it needs to be constructed with more in mind.

A *granule structure* for $\mathfrak{S}$ is a pair $\sigma = (\mathsf{Dom}\langle\sigma\rangle, \mathsf{GnletoDom}_\sigma)$ in which $\mathsf{Dom}\langle\sigma\rangle$ is a (not necessarily finite) set, called the *domain* of σ, and $\mathsf{GnletoDom}_\sigma : \mathsf{Granules}\langle\mathfrak{S}\rangle \to \mathbf{2}^{\mathsf{Dom}\langle\sigma\rangle}$ is a function, subject to the following conditions.

(grstr-i) $(\forall g_1, g_2 \in \mathsf{Granules}\langle\mathfrak{S}\rangle)((g_1 \mathrel{\bar{\sqsubseteq}}_{\mathfrak{S}} g_2) \Rightarrow (\mathsf{GnletoDom}_\sigma(g_1) \subseteq \mathsf{GnletoDom}_\sigma(g_2)))$.

(grstr-ii) $(\forall G \in \mathsf{Glty}\langle\mathfrak{S}\rangle \setminus \{\top_{\mathbf{Glty}\langle\mathfrak{S}\rangle}\})(\forall g_1, g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G\rangle)((g_1 \neq g_2) \Rightarrow (\mathsf{GnletoDom}_\sigma(g_1) \cap \mathsf{GnletoDom}_\sigma(g_2) = \emptyset))$.

(grstr-iii) $(\forall g_1, g_2 \in \mathsf{Granules}\langle\mathfrak{S}\rangle)((\mathsf{GnletoDom}_\sigma(g_1) = \mathsf{GnletoDom}_\sigma(g_2)) \Leftrightarrow [g_1]_{\mathfrak{S}} = [g_2]_{\mathfrak{S}})$.

The set of all granule structures for $\mathfrak{S}$ is denoted $\mathsf{GranStruct}\langle\mathfrak{S}\rangle$.

(grstr-i) mandates that, with respect to the order $\bar{\sqsubseteq}_{\mathfrak{S}}$, granule subsumption is modelled as set inclusion.

(grstr-ii) mandates that, except for the top granularity $\top_{\mathbf{Glty}\langle\mathfrak{S}\rangle}$ of $\mathsf{Glty}\langle\mathfrak{S}\rangle$, each pair of distinct granules of the same granularity G be nonoverlapping.

(grstr-iii) asserts that exactly those granules which are equivalent map to the same domain set.

3.6. *Motivation for Rule-Based Semantics*

It is possible to use a single structure to define the semantics of each granulated attribute schema. Indeed, for $\mathfrak{G}_{A_{\mathsf{Plc}}}$, a suitable domain might be $\sigma_{\mathsf{Earth}} = (\mathsf{Dom}\langle\sigma_{\mathsf{Earth}}\rangle, \mathsf{GnletoDom}_{\sigma_{\mathsf{Earth}}})$ in which $\mathsf{Dom}\langle\sigma_{\mathsf{Earth}}\rangle$ is the set of all coordinates on a sphere (representing the earth), with $\mathsf{GnletoDom}_{\sigma_{\mathsf{Earth}}}$ the function which maps a place (*qua* granule) to the set of coordinates on earth which it covers. Coordinates in the plane $\mathbb{R}^2$ might also be used if the entire model is of a smaller region, such as a country. From the point of view of embodying relevant information, it is a near-perfect model. The drawback is that it is much too detailed for most use, and is enormous to store, access, and maintain. For most applications, it is only necessary to know, for example, that the city *Concepción* lies in *Región_VII*; the exact physical coordinates are unnecessary.

A simpler, yet structural, model is also an alternative, provided only very simple constraints are used. For example, the approach of Bravo and Rodríguez (2014) employs a single structure, while representing constraints such as the subsumption embodied in the granule preorder of (grasgn-i) and the pairwise disjointness embodied in (grstr-ii). However, to model more complex spatial operations, such as join, requires a much more complete structure; complex enough to be burdened with many of the problems sketched for σ_{Earth} above. Far preferable is to be able to write constraints, such as (r-Chile), without detailed knowledge of the underlying geographic regions. It is sufficient to know that the fifteen regions cover Chile without overlap; additional information about their exact coordinates is not material to the model.

The solution forwarded here is to model the semantics as a set of rules, such as (r-Chile), and then to define the semantics as the set of structures which satisfy all of those rules. Thus, the semantics forwarded here is not defined by a single structure, but rather as a a set of possibilities satisfying some set of constraints. In that way, detailed knowledge of, for example, geographic regions, becomes unnecessary. This approach is next developed in detail.

3.7. *Granule Expressions and their Semantics*

Granule expressions operate on lattice-like expressions involving granules (in $\mathsf{Granules}\langle\mathfrak{S}\rangle$). Such an expression may be evaluated on a granule structure σ, returning a subset of domain elements (i.e. a subset of $\mathsf{Dom}\langle\sigma\rangle$). The evaluation is the natural one, with the lattice-like join operation $\bigsqcup_{\mathfrak{S}}$ on granules associated with set union $\bigcup$ on subsets of $\mathsf{Dom}\langle\sigma\rangle$, and the lattice-like meet operation $\bigsqcap_{\mathfrak{S}}$ on granules associated with set intersection $\bigcap$ on subsets of $\mathsf{Dom}\langle\sigma\rangle$.

Formally, the *granule expressions* over $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$ are defined as the smallest set $\mathsf{GrExpr}\langle\mathfrak{S}\rangle$ which is closed under the following operations.

(grex-i) $\mathsf{Granules}\langle\mathfrak{S}\rangle \subseteq \mathsf{GrExpr}\langle\mathfrak{S}\rangle$.

(grex-ii) If $S \subseteq_f \mathsf{GrExpr}\langle\mathfrak{S}\rangle$, then $(\bigsqcup_{\mathfrak{S}} S) \in \mathsf{GrExpr}\langle\mathfrak{S}\rangle$.

(grex-iii) If $S \subseteq_f \mathsf{GrExpr}\langle\mathfrak{S}\rangle$, then $(\bigsqcap_{\mathfrak{S}} S) \in \mathsf{GrExpr}\langle\mathfrak{S}\rangle$.

(grex-iv) If $e_1, e_2 \in \mathsf{GrExpr}\langle\mathfrak{S}\rangle$, then $\mathsf{RelCompl}_{\mathfrak{S}}\langle e_1, e_2\rangle \in \mathsf{GrExpr}\langle\mathfrak{S}\rangle$.

$\mathsf{RelCompl}$ denotes relative complement; the semantics of $\mathsf{RelCompl}_{\mathfrak{S}}\langle e_1, e_2\rangle$ is the set of all domain elements in $\mathsf{GrExSem}_{\sigma}(e_1)$ but not in $\mathsf{GrExSem}_{\sigma}(e_2)$.

Relative to a granule structure $\sigma = (\mathsf{Dom}\langle\sigma\rangle, \mathsf{GnletoDom}_\sigma)$ for $\mathfrak{S}$, the *semantics* of the members of $\mathsf{GrExpr}\langle\mathfrak{S}\rangle$ are defined by the function $\mathsf{GrExSem}_\sigma : \mathsf{GrExpr}\langle\mathfrak{S}\rangle \to \mathbf{2}^{\mathsf{Dom}\langle\sigma\rangle}$ which is given on elements as follows.

(semgrex-i) For $g \in \mathsf{Granules}\langle\mathfrak{S}\rangle$, $\mathsf{GrExSem}_\sigma(g) = \mathsf{GnletoDom}_\sigma(g)$.

(semgrex-ii) For $S \subseteq_f \mathsf{GrExpr}\langle\mathfrak{S}\rangle$,

$$\mathsf{GrExSem}_\sigma((\bigsqcup_{\mathfrak{S}} S)) = \bigcup\{\mathsf{GnletoDom}_\sigma(s) \mid s \in S\}.$$

(semgrex-iii) For $S \subseteq_f \mathsf{GrExpr}\langle\mathfrak{S}\rangle$,

$$\mathsf{GrExSem}_\sigma((\bigsqcap_{\mathfrak{S}} S)) = \bigcap\{\mathsf{GnletoDom}_\sigma(s) \mid s \in S\}.$$

(semgrex-iv) For $e_1, e_2 \in \mathsf{GrExpr}\langle\mathfrak{S}\rangle$,

$$\mathsf{GrExSem}_\sigma(\mathsf{RelCompl}_{\mathfrak{S}}\langle e_1, e_2\rangle) = \mathsf{GrExSem}_\sigma(e_1) \setminus \mathsf{GrExSem}_\sigma(e_2).$$

In short, the semantics of a granule g is just the set of domain elements which form its image under the mapping $\mathsf{GnletoDom}_\sigma$, while the semantics of join and meet expressions are converted via set union and intersection, respectively.

3.8. *Granule Rules as Constraints*

Constraints which determine possible structures for a multigranular schema are specified by sentences called *(granule) rules*. These constraints are built from expressions in $\mathsf{GrExpr}\langle\mathfrak{S}\rangle$. Formally, a *simple granule rule* is of the form $e_1 \sqsubseteq e_2$, where $e_1, e_2 \in \mathsf{GrExpr}\langle\mathfrak{S}\rangle$. The set of all such rules is denoted $\mathsf{SimpGrRules}\langle\mathfrak{S}\rangle$.

A *granule rule* is any nonempty conjunction of simple granule rules. In other words, a granule rule is of the form $\varphi_1 \wedge \varphi_2 \wedge \ldots \wedge \varphi_k$ with $\{\varphi_i \mid i \in [1,k]\} \subseteq \mathsf{SimpGrRules}\langle\mathfrak{S}\rangle$. This may also be written $\bigwedge\{\varphi_i \mid i \in [1,k]\}$, with $\bigwedge$ denoting logical conjunction. The set of all granule rules for $\mathfrak{S}$ is denoted $\mathsf{GrRules}\langle\mathfrak{S}\rangle$.

For $e_1, e_2 \in \mathsf{GrExpr}\langle\mathfrak{S}\rangle$, $e_1 = e_2$ is an abbreviation for $(e_1 \sqsubseteq_{\mathfrak{S}} e_2) \wedge (e_2 \sqsubseteq_{\mathfrak{S}} e_1)$.

Without loss of generality, it will furthermore be assumed that distinct, equivalent granules are never used together in the same rule.

The semantics for rules follow, at least in spirit, those of traditional mathematical logic (Monk, 1976). Formally, a granule structure $\sigma \in \mathsf{GranStruct}\langle\mathfrak{S}\rangle$ is a *model* of the rule $\varphi \in \mathsf{GrRules}\langle\mathfrak{S}\rangle$, written $\varphi \models_{\mathfrak{S}} \sigma$, if the appropriate condition below is met.

(rulemod-i) If φ is of the form $e_1 \sqsubseteq e_2$, then σ is a model of φ precisely in the case that $\mathsf{GrExSem}_\sigma(e_1) \subseteq \mathsf{GrExSem}_\sigma(e_2)$ holds.

(rulemod-ii) If φ is a conjunction of the form $\bigwedge\{\varphi_i \mid i \in [1,k]\}$, then σ is a model of φ precisely in the case that it is a model of each φ_i for $i \in [1,k]$.

The set of all models of the rule φ is denoted $\mathsf{Models}_{\mathfrak{S}}\langle\varphi\rangle$. If Φ is a set of rules, $\mathsf{Models}_{\mathfrak{S}}\langle\Phi\rangle$ denotes $\bigcap\{\mathsf{Models}_{\mathfrak{S}}\langle\varphi\rangle \mid \varphi \in \Phi\}$. If σ is a model of φ (resp. Φ), then $\varphi \models_{\mathfrak{S}} \sigma$ (resp. $\Phi \models_{\mathfrak{S}} \sigma$) may also be written.

φ (resp. Φ) is $\mathfrak{S}$-*satisfiable* if it admits a model; i.e. if $\mathsf{Models}_{\mathfrak{S}}\langle\varphi\rangle$ (resp. $\mathsf{Models}_{\mathfrak{S}}\langle\Phi\rangle$) is nonempty.

The $\mathfrak{S}$-*closure* (or just *closure* if the context is clear) of $\Phi \subseteq \mathsf{GrRules}\langle\mathfrak{S}\rangle$, denoted Φ^+, is $\{\varphi \in \mathsf{GrRules}\langle\mathfrak{S}\rangle \mid \mathsf{Models}_{\mathfrak{S}}\langle\Phi\rangle \subseteq \mathsf{Models}_{\mathfrak{S}}\langle\varphi\rangle\}$.

3.9. *Built-in Rules*

There are certain rules which are enforced by the very definition of a structure, and so must hold in every model of a consistent set $\Phi \subseteq \mathsf{GrRules}\langle\mathfrak{S}\rangle$. Define the *built-in rules* of $\mathfrak{S}$, denoted $\mathsf{BuiltInRules}\langle\mathfrak{S}\rangle$, by the following.

(bir-i) For $g_1, g_2 \in \mathsf{Granules}\langle\mathfrak{S}\rangle$, if $g_1 \mathrel{\bar{\sqsubseteq}_{\mathfrak{S}}} g_2$, then $(g_1 \sqsubseteq_{\mathfrak{S}} g_2) \in \mathsf{BuiltInRules}\langle\mathfrak{S}\rangle$.

(bir-ii) For $g_1, g_2 \in \mathsf{Granules}\langle\mathfrak{S}\rangle$, if there is a $G \in \mathsf{Glty}\langle\mathfrak{S}\rangle$ and $g_1', g_2' \in \mathsf{Granules}\langle\mathfrak{S}|G\rangle$ with $g_1 \mathrel{\bar{\sqsubseteq}_{\mathfrak{S}}} g_1'$ and $g_2 \mathrel{\bar{\sqsubseteq}_{\mathfrak{S}}} g_2'$, then $(\bigsqcap_{\mathfrak{S}}\{g_1, g_2\} = \bot_{\mathfrak{S}}) \in \mathsf{BuiltInRules}\langle\mathfrak{S}\rangle$.

That these rules are satisfied by every model follows immediately from (grstr-i) and (grstr-ii) of Section 3.5.

Additional elementary-subsumption and basic-disjointness rules may of course be added to the constraints of $\mathfrak{S}$. However, additional such rules may also be excluded via a closed-world assumption; see Section 3.20 below.

The built-in rules are instances of basic rules, as defined below.

3.10. *Basic Rules*

Although it is theoretically appealing to work with a very general class of rules, such as $\mathsf{GrRules}\langle\mathfrak{S}\rangle$, this generality may prove to be unrealistic to implement fully. It is therefore useful to identify a smaller set of constraints which are more manageable while still possessing enough expressive power. To this end, the basic rules are introduced.

Define the *primitive basic rules* over $\mathfrak{S}$ as those fitting one of the following two types.

(pbrule-i) A *basic subsumption rule* is of the form $g \sqsubseteq_{\mathfrak{S}} \bigsqcup_{\mathfrak{S}} S$, for $\{g\} \cup S \subseteq \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$.

(pbrule-ii) A *basic disjointness rule* is of the form $\bigsqcap_{\mathfrak{S}}\{g_1, g_2\} = \bot_{\mathfrak{S}}$ for $g_1, g_2 \in \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$ and $[g_1]_{\mathfrak{S}} \neq [g_2]_{\mathfrak{S}}$.

There are three further kinds of rules which, while definable in terms of the primitive basic rules, are so fundamental in usage that they deserve their own names and representations.

(xbrule-i) An *elemental subsumption rule* is of the form $g_1 \sqsubseteq_{\mathfrak{S}} g_2$ with $g_1, g_2 \in \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$. Its definition in terms of primitive rules is $g_1 \sqsubseteq_{\mathfrak{S}} \bigsqcup_{\mathfrak{S}}\{g_2\}$, so it is a special case of a basic subsumption rule. A basic subsumption rule which is not elemental is called *a complex subsumption rule*.

(xbrule-ii) A *basic join rule* is of the form $g = \bigsqcup_{\mathfrak{S}} S$, for $\{g\} \cup S \subseteq \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$. Its definition in terms of primitive rules and elemental subsumption rules is

$$(g \sqsubseteq_{\mathfrak{S}} \textstyle\bigsqcup_{\mathfrak{S}} S) \wedge (\bigwedge\{g_i \sqsubseteq_{\mathfrak{S}} g \mid g_i \in S\}).$$

It is easy to verify that both of these definitions respect the semantics defined in Sections 3.7 and 3.8.

In order to define the third type of extended basic rule, a further definition is useful. Given $S \subseteq \mathsf{Granules}\langle A\rangle$, define

$$\mathsf{PWDisjnt}_{\mathfrak{S}}\langle S\rangle = \textstyle\bigwedge\{\bigsqcap_{\mathfrak{S}}\{g_1, g_2\} = \bot_{\mathfrak{S}} \mid (g_1, g_2 \in S) \text{ and } (g_1, g_2 \notin \mathsf{GnleEq}_{\mathfrak{S}})\}$$

to be the granule rule which asserts that every pair $g_1, g_2 \subseteq S$ of elements in S which are not granule equivalent are disjoint, in the sense that their meet is $\bot_{\mathfrak{S}}$. (In the above formula, "$\bigwedge$" denotes logical conjunction on a set.) If $\mathsf{PWDisjnt}_{\mathfrak{S}}\langle S\rangle$ holds, then it is said that S is *pairwise disjoint (in $\mathfrak{S}$)*. Now, the last type of extended basic rule may be defined.

(xbrule-iii) A *basic disjoint-join rule* is written as $g = \bigsqcup\!\!\!\!\perp_{\mathfrak{S}} S$, for $\{g\} \cup S \subseteq \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$. Its definition in terms of basic join rules and $\mathsf{PWDisjnt}$ is $(g = \bigsqcup_{\mathfrak{S}} S) \wedge \mathsf{PWDisjnt}_{\mathfrak{S}}\langle S\rangle$.

The collection of all rules defined by (pbrule-i)–(pbrule-ii) and (xbrule-i)–(xbrule-iii) is called the *basic rules* over $\mathfrak{S}$, and is denoted $\mathsf{BaRules}\langle\mathfrak{S}\rangle$.

These basic rules have the semantics provided in Section 3.8.

The obvious question of when a set of rules admits a model must be addressed. The answer is not trivial, and depends upon whether there is a Boolean algebra in which the rules hold.

3.11. $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$-*Algebras*

In order to characterize the satisfiability of a set Φ of rules (as defined in Section 3.8, it is convenient to work with Boolean algebras whose elements include equivalence classes of granules. The rules are then tested by evaluating them within that algebra.

Using the notation from Section 2.1 concerning equivalence relations, for $\equiv_{\mathbf{Gnle}\langle\mathfrak{S}\rangle}$ and $g \in \mathsf{Granules}\langle\mathfrak{S}\rangle$, $[g]_{\mathfrak{S}}$ denotes the equivalence class of g, and $\mathsf{Blocks}\langle\equiv_{\mathbf{Gnle}\langle\mathfrak{S}\rangle}\rangle$ denotes the set of blocks of $\mathbf{Gnle}\langle\mathfrak{S}\rangle$.

Define a $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$-algebra to be a Boolean algebra $\mathbf{L} = (L, \vee_L, \wedge_L, \complement_L, \perp_L, \top_L)$ with the following properties:

(gralat-i) $\mathsf{Blocks}\langle\equiv_{\mathbf{Gnle}\langle\mathfrak{S}\rangle}\rangle \subseteq L$.
(gralat-ii) $\perp_L = [\perp_{\mathfrak{S}}]_{\mathfrak{S}}$.
(gralat-iii) $\top_L = [\top_{\mathfrak{S}}]_{\mathfrak{S}}$.

Thus, some of the elements of a $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$-algebra $\mathbf{L}$ are (equivalence classes of) granules, but not all of them. In effect, the granules are embedded in $\mathbf{L}$.

3.12. *Evaluating Expressions and Rules in* $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$-*Algebras*

Let $\mathbf{L} = (L, \vee_L, \wedge_L, \complement_L, \perp_L, \top_L)$ be a $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$-algebra. For $\epsilon \in \mathsf{GrExpr}\langle\mathfrak{S}\rangle$, the *evaluation of ϵ in L*, denoted $\mathsf{Eval}\langle\epsilon:\mathbf{L}\rangle$, is defined as follows.

(evalexpr-i) If $\epsilon \in \mathsf{Granules}\langle\mathfrak{S}\rangle$, then $\mathsf{Eval}\langle\epsilon:\mathbf{L}\rangle = [\epsilon]_{\mathfrak{S}}$.
(evalexpr-ii) If ϵ is of the form $(\bigsqcup_{\mathfrak{S}} S)$, then $\mathsf{Eval}\langle\epsilon:\mathbf{L}\rangle = \bigvee_{\mathbf{L}}\{\mathsf{Eval}\langle s:\mathbf{L}\rangle \mid s \in S\}$.
(evalexpr-iii) If ϵ is of the form $(\bigsqcap_{\mathfrak{S}} S)$, then $\mathsf{Eval}\langle\epsilon:\mathbf{L}\rangle = \bigwedge_{\mathbf{L}}\{\mathsf{Eval}\langle s:\mathbf{L}\rangle \mid s \in S\}$.
(evalexpr-iv) If ϵ is of the form $\mathsf{RelCompl}\langle e_1, e_2\rangle$, then $\mathsf{Eval}\langle\epsilon:\mathbf{L}\rangle = \mathsf{Eval}\langle e_1:\mathbf{L}\rangle \setminus \mathsf{Eval}\langle e_2:\mathbf{L}\rangle$.

The *Boolean-algebra models* of $\varphi \in \mathsf{GrRules}\langle\mathfrak{S}\rangle$, written $\mathsf{BAlgModels}_{\mathfrak{S}}\langle\varphi\rangle$, are those $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$-algebras which satisfy the applicable condition below.

(evalrule-i) If φ is of the form $e_1 = e_2$, then $\mathbf{L} \in \mathsf{BAlgModels}_{\mathfrak{S}}\langle\varphi\rangle$ if $\mathsf{Eval}\langle e_1:\mathbf{L}\rangle = \mathsf{Eval}\langle e_2:\mathbf{L}\rangle$.

(evalrule-ii) If φ is of the form $e_1 \sqsubseteq e_2$, then $\mathbf{L} \in \mathsf{BAlgModels}_{\mathfrak{S}}\langle\varphi\rangle$ if

$$\mathsf{Eval}\langle e_1 : \mathbf{L}\rangle \leqslant_{\mathbf{L}} \mathsf{Eval}\langle e_2 : \mathbf{L}\rangle.$$

For $\Phi \subseteq \mathsf{GrRules}\langle\mathfrak{S}\rangle$, $\mathsf{BAlgModels}_{\mathfrak{S}}\langle\Phi\rangle$ is defined to be

$$\bigcap\{\mathsf{BAlgModels}_{\mathfrak{S}}\langle\varphi\rangle \mid \varphi \in \Phi\}.$$

3.13. *Existence of a Model*

For $\Phi \subseteq \mathsf{GrRules}\langle\mathfrak{S}\rangle$, $\mathsf{Models}_{\mathfrak{S}}\langle\Phi\rangle$ is nonempty iff $\mathsf{BAlgModels}_{\mathfrak{S}}\langle\Phi\rangle$ has that property. In other words, Φ is $\mathfrak{S}$-satisfiable iff there is a $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$-algebra in which every rule in Φ holds.

Proof. Assume that Φ is $\mathfrak{S}$-satisfiable. Let $\sigma = (\mathsf{Dom}\langle\sigma\rangle, \mathsf{GnletoDom}_\sigma)$ be a granule structure for $\mathsf{GrAsgn}\langle\mathfrak{S}\rangle$, and let $\mathbf{T} = \{\mathsf{GnletoDom}_\sigma(g) \mid g \in \mathsf{Granules}\langle\mathfrak{S}\rangle\}$. As sketched in Section 2.4, $\mathsf{FoSLat}\langle\mathbf{T}\rangle = (\mathsf{FClosure}\langle\mathbf{T}\rangle, \cup, \cap, \complement_{\mathbf{T}}, \emptyset, \bigcup\mathbf{T})$ is a Boolean algebra. Just by construction, $\mathsf{FoSLat}\langle\mathbf{T}\rangle \in \mathsf{BAlgModels}_{\mathfrak{S}}\langle\Phi\rangle$.

Conversely, let $\mathbf{L} = (L, \vee_{\mathbf{L}}, \wedge_{\mathbf{L}}, \complement_{\mathbf{L}}, \perp_{\mathbf{L}}, \top_{\mathbf{L}}) \in \mathsf{BAlgModels}_{\mathfrak{S}}\langle\Phi\rangle$. In view of Section 2.5, there is a set $\mathbf{T}'$ of subsets with the property that $\mathsf{FoSLat}\langle\mathbf{T}'\rangle = (\mathsf{FClosure}_\emptyset\langle\mathbf{T}'\rangle, \cup, \cap, \complement_{\mathbf{T}'}, \emptyset, \bigcup\mathbf{T}')$ is isomorphic to L. Let $\iota : L \to \mathsf{FClosure}_\emptyset\langle\mathbf{T}'\rangle$ be the function which underlies this isomorphism. To complete the proof and obtain a granule structure $\sigma' = (\mathsf{Dom}\langle\sigma'\rangle, \mathsf{GnletoDom}_{\sigma'}) \in \mathsf{Models}_{\mathfrak{S}}\langle\Phi\rangle$, it suffices to choose $\mathsf{Dom}\langle\sigma'\rangle = \bigcup\mathbf{T}'$ and define $\mathsf{GnletoDom}_{\sigma'}$ on elements by $g \mapsto \iota([g]_{\mathfrak{S}})$. □

3.14. *Guaranteeing Consistency of Sets of Rules*

The conditions identified in Section 3.13 may seem difficult to verify, thus limiting the practicality of the approach. For abstract specifications of constraint sets this is indeed the case; the question of whether a set of constraints is satisfiable is NP-hard; see Hegner (1994) for details. For more restrictive classes of constraints, such as the basic rules introduced in Section 3.10, more efficient algorithms for testing consistency are being developed and will appear in a forthcoming paper.

However, in many practical modelling situations, satisfiability is guaranteed, because a structure which satisfies the constraints underlies the granularity schema itself. For the granularity schema $\mathfrak{G}_{A_{\mathsf{Plc}}}$ of Sections 3.2 and 3.3, the physical structure $\sigma_{\mathsf{Earth}} = (\mathsf{Dom}\langle\sigma_{\mathsf{Earth}}\rangle, \mathsf{GnletoDom}_{\sigma_{\mathsf{Earth}}})$, as defined in Section 3.6 exists, even though it is not part of the full model. The fact that it exists, and that the rules are based upon it, is enough to guarantee that the rules are satisfiable.

3.15. *Armstrong Models*

It is useful to be able to make closed-world assumptions (CWA) on rules; that is to take as false all which cannot be proven true (see Section 3.20 below). It has long been recognized that such an assumption can easily lead to contradictions (Reiter, 1978; Example 8); to show that this cannot happen with rules as defined here, the notion of an Armstrong model is central.

For a class $\mathcal{C}$ of constraints, an Armstrong model for a set $\Phi \subseteq \mathcal{C}$ is one which satisfies all sentences in Φ^+, but no others in $\mathcal{C}$. Armstrong models were first studied for functional dependencies in Armstrong (1974) and later for other types of dependencies; for an overview, see Fagin (1982a). However, the notion may be formulated in a very general context, independent of any particular concept of database dependency (Fagin, 1982b).

For $\mathcal{C} = \mathsf{GrRules}\langle\mathfrak{S}\rangle$, the setting of interest here, $M \in \mathsf{GranStruct}\langle\mathfrak{S}\rangle$ is an *Armstrong model (with respect to* $\mathsf{GrRules}\langle\mathfrak{S}\rangle$*)* for a set $\Phi \subseteq \mathsf{GrRules}\langle\mathfrak{S}\rangle$ if for any $\Psi \in \mathsf{GrRules}\langle\mathfrak{S}\rangle$, $M \in \mathsf{Mod}(\Psi)$ iff $\Psi \subseteq \Phi^+$.

The next theorem is proven in Fagin (1982b). It is copied almost verbatim, with only minor notational changes, as parts (a) and (b) of Theorem 3.1 of that paper.

3.16. *Faithfulness and Armstrong Models*

Let $\mathcal{S}$ be a set of sentences. The following properties of $\mathcal{S}$ are equivalent.

(a) Existence of a faithful operator. There is an operator $\boxplus$ that maps nonempty families of models into models, such that if φ is a sentence in $\mathcal{S}$ and $\{R_i \mid i \in I\}$ is a nonempty family of models, then φ holds for $\boxplus\{R_i \mid i \in I\}$ if and only if holds for each R_i.

(b) Existence of Armstrong models. Whenever Ψ is a consistent subset of $\mathcal{S}$ and Ψ^+ is the set of sentences in $\mathcal{S}$ that are logical consequences of Ψ, then there is a model (an "Armstrong model") that obeys Ψ^+ and no other sentences in $\mathcal{S}$.

3.17. *A Faithful Operator for Granule Structures*

In order to apply Section 3.16 to the context of granule rules, it is necessary to identify a suitable faithful operator. To this end, let $S = \{\sigma_i = (\mathsf{Dom}\langle\sigma_i\rangle, \mathsf{GnletoDom}_{\sigma_i}) \mid i \in I\} \subseteq \mathsf{GranStruct}\langle\mathfrak{S}\rangle$ be a nonempty set of structures, and assume, without loss of generality, that $\mathsf{Dom}\langle\sigma_i\rangle \cap \mathsf{Dom}\langle\sigma_j\rangle = \emptyset$ for all $i, j \in I$ with $i \neq j$.[3] Define the *product* of S to be the structure $\boxplus\langle S\rangle = (\mathsf{Dom}\langle\boxplus\langle S\rangle\rangle, \mathsf{GnletoDom}_{\boxplus\langle S\rangle})$ with $\mathsf{Dom}\langle\boxplus\langle S\rangle\rangle = \bigcup\{\mathsf{Dom}\langle\sigma_i\rangle \mid i \in I\}$ with $\mathsf{GnletoDom}_{\boxplus\langle S\rangle} : \mathsf{Granules}\langle\mathfrak{S}\rangle \to \mathbf{2}^{\bigcup\{\mathsf{Dom}\langle\sigma_i\rangle \mid i \in I\}}$ defined on elements by $g \mapsto \bigcup\{\mathsf{GnletoDom}_{\sigma_i}(g) \mid i \in I\}$. Let $\mathsf{Products}\langle\mathsf{GranStruct}\langle\mathfrak{S}\rangle\rangle$ denote the collection of all such products.

3.18. *Armstrong Models for Sets of Granule Rules*

Every consistent subset $\Phi \subseteq \mathsf{GrRules}\langle\mathfrak{S}\rangle$ has an Armstrong model (with respect to $\mathsf{GrRules}\langle\mathfrak{S}\rangle$).

Proof. It suffices to show that the product operator defined in Section 3.17 is faithful, and then to apply 3.16. Let $S = \{\sigma_i = (\mathsf{Dom}\langle\sigma_i\rangle, \mathsf{GnletoDom}_{\sigma_i}) \mid i \in I\} \subseteq \mathsf{GranStruct}\langle\mathfrak{S}\rangle$

[3]This restriction may be relaxed by working with disjoint union instead of ordinary union; the details are left to the reader.

be a nonempty set of structures satisfying $\mathsf{Dom}\langle\sigma_i\rangle \cap \mathsf{Dom}\langle\sigma_j\rangle = \emptyset$ for all $i, j \in I$ with $i \neq j$. For a rule of the form $e_1 \sqsubseteq_{\mathfrak{S}} e_2$ for $e_1, e_2 \in \mathsf{GrExpr}\langle\mathfrak{S}\rangle$, if $\mathsf{GnletoDom}_{\sigma_i}(e_1) \subseteq \mathsf{GnletoDom}_{\sigma_i}(e_2)$ holds for all $i \in I$, then $\bigcup\{\mathsf{GnletoDom}_{\sigma_i}(g) \mid i \in I\} \subseteq \bigcup\{\mathsf{GnletoDom}_{\sigma_i}(s) \mid i \in I \text{ and } s \in S\}$. On the other hand, if $\mathsf{GnletoDom}_{\sigma_i}(e_1) \subseteq \mathsf{GnletoDom}_{\sigma_i}(e_2)$ fails to hold for some $i \in I$, then $\bigcup\{\mathsf{GnletoDom}_{\sigma_i}(g) \mid i \in I\} \subseteq \bigcup\{\mathsf{GnletoDom}_{\sigma_i}(s) \mid i \in I \text{ and } s \in S\}$ cannot hold, since $\mathsf{GnletoDom}_{\sigma_i} \cap \mathsf{GnletoDom}_{\sigma_j} = \emptyset$ for $i \neq j$. Thus, $\boxplus$ is faithful for constraints of the form $e_1 \sqsubseteq_{\mathfrak{S}} e_2$, as required. The extension to rules of the form $\bigwedge\{\varphi_i \mid i \in [1, k]\}$, with φ_j of the form $(e_{j1} \sqsubseteq_{\mathfrak{S}} e_{j2})$, is a simple extension of the above. The details are omitted. □

3.19. *Negation of Rules*

In order to develop a proper theory of closed-world semantics (see Section 3.20 below), it is necessary to work with negations of rules. The semantics are the natural one extension of (rulemod-i)–(rulemod-ii) of 3.8. Given $\varphi \in \mathsf{GrRules}\langle\mathfrak{S}\rangle$, $\sigma \in \mathsf{GranStruct}\langle\mathfrak{S}\rangle$ is a model of $\neg(\varphi)$ simply means that φ does not hold in σ; so $\mathsf{Models}_{\mathfrak{S}}\langle\neg\varphi\rangle = \{\sigma \in \mathsf{GranStruct}\langle\mathfrak{S}\rangle \mid \sigma \notin \mathsf{Models}_{\mathfrak{S}}\langle\varphi\rangle\}$. Following standard mathematical conventions, $e_1 \neq e_2$ and $e_1 \not\sqsubseteq_{\mathfrak{S}} e_2$ are shorthand for $\neg(e_1 = e_2)$ and $\neg(e_1 \sqsubseteq_{\mathfrak{S}} e_2)$, respectively. Given $\Phi \subseteq \mathsf{GrRules}\langle\mathfrak{S}\rangle$ and $\varphi \in \mathsf{GrRules}\langle\mathfrak{S}\rangle$, observe that $\Phi \models_{\mathfrak{S}} \neg(\varphi)$ means that φ never holds when Φ holds; i.e. that $\Phi \cup \{\varphi\}$ is unsatisfiable.

3.20. *Closed-World Assumptions*

The use of sets of rules to specify the semantics of a granulated attribute schema provides great flexibility; constraints may be specified without the need to identify and represent a specific (and typically very detailed) structure, such as spatial coordinates for geographic regions. Nevertheless, it is useful to require that some basic information always be complete. This is accomplished via *closed-world assumptions (CWAs)* (Reiter, 1978; Clark, 1978), in which certain (or all) statements (here, rules) which cannot be proven true are taken to be false.

A *CWA pair* for $\mathfrak{S}$ is an ordered pair $\langle\Phi, \Psi\rangle \subseteq \mathsf{GrRules}\langle\mathfrak{S}\rangle \times \mathsf{GrRules}\langle\mathfrak{S}\rangle$. In the *CWA closure of* $\langle\Phi, \Psi\rangle$, every rule in Ψ which is not a consequence of Φ is taken to be false. Formally, define this closure to be $\overline{\mathsf{CWA}}_{\mathfrak{S}}\langle\Phi, \Psi\rangle = \Phi \cup \{\neg\varphi \mid \varphi \in \Psi \text{ and } \Phi \not\models_{\mathfrak{S}} \varphi\}$. In general, a CWA may lead to contradictions (Reiter, 1978; Example 8). However, this cannot happen in the context of granule rules, thanks to the existence of Armstrong models. Indeed, any Armstrong model of Φ, as guaranteed by 3.18, must also satisfy all sentences in $\overline{\mathsf{CWA}}_{\mathfrak{S}}\langle\Phi, \Psi\rangle$.

Given a CWA pair $\langle\Phi, \Psi\rangle$, a rule $\varphi \in \mathsf{GrRules}\langle\mathfrak{S}\rangle$ is *resolvable* from $\overline{\mathsf{CWA}}_{\mathfrak{S}}\langle\Phi, \Psi\rangle$ if either $\overline{\mathsf{CWA}}_{\mathfrak{S}}\langle\Phi, \Psi\rangle \models_{\mathfrak{S}} \varphi$ or else $\overline{\mathsf{CWA}}_{\mathfrak{S}}\langle\Phi, \Psi\rangle \models_{\mathfrak{S}} \neg\varphi$. In other words, φ is resolvable if its truth value may be determined under from Φ with the CWA on Ψ. A set $\Omega \subseteq \mathsf{GrRules}\langle\mathfrak{S}\rangle$ is *resolvable* from $\overline{\mathsf{CWA}}_{\mathfrak{S}}\langle\Phi, \Psi\rangle$ if every $\varphi \in \Omega$ has that property.

As a specific example for $\overline{\mathsf{CWA}}_{\mathfrak{S}}\langle\Phi, \Psi\rangle$, let Ψ be the set consisting of all elemental subsumption rules and all basic disjointness rules (see Section 3.10). The resulting CWA then enforces that all such rules which are not in $(\mathsf{BuiltInRules}\langle\mathfrak{S}\rangle \cup \Phi)^+$ are taken to be false. In particular, if Φ adds no new elementary subsumption or basic disjointness rules, then this CWA enforces that the only such rules are those in $\mathsf{BuiltInRules}\langle\mathfrak{S}\rangle$.

3.21. *Constrained Granulated Attribute Schemata*

A *constrained granularity schema* is a four-tuple $(\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle, \mathsf{Constr}\langle\mathfrak{S}\rangle, \mathsf{cwa}\langle\mathfrak{S}\rangle)$ in which $(\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle)$ is a granulated attribute schema, with $\mathsf{Constr}\langle\mathfrak{S}\rangle, \mathsf{cwa}\langle\mathfrak{S}\rangle \subseteq \mathsf{GrRules}\langle\mathfrak{S}\rangle$. For these last two sets, $\mathsf{Constr}\langle\mathfrak{S}\rangle$ is the main set of constraints governing the schema, with $\langle\mathsf{Constr}\langle\mathfrak{S}\rangle, \mathsf{cwa}\langle\mathfrak{S}\rangle\rangle$ the CWA pair which governs the schema. Thus, $\mathsf{cwa}\langle\mathfrak{S}\rangle$ is the set of constraints which are taken to be false, under the CWA, if they do not follow from $\mathsf{Constr}(\mathfrak{S})$.

$\mathsf{CWAConstr}\langle\mathfrak{S}\rangle$ is shorthand for $\overline{\mathsf{CWA}}_{\mathfrak{S}}\langle\mathsf{Constr}\langle\mathfrak{S}\rangle, \mathsf{cwa}\langle\mathfrak{S}\rangle\rangle$.

As a slight abuse of notation, $\mathfrak{S}$ will be used to denote this constrained granularity schema (as well as the unconstrained one $(\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle)$).

If $\Omega \subseteq \mathsf{GrRules}\langle\mathfrak{S}\rangle$ is resolvable from $\langle\mathsf{Constr}\langle\mathfrak{S}\rangle, \mathsf{cwa}\langle\mathfrak{S}\rangle\rangle$, say that Ω is *resolvable* within $\mathfrak{S}$.

3.22. *Comparison to Previous Work on Multigranular Attributes*

Bravo and Rodríguez (2014) use an approach based upon structures, similar in some ways to that described in Section 3.5. However, in that work, only basic subsumption rules (constraints corresponding to those of (brule-i) of Section 3.10) are considered. In that more limited case, a single structure as a model suffices, since knowing the complete ordering is reasonable. In Hegner and Rodríguez (2016), join constraints (including the disjoint variety) similar to those described in (brule-ii) and (brule-iii) of Section 3.10 are modelled. In contrast to the approach taken here, in that work rules are defined first, and consistency is defined later, in a less than completely rigorous fashion. The approach taken here is the more natural one, paralleling the approach taken in mathematical logic, with all steps spelled out carefully.

3.23. *Comparison to the Partition Model*

In the partition model for the representation of attributes, as described in Spyratos (1987) and Molnár (2007), there is a single base set, with the granules (there called domain elements) of each attribute modelled as a partition on that common base set. As is the case with the model presented here, each granule is modelled as a block of the partition. The relationship between values of different attributes is then recaptured via intersection of the blocks which represent them. Thus, there is a significant connection between the two models. However, to use this approach, it would be necessary to require that for every granularity G and for every structure σ, every element $x \in \mathsf{Dom}\langle\sigma\rangle$ lie in $\mathsf{GnletoDom}_{\sigma}(g)$ for some $g \in \mathsf{Granules}\langle\mathfrak{S}|G\rangle$. For granularities whose granules do not join to the top granule, this would be an unnecessary condition, requiring the introduction of artificial granules. For example, consider the granularity City for a multigranular attribute which models all of Chile. Most of the land of Chile does not lie within any city. Thus, if the domain $\mathsf{Dom}\langle\sigma_{\mathsf{Chile}}\rangle$ is all of Chile, then most of the elements of $\mathsf{Dom}\langle\sigma_{\mathsf{Chile}}\rangle$ will not lie in any granule of the granularity City. This can be repaired, to achieve a total partition of $\mathsf{Dom}\langle\sigma_{\mathsf{Chile}}\rangle$, by introducing a special granule *NoCity*, and putting all surface points not lying within a city in that granule. However, that is an awkward solution, if for no other reason than *NoCity* is not a city, despite lying in the granularity City. It introduces other complications as well.

To address this issue, define a *partial partition* of a set X to be a set $\mathbf{P}$ consisting of nonempty disjoint subsets of X. In other words, a partial partition consists of some, but not necessarily all, of the blocks of an ordinary partition. It is easy to see from (grstr-ii) of Section 3.5 that for a structure σ, each granularity induces a partial partition on $\mathsf{Dom}\langle\sigma\rangle$. Thus, the approach presented here is related to the partition model in that it provides a model of granules based upon partial partitions.

4. Thematic Attributes and Aggregation

As mentioned in Section 1, granular attributes may be classified as *basic* (often spatio-temporal) and *thematic*, which record values and upon which aggregation is applied. For example, in the schema $R_{\mathsf{sumb}}\langle A_{\mathsf{Plc}}, A_{\mathsf{Tim}}, B_{\mathsf{Bth}}\rangle$, A_{Plc} and A_{Tim} are basic, while B_{Bth} is thematic. While thematic attributes often have a simpler granular structure than their basic counterparts, they must also be embellished with sufficient structure to allow aggregation operators to be applied to their granules. In this section, the special aspects of thematic attributes is developed.

4.1. *Complete Subset Attribute Schemata*

A *complete subset attribute schema* over a set S is a unified attribute schema $\mathfrak{S} = (\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle)$ with the following properties.

(csas-i) $\mathbf{Glty}\langle\mathfrak{S}\rangle$ has a least (finest) granularity, called the *base granularity* and denoted $\mathsf{BaseGlty}\langle\mathfrak{S}\rangle$. All other granularities, including $\top_{\mathbf{Glty}\langle\mathfrak{S}\rangle}$, are called *grouping granularities*; the set of all such granularities is denoted $\mathsf{GrpGlty}\langle\mathfrak{S}\rangle$.

(csas-ii) $\mathsf{Granules}\langle\mathfrak{S}|\mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle = S$.

(csas-iii) For each $G \in \mathsf{GrpGlty}\langle\mathfrak{S}\rangle$ and each $g \in \mathsf{Granules}\langle\mathfrak{S}|G\rangle$, g is of the form $\langle G, v\rangle$ with $v \in \mathbf{2}^S \setminus \emptyset$. In the pair $\langle G, v\rangle$, G is called the *granularity tag* and v is called the *value*. As a convenient notation, define $\mathsf{Tag}_{\mathfrak{S}}(\langle G, v\rangle) = G$ and $\mathsf{Val}_{\mathfrak{S}}(\langle G, v\rangle) = v$. For granularity $G \in \mathsf{GrpGlty}\langle\mathfrak{S}\rangle$, as a slight abuse of notation, $\mathsf{Val}(G)$ is used to denote $\{\mathsf{Val}(g) \mid g \in \mathsf{Granules}\langle\mathfrak{S}|G\rangle\}$.

(csas-iv) For each $G \in \mathsf{GrpGlty}\langle\mathfrak{S}\rangle$, $\mathsf{Val}(G)$ forms a partition of $\mathsf{BaseGlty}\langle\mathfrak{S}\rangle$. In particular, $\mathsf{Granules}\langle\mathfrak{S}|\top_{\mathbf{Glty}\langle\mathfrak{S}\rangle}\rangle = \{\langle\top_{\mathbf{Glty}\langle\mathfrak{S}\rangle}, S\rangle\}$.

(csas-v) The granule preorder $\mathbf{Gnle}\langle\mathfrak{S}\rangle = (\mathsf{Granules}\langle\mathfrak{S}\rangle, \bar{\sqsubseteq}_{\mathfrak{S}}, \top_{\mathfrak{S}}, \bot_{\mathfrak{S}})$ is defined by $g_1 \bar{\sqsubseteq}_{\mathfrak{S}} g_2$ iff either $\mathsf{Tag}(g_1) = \mathsf{BaseGlty}\langle\mathfrak{S}\rangle$, $\mathsf{Tag}(g_2) \in \mathsf{GrpGlty}\langle\mathfrak{S}\rangle$, and $\mathsf{Val}(g_1) \in \mathsf{Val}(g_2)$; or else $\mathsf{Tag}(g_1), \mathsf{Tag}(g_2) \in \mathsf{GrpGlty}\langle\mathfrak{S}\rangle$ with $\mathsf{Val}(g_1) \subseteq \mathsf{Val}(g_2)$.

Some special notation for things which are used frequently is in order. First of all, since the set S is not mentioned explicitly in the pair $\mathfrak{S} = (\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle)$, it is useful to have a notation to recover it. To this end, define the set of *base granules* of $\mathfrak{S}$ to be $\mathsf{BaseGranules}\langle\mathfrak{S}\rangle = \mathsf{Granules}\langle\mathfrak{S}|\mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle$. On the other hand, define the set of *group granules* of $\mathfrak{S}$ to be $\mathsf{GrpGranules}\langle\mathfrak{S}\rangle = \bigcup\{\mathsf{Granules}\langle\mathfrak{S}|G\rangle \mid G \in \mathsf{GrpGlty}\langle\mathfrak{S}\rangle\}$.

It is important to note that the values of granules in $\mathsf{GrpGlty}\langle\mathfrak{S}\rangle$ are **sets** of elements from $\mathsf{BaseGlty}\langle\mathfrak{S}\rangle$. It is necessary to maintain this distinction in order to facilitate the application of this concept to aggregation (see 4.12). It is nevertheless useful to introduce a notation which unifies them. To this end, for $g \in \mathsf{Granules}\langle\mathfrak{S}\rangle$, define $\tilde{g}$ to be $\mathsf{Val}(g)$ if $g \in \mathsf{GrpGlty}\langle\mathfrak{S}\rangle$, $\{g\}$ if $g \in \mathsf{BaseGlty}\langle\mathfrak{S}\rangle$, and $\emptyset$ if $g = \bot_{\mathfrak{S}}$. In other words, the transformation $g \mapsto \tilde{g}$ maps each base granule to a singleton set containing that granule, each group granule to its value, and $\bot_{\mathfrak{S}}$ to $\emptyset$. Thus, $\{\tilde{g} \mid g \in \mathsf{Granules}_{\not\bot}\langle\mathfrak{S}\rangle\}$ consists, uniformly, of subsets of $\mathsf{BaseGlty}\langle\mathfrak{S}\rangle$. With this notation, (csas-v) is expressed more compactly as $g_1 \mathrel{\bar{\sqsubseteq}_{\mathfrak{S}}} g_2$ iff $\tilde{g_1} \subseteq \tilde{g_2}$.

The *standard structure* $\mathsf{StdStr}\langle\mathfrak{S}\rangle$ for $\mathfrak{S}$ has $\mathsf{Dom}\langle\mathsf{StdStr}\langle\mathfrak{S}\rangle\rangle = \mathsf{BaseGlty}\langle\mathfrak{S}\rangle$ with $\mathsf{GnletoDom}_{\mathsf{StdStr}\langle\mathfrak{S}\rangle}$ defined by $g \mapsto \tilde{g}$ for all $g \in \mathsf{Granules}_{\not\bot}\langle\mathfrak{S}\rangle$. It is easy to verify that $\mathsf{StdStr}\langle\mathfrak{S}\rangle$ forms a structure for $\mathfrak{S}$ in the sense defined in Section 3.5.

For subset attribute schemata, the standard structure is always used to define the semantics. The *standard constraint set* for $\mathsf{StdStr}\langle\mathfrak{S}\rangle$ is $\{\varphi \in \mathsf{BaRules}\langle\mathfrak{S}\rangle \mid \mathsf{StdStr}\langle\mathfrak{S}\rangle \in \mathsf{Mod}(\varphi)\}$, and is denoted $\mathsf{StdConstr}\langle\mathfrak{S}\rangle$. The *standard constrained granulated attribute schema* (or just *standard schema*) of $\mathfrak{S}$ is as defined in Section 3.21, with $\mathsf{Constr}(\boldsymbol{\mathfrak{S}}) = \mathsf{StdConstr}\langle\mathfrak{S}\rangle$. In particular, for $g_1, g_2 \in \mathsf{Granules}\langle\mathfrak{S}\rangle$, $g_1 \sqsubseteq_{\mathfrak{S}} g_2$ iff $g_1 \mathrel{\bar{\sqsubseteq}_{\mathfrak{S}}} g_2$ iff $\tilde{g}_1 \subseteq \tilde{g}_2$.

In the case of a complete subset attribute schema $\mathfrak{S}$, for any $G \in \mathsf{GrpGlty}\langle\mathfrak{S}\rangle$, $\{\mathsf{GnletoDom}_\sigma(g) \mid g \in \mathsf{Granules}\langle\mathfrak{S}|G\rangle\}$ forms a full (and not just partial) partition of $\mathsf{Dom}\langle\sigma\rangle$ (compare to Section 3.23).

4.2. *Examples*

For $m \in \mathbb{N}$, let $\mathsf{RndSch}^m_{\mathbb{Z}}$ denote the subset attribute schema with $\mathsf{BaseGlty}\langle\mathsf{RndSch}^{\mathsf{m}}_{\mathbb{Z}}\rangle$ denoted $\mathsf{Gran}\mathbb{Z}$, $\mathsf{BaseGranules}\langle\mathsf{RndSch}^m_{\mathbb{Z}}\rangle = \mathsf{Granules}\langle\mathsf{RndSch}^m_{\mathbb{Z}}|\mathsf{Gran}\mathbb{Z}\rangle = \mathbb{Z}$, and for each i with $1 \leqslant i \leqslant m$, the granularity round_i has
$\mathsf{Granules}\langle\mathsf{RndSch}^m_{\mathbb{Z}}|\mathsf{round}_i\rangle = \{\langle\mathsf{round}_i, [10^i \cdot (j-0.5), 10^i \cdot (j+0.5))\rangle \mid 1 \leqslant j \leqslant m\}$; it identifies rounding to the nearest 10^i. $[n_1, n_2)$ represents the *clopen* interval $\{x \in \mathbb{Z} \mid n_1 \leqslant x < n_2\}$. To illustrate less formally, the values of round_2 consist of clopen intervals of the form $[\ell - 50, \ell + 50)$ for ℓ a multiple of 100. $\mathsf{Gran}\mathbb{Z}$ is also denoted round_0, since it recaptures rounding to the nearest $10^0 = 1$; i.e. no rounding at all. In addition, there is the top granularity $\top_{\mathbf{Glty}\langle\mathsf{RndSch}^{\mathbf{m}}_{\mathbb{Z}}\rangle}$, which is also denoted round_∞, since it rounds every integer to the same value, thus preserving no information; its sole granule has as its value the entire set $\mathbb{Z}$.

It is important to observe that the granules of $\mathsf{Gran}\mathbb{Z} = \mathsf{round}_0$ are actual numbers; that is, elements of $\mathbb{Z}$. On the other hand, elements of round_i for $i > 0$ have values which are *sets* of integers; for $g \in \mathsf{Granules}\langle\mathsf{RndSch}^m_{\mathbb{Z}}|\mathsf{round}_i\rangle$, $\mathsf{Val}(g)$ consists of all integers which round (for i significant digits) to the same values. This distinction is necessary in order to allow the modelling of aggregation operators (see Section 4.12) to be as natural as possible.

The ordering on grouping granules is defined by subset inclusion of their values. For example, $\langle \mathsf{round}_1, [375, 385)\rangle \sqsubseteq \langle \mathsf{round}_2, [350, 450)\rangle$, with $[350, 450]$ representing all integers which round to 400, when the rounding is to two significant digits. Similarly, $[375, 385)$ represents all integers which round to 380, when rounding is to the nearest 10. When one of the granules is an integer (i.e. a base granule) and not a set of integers, the ordering is defined by membership; e.g. $400 \sqsubseteq \langle \mathsf{round}_2, [350, 450)\rangle$. It is critical to observe that this ordering does not embody the usual ordering of integers; $300 \sqsubseteq 400$ does **not** hold in this model. Indeed, $\bigsqcap_{\mathsf{RndSch}^m_{\mathbb{Z}}} \{300, 400\} = \bot_{\mathfrak{S}}$, since both numbers belong to the same granularity. Rather, $300 \leqslant 400$ is recaptured by the *thematic ordering*, discussed in Section 4.5 below.

A similar construction applies in the case that the base set is the natural numbers $\mathbb{N}$, instead of the integers $\mathbb{Z}$, to yield the subset attribute schema $\mathsf{RndSch}^m_{\mathbb{N}}$. The grouping granules are just those for $\mathsf{RndSch}^m_{\mathbb{Z}}$, intersected with $\mathbb{N}$; i.e. discarding the negative numbers. Note in particular that the clopen interval $[-10^i, 10^i) \in \mathsf{Val}_{\mathsf{RndSch}^m_{\mathbb{Z}}}(\mathsf{round}_i)$ is truncated to $[0, 10^i) \in \mathsf{Val}_{\mathsf{RndSch}^m_{\mathbb{N}}}(\mathsf{round}i)$. All clopen intervals in $\mathsf{Granules}\langle \mathsf{RndSch}^m_{\mathbb{Z}} | \mathsf{round}_i\rangle$ containing only nonnegative numbers are retained in $\mathsf{Granules}\langle \mathsf{RndSch}^m_{\mathbb{N}} | \mathsf{round}_i\rangle$, while those containing only negative numbers are discarded.

The attribute B_{Bth}, Section 1, and elaborated further at the end of Section 3.1, provides a concrete application, with $\mathfrak{S}_{B_{\mathsf{Bth}}} = \mathsf{RndSch}^m_{\mathbb{N}}$ for a suitable choice of m.

These ideas may also be extended to obtain $\mathsf{RndSch}^m_{\mathbb{R}}$ over the real numbers, although in that case the least granularity is not round_0. The details are left to the reader.

4.3. *Uniqueness of Subsuming Granules*

Let $\mathfrak{S}$ be any multigranular attribute schema (not necessarily a subset schema). Given $g_1, g_2, g_2' \in \mathsf{Granules}\langle\mathfrak{S}\rangle$ with $g_1 \sqsubseteq_{\mathfrak{S}} g_2$, $g_1 \sqsubseteq_{\mathfrak{S}} g_2'$, and $G_2 \in \mathsf{Glty}\langle\mathfrak{S}\rangle$ with $g_2, g_2' \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$, it must be the case that $g_2 = g_2'$.

Proof. Let g_1, g_2, g_2' and G_2 be as stated. By (grstr-i), if $g_1 \neq g_2$, then $\bigsqcap_{\mathfrak{S}} \{g_2, g_2'\} = \bot_S$. However, $g_1 \sqsubseteq_{\mathfrak{S}} \bigsqcap_{\mathfrak{S}} \{g_2, g_2'\}$, whence it must be the case that $g_2 = g_2'$. □

4.4. *Coarsening*

The concepts in this paragraph apply to any multigranular attribute schema $\mathfrak{S}$; not just those which are subset schemata.

In order to support the management of source data at differing granularities, it is often necessary to reduce them to a common granularity. The operation of coarsening, which transforms a granule to a one at a coarser granularity, is central to this idea. Formally, for $G_1, G_2 \in \mathsf{Glty}\langle\mathfrak{S}\rangle$, the function $\mathsf{Coarsen}_{\mathfrak{S}} G_1 G_2 : \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle \to \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$ is defined on $g_1 \in \mathsf{Granules}\langle\mathfrak{S}|G_1\rangle$ iff there is a $g_2 \in \mathsf{Granules}\langle\mathfrak{S}|G_2\rangle$ with $g_1 \sqsubseteq_{\mathfrak{S}} g_2$. In view of 4.3, this g_2 is unique whenever it exists. In this case $g_2 = \mathsf{Coarsen}_{\mathfrak{S}}\, G_1 G_2(g_1)$. In general, $\mathsf{Coarsen}_{\mathfrak{S}}\, G_1 G_2$ is a partial function; it is total precisely in the case that $G_1 \leqslant_{\mathbf{Glty}\langle\sigma\rangle} G_2$ (use (grstr-iii) of Section 3.5).

Closely related is the partial function $\mathsf{Map} : \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle \times \mathsf{Glty}\langle\mathfrak{S}\rangle \to \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$ (Bravo and Rodríguez, 2014) defined on elements by $(G, g) \mapsto \mathsf{Coarsen}_{\mathfrak{S}}\, \mathsf{Glty}_{\mathfrak{S}}(g) G(g)$.

In the spatial context of A_{Plc}, the city of Concepción lies in Region VIII of Chile. This would be represented by the coarsening $\mathsf{Coarsen}_{\mathfrak{S}_{A_{\mathsf{Plc}}}}\, \mathsf{CityRegion}(\mathit{Concepción}) = \mathit{Región\_VIII}$. Similarly, in the temporal context of A_{Tim}, quarter 1 of year 2014 lies with 2014; this would be represented by the coarsening

$$\mathsf{Coarsen}_{\mathfrak{S}_{A_{\mathsf{Tim}}}}\, \mathsf{QuarterYrYear}(\mathit{Q1Y2014}) = \mathit{2014}.$$

In this work, coarsening is used primarily on the granules of complete subset attribute schema. For example, in the context of Section 4.2,

$$\mathsf{Coarsen}_{\mathfrak{S}_{B_{\mathsf{Bth}}}}\, \mathsf{round}_1\mathsf{round}_2(\langle\mathsf{round}_1, [375, 385)\rangle) = \langle\mathsf{round}_2, [350, 450)\rangle \text{ and}$$
$$\mathsf{Coarsen}_{\mathfrak{S}_{B_{\mathsf{Bth}}}}\, \mathsf{round}_0\mathsf{round}_1(376) = \langle\mathsf{round}_1, [375, 385)\rangle.$$

4.5. *Thematic Attributes and Orderings*

Following common usage in geographic information systems (Bonham-Carter, 1995), a *thematic attribute* is used to record values associated with aggregating non-thematic (e.g. spatial or temporal) attributes. The attribute B_{Bth} and its associated granularity schema $\mathfrak{S}_{B_{\mathsf{Bth}}}$ form a typical example.

In this work, the underlying schemata of such attributes are modelled as subset attribute schemata with the further property that the granules of the least granularity are endowed with an additional order, called the *thematic order*. In the example of Section 4.2 above, in which $\mathsf{BaseGranules}\langle\mathfrak{S}_{B_{\mathsf{Bth}}}\rangle = \mathbb{Z}$, this order is the usual total order on the integers. This is quite distinct from the ordinary granule ordering; indeed for any granulated attribute schema $\mathfrak{S}$, two granules of the same granularity are never related via either $\bar{\sqsubseteq}_{\mathfrak{S}}$ or $\sqsubseteq_{\mathfrak{S}}$.

Let $\mathfrak{S}$ be a subset attribute schema and let $\leqslant_{\mathfrak{S}}$ be a partial order on $\mathsf{BaseGranules}\langle\mathfrak{S}\rangle$. Extend this order to all of $\mathsf{Granules}\langle\mathfrak{S}\rangle$ by defining $g_1 \leqslant^+_{\mathfrak{S}} g_2$ to hold iff $(\forall x \in \tilde{g}_1)(\forall y \in \tilde{g}_2)(x \leqslant_{\mathfrak{S}} y)$. Then, define an *attribute schema with thematic order* to be a triple $\mathfrak{S} = (\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle, \leqslant_{\mathfrak{S}})$ in which $(\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle)$ is a unified subset attribute schema and $\leqslant_{\mathfrak{S}}$ is a partial order on $\mathsf{BaseGlty}\langle\mathfrak{S}\rangle$ with the property that for each $G \in \mathsf{GrpGlty}\langle\mathfrak{S}\rangle$, the function $\mathsf{Coarsen}_{\mathfrak{S}}\langle\mathsf{BaseGlty}\langle\mathfrak{S}\rangle \lhd G\rangle : \mathsf{Granules}\langle\mathfrak{S}|G\rangle \to \mathsf{Granules}\langle\mathfrak{S}|\mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle$ is order embedding (Davey and Priestly, 2002; 1.34(ii)). In other words, the following condition must hold.

(tho-i) $(\forall G \in \mathsf{Glty}\langle\mathfrak{S}\rangle)$
$(\forall(g_1, g_2) \in \mathsf{Granules}\langle\mathfrak{S}|\mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle \times \mathsf{Granules}\langle\mathfrak{S}|\mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle$,
$(g_1 \leqslant_{\mathfrak{S}} g_2$
iff $\mathsf{Coarsen}_{\mathfrak{S}}\langle\mathsf{BaseGlty}\langle\mathfrak{S}\rangle \lhd G\rangle(g_1) \leqslant^+_{\mathfrak{S}} \mathsf{Coarsen}_{\mathfrak{S}}\langle\mathsf{BaseGlty}\langle\mathfrak{S}\rangle \lhd G\rangle(g_2))$.

4.6. *Convex Subsets of Posets*

Let $\mathbf{P} = (P, \leqslant_{\mathbf{P}})$ be a poset. A subset $Q \subseteq P$ is said to be *convex* in $\mathbf{P}$ (Davey and Priestly, 2002; Exer. 2.28) if for any $x, y \in Q$ and $z \in P$, if $x \leqslant z \leqslant y$, then $z \in Q$.

For sets of numbers, such as $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}$, the convex sets are just intervals.

4.7. *Convexity Implies Thematic Order*

Let $\mathfrak{S} = (\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle)$ be a unified subset attribute schema, and let $\leqslant_{\mathfrak{S}}$ be a partial order on $\mathsf{Granules}\langle\mathfrak{S}|\mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle$. If every element of $\mathsf{GrpGranules}\langle\mathfrak{S}\rangle$ is convex, then $\mathfrak{S} = (\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle, \leqslant_{\mathfrak{S}})$ is an attribute schema with thematic order.

Proof. Straightforward verification. □

4.8. *Example*

Using Section 4.7, it is clear that $\mathsf{RndSch}^m_{\mathbb{Z}} = (\mathbf{Glty}\langle\mathsf{RndSch}^{\mathbf{m}}_{\mathbb{Z}}\rangle, \mathsf{GrAsgn}\langle\mathsf{RndSch}^m_{\mathbb{Z}}\rangle, \leqslant)$, with $\leqslant$ the usual numerical ordering on $\mathbb{Z}$, is an attribute schema with thematic order for any natural number m.

4.9. *Notational Convention*

Throughout the rest of this section, unless stated specifically to the contrary, take $\mathfrak{S} = (\mathbf{Glty}\langle\mathfrak{S}\rangle, \mathsf{GrAsgn}\langle\mathfrak{S}\rangle, \leqslant_{\mathfrak{S}})$ to be an attribute schema with thematic order, as defined in Section 4.5.

4.10. *Granular Refinement for Thematic Attributes*

Refinement is a sort of inverse to coarsening. The idea is to take a compound granule $g \in \mathsf{GrpGranules}\langle\mathfrak{S}\rangle$ and map it to a base granule $g' \in \mathsf{BaseGranules}\langle\mathfrak{S}\rangle$ with the property that $g' \in g$. For example, if $\tilde{g}$ is the clopen interval $[350, 450)$, it might be mapped to $g' = 400$. This example is explored more thoroughly in Section 4.11 below.

Formally, a *refinement family* for $\mathfrak{S}$ is a set

$\{\mathsf{Refine}_{\mathfrak{S}}\langle G \rhd \mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle : \mathsf{Granules}\langle\mathfrak{S}|G\rangle \to \mathsf{Granules}\langle\mathfrak{S}|\mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle \mid G \in \mathsf{Glty}\langle\mathfrak{S}\rangle\}$ of functions with the following properties.

(ref-i) For all $G \in \mathsf{Glty}\langle\mathfrak{S}\rangle$,

$$\mathsf{Coarsen}_{\mathfrak{S}}\langle\mathsf{BaseGlty}\langle\mathfrak{S}\rangle \lhd G\rangle \circ \mathsf{Refine}_{\mathfrak{S}}\langle G \rhd \mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle = \mathbf{1}_{\mathsf{Granules}\langle\mathfrak{S}|G\rangle},$$

where $\mathbf{1}_{\mathsf{Granules}\langle\mathfrak{S}|G\rangle}$ is the identity function on $\mathsf{Granules}\langle\mathfrak{S}|G\rangle$. In other words, $\mathsf{Refine}_{\mathfrak{S}}\langle G \rhd \mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle$ is a left inverse of $\mathsf{Coarsen}_{\mathfrak{S}}\langle\mathsf{BaseGlty}\langle\mathfrak{S}\rangle \lhd G\rangle$.

(ref-ii) For all $G_1, G_2 \in \mathsf{Glty}\langle\mathfrak{S}\rangle$ with $G_1 \leqslant_{\mathbf{Glty}\langle\mathfrak{S}\rangle} G_2$,

$$\mathsf{Coarsen}_{\mathfrak{S}}\langle G_1 \lhd G_2\rangle = \mathsf{Coarsen}_{\mathfrak{S}}\langle\mathsf{BaseGlty}\langle\mathfrak{S}\rangle \lhd G_2\rangle \circ \mathsf{Refine}_{\mathfrak{S}}\langle G_1 \rhd \mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle.$$

For $G_1, G_2 \in \mathsf{Glty}\langle\mathfrak{S}\rangle$ with $G_1 \leqslant_{\mathsf{Glty}\langle\mathfrak{S}\rangle} G_2$, and a refinement family as just defined, define $\mathsf{Refine}_{\mathfrak{S}}\langle G_2 \rhd G_1\rangle = \mathsf{Coarsen}_{\mathfrak{S}}\langle \mathsf{BaseGlty}\langle\mathfrak{S}\rangle \lhd G_1\rangle \circ \mathsf{Refine}_{\mathfrak{S}}\langle G_2 \rhd \mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle$.

Observe that $\mathsf{Refine}_{\mathfrak{S}}\langle \mathsf{BaseGlty}\langle\mathfrak{S}\rangle \rhd \mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle$ must be the identity function on $\mathsf{BaseGranules}\langle\mathfrak{S}\rangle$, so that this definition reduces to that given by the refinement family for $G_1 = \mathsf{BaseGlty}\langle\mathfrak{S}\rangle$.

4.11. *Example*

Continue with the schema $\mathsf{RndSch}^m_{\mathbb{Z}}$ introduced in Section 4.2 and discussed further in Section 4.8. The natural refinement operator sends a clopen interval to its midpoint. For example, $\langle \mathsf{round}_2, [350, 450)\rangle \in \mathsf{Granules}\langle \mathsf{RndSch}^m_{\mathbb{Z}} | \mathsf{round}_2\rangle$ is mapped to 400.

Formally, for a clopen interval $[\ell, h)$, define $\mathsf{MidPt}\langle[\ell, h)\rangle$ to be just $(\ell + h)/2$. Then, for each $G \in \mathsf{Glty}\langle \mathsf{RndSch}^m_{\mathbb{Z}}\rangle \setminus \{\mathsf{BaseGlty}\langle \mathsf{RndSch}^{\mathsf{m}}_{\mathbb{Z}}\rangle\}$, define $\mathsf{Refine}^{\mathsf{mid}}_{\mathsf{RndSch}^m_{\mathbb{Z}}}\langle G \rhd \mathsf{Gran}\mathbb{Z}\rangle$ on granule values to be just MidPt restricted to the applicable intervals. Define $\mathsf{Refine}^{\mathsf{mid}}_{\mathsf{RndSch}^m_{\mathbb{Z}}}\langle \mathsf{BaseGlty}\langle \mathsf{RndSch}^{\mathsf{m}}_{\mathbb{Z}}\rangle \rhd \mathsf{BaseGlty}\langle \mathsf{RndSch}^{\mathsf{m}}_{\mathbb{Z}}\rangle\rangle$ to be the identity on $\mathbb{Z}$. It is then easy to verify that $\{\mathsf{Refine}^{\mathsf{mid}}_{\mathsf{RndSch}^m_{\mathbb{Z}}}\langle G \rhd \mathsf{Gran}\mathbb{Z}\rangle \mid G \in \mathsf{Granules}\langle\mathfrak{S}\rangle\}$ is a refinement family for $\mathsf{RndSch}^m_{\mathbb{Z}}$. It is the most natural one, since it sends a rounded number, represented as an interval $[\ell, h)$, as the number $(\ell + h)/2$ which represents the rounding.

However, other choices are possible. Instead of the midpoint, the function LowPt defined on elements by $[\ell, h) \mapsto \ell$ could also be used. This is a form of rounding down. It is not as useful in practice because it generally results in larger overall errors than does rounding to midpoint.

4.12. *Aggregation Operators*

Data in a multigranular context are often statistical in nature. As such, thematic values corresponding to coarser spatial or temporal regions may be aggregations of those for finer ones. Therefore, a general formulation of an aggregation operator is central to any effort to model data integration in such a context.

Let $\mathbf{P} = (P, \leqslant_{\mathbf{P}})$ be any poset. An *aggregation operator* on $\mathbf{P}$ is a function $\bigoplus : \mathsf{MultisetsOf}\langle P\rangle \to P$ such that the following two properties hold.

Unary idempotence: For any $x \in P$, $\bigoplus\{x\} = x$.

Group associativity: For any finite multiset $S \subseteq P$ and any multi-partition $\wr S_i \mid i \in I \wr$ of S, $\bigoplus \wr \bigoplus(S_i) \mid i \in I \wr = \bigoplus S$.

These properties only identify that which is necessary for an operator of the form $\bigoplus : \mathsf{MultisetsOf}\langle P\rangle \to P$ to “make sense” as an aggregation operator. They do not characterize quality or identify desirable properties in any way. For a discussion of the latter, see Calvo *et al.* (2002) and Lenz and Thalheim (2009).

There are two additional properties which enhance an aggregation operator, but which are not required in all situations.

Monotonicity: For any finite multisets $S_1, S_2 \subseteq P$, if there is an injective multifunction $h : S_1 \to S_2$ such that $(\forall g \in S_1)(g \leqslant_P h(g))$, then $\bigoplus S_1 \leqslant_P \bigoplus S_2$.

Duplicate invariance: For any finite multiset $S \subseteq P$, $\bigoplus S = \bigoplus \mathsf{SetOf}\langle S \rangle$.
If $\oplus$ has the monotonicity (resp. duplicate invariance) property, then it is said to be *monotonic* (resp. *duplicate invariant*).

On the natural numbers $\mathbb{N}$, summation $\sum$ is an aggregation operator which is unary idempotent and group associative (as are all aggregation operators as defined here), as well as monotonic, but not duplicate invariant.

On the integers $\mathbb{Z}$, summation $\sum$ is still an aggregation operator, but it is neither monotonic nor duplicate invariant. An example of an attribute which would use such an operator would be $\mathsf{NetBirths}$, that is, $\mathsf{Births}–\mathsf{Deaths}$.

On both $\mathbb{N}$ and $\mathbb{Z}$, max is an aggregation operator which is both monotonic and duplicate invariant.

To obtain an example of an aggregation operator which is duplicate invariant but not monotonic, consider the real numbers $\mathbb{R}$ with the aggregation operator which takes a set $S \subseteq \mathbb{R}$ and returns the number which is closest to zero. If there is a tie between a negative number and a positive one, choose the positive one.

The operator min has the same properties as max, if the order used is $\geqslant$ instead of $\leqslant$.

Operations which do not respect group associativity, such as averaging, are not aggregation operators in the sense defined here.

4.13. *Aggregation Operators and Thematic Orderings*

An *aggregation operator* on a thematic attribute schema $\mathfrak{S}$ is just an aggregation operator, in the sense of Section 4.12, on the poset $(\mathsf{BaseGranules}\langle\mathfrak{S}\rangle, \leqslant_{\mathfrak{S}})$.

Data from different sources may be delivered with values for the thematic attributes provided in different granularities (e.g. with different degrees of rounding). Therefore, it is desirable to allow aggregation on multisets consisting not only of elements in $\mathsf{BaseGranules}\langle\mathfrak{S}\rangle$, but on all members of $\mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$, and furthermore on mixes of elements of differing granularities. If $\mathfrak{S}$ has a refinement family; in particular, if the thematic order $\leqslant_{\mathfrak{S}}$ is convex, and a refinement operator $\mathsf{Refine}_{\mathfrak{S}}\langle \cdot \rhd \cdot \rangle$ is provided, this is accomplished quite effortlessly. Specifically, for $S \subseteq \mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$ and $\bigoplus$ an aggregation operator on $\mathsf{BaseGranules}\langle\mathfrak{S}\rangle$, extend $\bigoplus$ to all of $\mathsf{Granules}_{\not\perp}\langle\mathfrak{S}\rangle$ via $\bigoplus S = \bigoplus\{\mathsf{Refine}_{\mathfrak{S}}\langle \mathsf{Glty}_{\mathfrak{S}}(g) \rhd \mathsf{BaseGlty}\langle\mathfrak{S}\rangle\rangle(g) \mid g \in S\}$.

The idea is best illustrated by example. In the context of $\mathsf{RndSch}^3_{\mathbb{N}}$, suppose data from different sources provide the following set of values $\{126, [3420, 3429), [2200, 2300)\}$. The first value is from round_0, the second from round_1, and the third from round_2. Suppose further that these numbers are aggregated using the summation operator $\sum$. The result is

$$\begin{aligned}\textstyle\sum\wr\ & \mathsf{Refine}^{\mathsf{mid}}_{\mathsf{RndSch}^3_{\mathbb{N}}}\langle \mathsf{round}_0 \rhd \mathsf{round}_0\rangle(1264),\\ & \mathsf{Refine}^{\mathsf{mid}}_{\mathsf{RndSch}^3_{\mathbb{N}}}\langle \mathsf{round}_1 \rhd \mathsf{round}_0\rangle(\langle \mathsf{round}_1, [3305, 3315)\rangle),\\ & \mathsf{Refine}^{\mathsf{mid}}_{\mathsf{RndSch}^3_{\mathbb{N}}}\langle \mathsf{round}_2 \rhd \mathsf{round}_0\rangle(\langle \mathsf{round}_2, [2250, 2350)\rangle)\int\\ & \qquad = \textstyle\sum \wr 1264, 3310, 2300 \int = 6874.\end{aligned}$$

If it is subsequently desired to provide a value rounded to the nearest 100; i.e. a value in $\mathsf{Granules}\langle \mathsf{RndSch}^3_{\mathbb{N}} | \mathsf{round}_2 \rangle$, this value may be obtained by coarsening the sum:

$$\mathsf{Coarsen}_{\mathsf{RndSch}^3_{\mathbb{N}}} \langle \mathsf{round}_0 \lhd \mathsf{round}_2 \rangle (6874) = \langle \mathsf{round}_2, [6850, 6950) \rangle,$$

which amounts to saying that 6874 is rounded to 6900. Note that

$$6900 = \mathsf{Refine}^{\mathsf{mid}}_{\mathsf{RndSch}^3_{\mathbb{N}}} \langle \mathsf{round}_0 \rhd \mathsf{round}_2 \rangle (\langle \mathsf{round}_2, [6850, 6950) \rangle.$$

In practice, one may say that the granule $\langle \mathsf{round}_2, [6850, 6950) \rangle$ "is" 6900, although technically speaking, it is a representation of numbers which round to 6900. The granule $\langle \mathsf{round}_2, [6850, 6950) \rangle$ is not the same as the granule $6900 \in \mathsf{Granules}\langle \mathsf{RndSch}^3_{\mathbb{N}} | \mathsf{round}_0 \rangle$, even though they have the same "value" of 6900.

5. Constraints for Data Integrity

5.1. *Attributewise Specification of Classical Order Dependencies*

The dependencies which are developed in this paper are presented in an *attributewise* fashion, in the sense that all attributes on the left-hand side (LHS), save for one, are held constant. To illustrate, the idea is first sketched within the context of order dependencies (Ginsburg and Hull, 1983; Ng, 2001; Szlichta *et al.*, 2012), which generalize functional dependencies (FDs) to a framework which includes order. The domain of each attribute A is endowed with a partial order $\leqslant_A$, with the *order dependency (OD)* $A_1 A_2 \dots A_k \overset{\leqslant}{\to} B$ holding iff for any two tuples t_1, t_2 with the property that whenever $t_1[A_i] \leqslant_{A_i} t_2[A_i]$ for $1 \leqslant i \leqslant k$, then $t_1[B] \leqslant_B t_2[B]$ as well. If, for each attribute A, $\leqslant_A$ is taken to be the trivial order in which $x \leqslant_A y$ iff $x = y$, then ordinary FDs are recovered.

Now, for $i \in [1, k]$, define $S_i = \{A_1, A_2, \dots, A_k\} \setminus \{A_i\}$, and call two tuples t_1, t_2 S_i-equivalent if $t_1[S_i] = t_2[S_i]$. Choosing S as above, the *attributewise dependency* defined by S_i on the OD $A_1 A_2 \dots A_k \overset{\leqslant}{\to} B$, denoted $\underline{A}_1 \dots \underline{A}_{i-1} A_i \underline{A}_{i+1} \dots \underline{A}_k \overset{\leqslant}{\to} B$ is defined to hold precisely in the case that for any two S_i-equivalent tuples t_1, t_2, if $t_1[A_i] \leqslant_{A_i} t_2[A_i]$, then $t_1[B] \leqslant_B t_2[B]$. In other words, all attributes save for A_i are held constant; only A_i is allowed to vary.

For the dependencies developed in this paper, such an attributewise representation is essential. This will be explained in more detail in 5.10, after these dependencies have been developed fully.

Before presenting the dependencies themselves, it is necessary to identify the relational framework which they constrain.

5.2. *Multigranular Relation Schemes*

Let $\mathfrak{U}$ be a set of granulated attributes. Assume that to each $A \in \mathfrak{U}$ is associated a constrained granulated attribute schema $(\mathbf{Glty}\langle \mathfrak{G}_{\mathbf{A}} \rangle, \mathsf{GrAsgn}\langle \mathfrak{G}_A \rangle, \mathsf{Constr}\langle \mathfrak{G}_A \rangle, \mathsf{cwa}\langle \mathfrak{G}_A \rangle)$, as described in Section 3.21.

Extending the classical definition (Maier, 1983; 1.2), for $k \in \mathbb{N}^+$, a (k-ary) *multigranular relation scheme* over $\mathfrak{U}$ is an expression of the form $R\langle \alpha \rangle$, where $\alpha =$

$\langle A_1, A_2, \ldots, A_k \rangle \in \mathfrak{U}^k$. The symbol R is called the *relation name*, and the list α is called an *attribute vector*.

A *data tuple* for the attribute vector $\alpha = \langle A_1, A_2, \ldots, A_k \rangle$ is a k-tuple $t \in \mathsf{Granules}\langle \mathfrak{G}_{A_1} \rangle \times \mathsf{Granules}\langle \mathfrak{G}_{A_2} \rangle \times \ldots \times \mathsf{Granules}\langle \mathfrak{G}_{A_k} \rangle$. The set of all data tuples for α is denoted $\mathsf{Tuples}\langle \alpha \rangle$. A *database* for the schema $R\langle \alpha \rangle$ is a set $M \subseteq \mathsf{Tuples}\langle \alpha \rangle$. The set of all databases for $R\langle \alpha \rangle$ is denoted $\mathsf{DB}(R\langle \alpha \rangle)$.

5.3. *The Context*

Throughout this section, unless stated specifically to the contrary, take $\mathfrak{S} = (\mathbf{Glty}\langle \mathfrak{S} \rangle, \mathsf{GrAsgn}\langle \mathfrak{S} \rangle, \mathsf{Constr}\langle \mathfrak{S} \rangle, \mathsf{cwa}\langle \mathfrak{S} \rangle)$ to be a constrained granulated attribute schema (see Section 3.21).

Likewise, take $\mathfrak{U}$ to be a finite set of granulated attributes, with a constrained granulated attribute schema $(\mathbf{Glty}\langle \mathfrak{G}_{\mathbf{A}} \rangle, \mathsf{GrAsgn}\langle \mathfrak{G}_A \rangle, \mathsf{Constr}\langle \mathfrak{G}_A \rangle, \mathsf{cwa}\langle \mathfrak{G}_A \rangle)$ associated with each $A \in \mathfrak{U}$.

The names in $\mathfrak{U}$ include at least those of the form A_i and B, with the association of a schema to the name as described in Section 5.2. It will further be assumed that B is a thematic attribute, with thematic order $\leqslant_{\mathfrak{G}_B}$, as described in Section 4.5, and that $\oplus$ is an aggregation operator for $\mathfrak{G}_B$ as defined in Section 4.13.

5.4. *Set Coarsening*

Define the function $\mathsf{CoarsenSetMUB}_{\mathfrak{S}} : \mathbf{2}^{\mathsf{Granules}_{\not\perp}\langle \mathfrak{S} \rangle} \to \mathbf{2}^{\mathsf{Glty}\langle \mathfrak{S} \rangle}$ to be that which maps $S \subseteq \mathsf{Granules}_{\not\perp}\langle \mathfrak{S} \rangle$ to the minimal elements (under $\leqslant_{\mathsf{Glty}\langle S \rangle}$) in the set

$$\{G \in \mathsf{Glty}\langle \mathfrak{S} \rangle \mid (\forall g \in S)(\mathsf{Coarsen}_{\mathfrak{S}}\langle \mathsf{Glty}_{\mathfrak{S}}(g) \lhd G \rangle(g)) \text{ is defined}\}.$$

In words, it returns the minimal granularities to which all elements of S coarsen. Since $\mathsf{Glty}\langle \mathfrak{S} \rangle$ has a greatest element $\top_{\mathsf{Glty}\langle \mathfrak{G}_{\mathbf{A}} \rangle}$, this set of minimal granularities can never be empty.

This operation will be applied only to granules of thematic attributes. In all examples of this paper (in Section 4), the associated granularity poset has LUBs. However, the formalism allows for the case in which there are several MUBs (minimal upper bounds) of given set $S \subseteq \mathsf{Granules}\langle \mathfrak{S} \rangle$.

5.5. *The TMCD*

The dependencies developed in this section are called *thematic multigranular comparison dependencies*, or *TMCDs*. Each one is specified in attributewise fashion, with the value of only one attribute on the LHS allowed to vary. The general notation is $\underline{A_1}\underline{A_2} \ldots \underline{A_{i-1}} A_i \underline{A_{i+1}} \ldots \underline{A_k} \overset{[}{]}{}_{\langle \beta : \eta \rangle}\!\!\longrightarrow \langle B, \oplus \rangle$, in which the A_i's are (ordinary) multigranular attributes, B is a thematic attribute, and $\oplus$ is an aggregation operator on B. The dependencies are classified according to the two parameters $\beta \in \{\bigsqcup\!\!\!\!\perp, \bigsqcup\}$ (the *join type*) and $\eta \in \{\sqsubseteq, =, \sqsupseteq\}$ (the *order type*). Thus, their are six fundamental variants. There is also a third possibility, $\beta = 1$, but it is a special case of $\beta = \bigsqcup\!\!\!\!\perp$, and not a fundamentally different case. It will be discussed later.

In a TCMD, it is not in general a single value for A_i (as in the case of FDs and ODs), but rather a set of values which match some join rule, which determine the value of B of a second tuple. The formula which defines the semantics of this TMCD is shown below, for a multigranular relation scheme $R\langle\alpha\rangle$ whose attribute vector α contains at least $\{A_1, A_2, \ldots, A_k, B\}$.

$$
\begin{aligned}
&(\forall t_1 \in \mathsf{Tuples}\langle\alpha\rangle)(\forall T_2 \subseteq_f \mathsf{Tuples}\langle\alpha\rangle)\\
&\qquad(\forall G \in \mathsf{CoarsenSetMUB}_{\mathfrak{G}_B}\langle\{t_1.B\} \cup \{t.B \mid t \in T_2\}\rangle)\\
&\Bigg(\Bigg(R\langle t_1\rangle \wedge \Big(\bigwedge_{t_2\in T_2} R\langle t_2\rangle\Big) \wedge \Big(\bigwedge_{\substack{t_2 \in T_2\\ j \in [1,k]\setminus\{i\}}} (t_1.A_j = t_2.A_j)\Big)\\
&\qquad \wedge \Big(t_1.A_i \begin{bmatrix}\sqsubseteq_{\mathfrak{G}_{A_i}}\\ =\\ \sqsupseteq_{\mathfrak{G}_{A_i}}\end{bmatrix} \bigsqcup^{?}_{t_2\in T_2}{}_{\mathfrak{G}_{A_i}}\, t_2.A_i\Big)\Bigg)\\
&\Rightarrow \Bigg(\mathsf{Coarsen}_{\mathfrak{G}_B}\langle \mathsf{Glty}_{\mathfrak{G}_B}(t_1.B) \lhd G\rangle(t_1.B) \begin{bmatrix}\leqslant_{\mathfrak{G}_B}\\ =\\ \geqslant_{\mathfrak{G}_B}\end{bmatrix}\\
&\qquad \mathsf{Coarsen}_{\mathfrak{G}_B}\langle \mathsf{BaseGlty}\langle\mathfrak{G}_\mathsf{B}\rangle \lhd G\rangle\\
&\qquad\quad \Big(\bigoplus_{t_2\in T_2} \mathsf{Refine}_{\mathfrak{G}_B}\langle \mathsf{Glty}_{\mathfrak{G}_B}(t_2.B) \rhd \mathsf{BaseGlty}\langle\mathfrak{G}_\mathsf{B}\rangle\rangle(t_1.B)\Big)\Bigg)\Bigg).
\end{aligned}
$$

The two parameters are incorporated in this representation. First of all, $\bigsqcup^{?}$ represents the choice of β, and thus may be either $\bigsqcup^{\perp}$ or $\bigsqcup$. Second, the two stacks of symbols, $\begin{bmatrix}\sqsubseteq_{\mathfrak{G}_{A_i}}\\ =\\ \sqsupseteq_{\mathfrak{G}_i A}\end{bmatrix}$ and $\begin{bmatrix}\leqslant_{\mathfrak{G}_B}\\ =\\ \geqslant_{\mathfrak{G}_B}\end{bmatrix}$, represent the choices for η. The same row must be chosen for each; for example, if $\sqsubseteq_{\mathfrak{G}_{A_i}}$ is chosen in the first stack, then $\leqslant_{\mathfrak{G}_B}$ must be chosen from the second.

For a particular choice of t_1 and T_2, the *match rule* is $t_1.A_i \circledast \bigsqcup^{?}_{t_2\in T_2}{}_{\mathfrak{G}_{A_i}} t_2.A_i$, with $\circledast$ the appropriate choice in $\{\sqsubseteq_{\mathfrak{G}_{A_i}}, =, \sqsupseteq_{\mathfrak{G}_{A_i}}\}$. This match rule must be in $\mathsf{Constr}\langle\mathfrak{G}_{A_i}\rangle^+$ for the TMCD to apply. For example, for a TCMD of type $\langle \bigsqcup^{\perp} : =\rangle$, the match rule for t_1 and T_2 is $t_1.A_i = \bigsqcup^{\perp}_{t_2\in T_2} t_2.A_i$. If it is in $\mathsf{Constr}\langle\mathfrak{G}_{A_i}\rangle^+$, and if the values for the other A_j attributes are constant; i.e. if $t_1.A_j = t_2.A_j = t'_2.A_j$ for all $t_2, t'_2 \in T_2$ and all $j \in [1,k]\setminus\{i\}$, then the aggregation $t_1.b = \bigoplus_{t_2\in T_2} t_2.B$ on B must hold, after suitable coarsening is applied. Table 1 summarizes these conditions.

Table 1
Properties of TMCDs; see Sections 5.7 and 5.9 for clarification of remarks.

Type	Match rule	Aggregation (w/o coarsening)	Remarks
$\langle \biguplus : = \rangle$	$t_1.A_i = \biguplus_{t_2 \in T_2} t_2.A_i$	$t_1.B = \bigoplus_{t_2 \in T_2} t_1.B$	
$\langle \biguplus : \sqsubseteq \rangle$	$t_1.A_i \sqsubseteq_{\mathfrak{G}_{A_i}} \biguplus_{t_2 \in T_2} t_2.A_i$	$t_1.B \leqslant_{\mathfrak{G}_B} \bigoplus_{t_2 \in T_2} t_2.B$	Typically monotonic
$\langle \biguplus : \sqsupseteq \rangle$	$t_1.A_i \sqsupseteq_{\mathfrak{G}_{A_i}} \biguplus_{t_2 \in T_2} t_2.A_i$	$t_1.B \geqslant_{\mathfrak{G}_B} \bigoplus_{t_2 \in T_2} t_2.B$	Typically monotonic
$\langle \bigsqcup : = \rangle$	$t_1.A_i = \bigsqcup_{t_2 \in T_2} t_2.A_i$	$t_1.B = \bigoplus_{t_2 \in T_2} t_1.B$	Typically duplicate invariant
$\langle \bigsqcup : \sqsubseteq \rangle$	$t_1.A_i \sqsubseteq_{\mathfrak{G}_{A_i}} \bigsqcup_{t_2 \in T_2} t_2.A_i$	$t_1.B \leqslant_{\mathfrak{G}_B} \bigoplus_{t_2 \in T_2} t_2.B$	Typically monotonic + duplicate invariant
$\langle \bigsqcup : \sqsupseteq \rangle$	$t_1.A_i \sqsupseteq_{\mathfrak{G}_{A_i}} \bigsqcup_{t_2 \in T_2} t_2.A_i$	$t_1.B \geqslant_{\mathfrak{G}_B} \bigoplus_{t_2 \in T_2} t_2.B$	Typically monotonic + duplicate invariant
$\langle 1 : = \rangle$	$t_1.A_i = t_2.A_i$	$t_1.B = t_2.B$	Special case of $\langle \biguplus : = \rangle$
$\langle 1 : \sqsubseteq \rangle$	$t_1.A_i \sqsubseteq_{\mathfrak{G}_{A_i}} t_2.A_i$	$t_1.B \leqslant_{\mathfrak{G}_B} t_2.B$	Special case of $\langle \biguplus : \sqsubseteq \rangle$
$\langle 1 : \sqsupseteq \rangle$	$t_1.A_i \sqsupseteq_{\mathfrak{G}_{A_i}} t_2.A_i$	$t_1.B \geqslant_{\mathfrak{G}_B} t_2.B$	Special case of $\langle \biguplus : \sqsupseteq \rangle$

5.6. *Example*

First consider the schema $R_{\mathsf{sumb}}\langle A_{\mathsf{Plc}}, A_{\mathsf{Tim}}, B_{\mathsf{Bth}}\rangle$, governed by the TMCD $A_{\mathsf{Plc}}\underline{A_{\mathsf{Tim}}}\,{}^{[}_{]}\xrightarrow[\langle\bigsqcup:=\rangle]{}\langle B_{\mathsf{Bth}}, \sum\rangle$, with $\sum$ denoting the summation operator. The aggregation is over places (A_{Plc}), for fixed periods of time (A_{Tim}). Think of using (r-Chile) from Section 1 as the match rule, for the fixed time interval *Q1Y2014*, with the tuples in T_1 coming from Part 1 of Fig. 1, and t_2 coming from Part 2. The rule captures formally that which was described informally in Section 1, that the number of births during *Q1Y2014* in all of Chile is the sum of those in the fifteen regions, after accounting for rounding.

The granularity G of the formula of Section 5.5 is chosen to make sure that all results are compared at the same granularity. The formalism does not require that G be unique, but in most practical examples, such as the rounding attribute schema $\mathsf{RndSch}^m_{\mathbb{N}}$ presented in Section 4.2, it will be. Returning to the concrete example, suppose that the birth data in Part 1 of Fig. 1 are rounded to the nearest 20, while the data in Part 2 are rounded to the nearest 50. It would be unrealistic to expect a sum of numbers, each rounded to the nearest 20, to agree with a summary which has been rounded to the nearest 50. To make the comparison more realistic, each is rounded to the finest rounding which is coarser than both; in this case, 100. (This assumes that rounding to 20, 50, and 100 are all supported by the thematic schema $\mathfrak{G}_{B_{\mathsf{Bth}}}$.)

Now consider the schema $R_{\mathsf{maxp}}\langle A_{\mathsf{Plc}}, A_{\mathsf{Tim}}, B_{\mathsf{Pop}}\rangle$, with A_{Plc} and A_{Tim} as in $R_{\mathsf{sumb}}\langle A_{\mathsf{Plc}}, A_{\mathsf{Tim}}, B_{\mathsf{Bth}}\rangle$, but B_{Pop} an attribute which records the maximum population of a geographic region during a given period of time. The applicable TMCD $A_{\mathsf{Plc}}\underline{A_{\mathsf{Tim}}}\,{}^{[}_{]}\xrightarrow[\langle\biguplus:=\rangle]{}\langle B_{\mathsf{Pop}}, \mathsf{max}\rangle$ in which max is the aggregation operator on $\mathbb{N}$ which selects the maximum value. Here the fact that the join may not be disjoint does not matter, since the aggregation operator max is duplicate invariant. For a further discussion of this, see Section 5.9 below.

5.7. *TMCDs Without Aggregation*

The choice of $\beta = 1$ represents the case of $\beta = \lfloor \perp \rfloor$ in which only one element is aggregated. Since it is a very fundamental special case which occurs frequently in practice, it deserves special attention. The resulting formula is shown below for a multigranular relation scheme $R\langle\alpha\rangle$ whose attribute vector α contains at least $\{A_1, A_2, \ldots, A_k, B\}$.

$$(\forall t_1, t_2 \in \mathsf{Tuples}\langle\alpha\rangle)(\forall G \in \mathsf{CoarsenSetMUB}_{\mathfrak{G}_B}\langle\{t_1.B, t_2.B\}\rangle)$$

$$\left(R\langle t_1\rangle \wedge R\langle t_2\rangle \wedge \left(t_1.A_i \begin{bmatrix} \sqsubseteq_{\mathfrak{G}_{A_i}} \\ = \\ \sqsupseteq_{\mathfrak{G}_{A_i}} \end{bmatrix} t_2.A_i\right)\right.$$

$$\Rightarrow \left(\mathsf{Coarsen}_{\mathfrak{G}_B}\langle \mathsf{BaseGlty}\langle\mathfrak{G}_\mathsf{B}\rangle \lhd G\rangle(t_1.B)\right.$$

$$\left.\left. \begin{bmatrix} \leqslant_{\mathfrak{G}_B} \\ = \\ \geqslant_{\mathfrak{G}_B} \end{bmatrix} \mathsf{Coarsen}_{\mathfrak{G}_B}\langle \mathsf{Glty}_{\mathfrak{G}_B}(t_2.B) \lhd G\rangle(t_2.B)\right)\right).$$

For the case of equality; e.g. $A_{\mathsf{Plc}}\underline{A_{\mathsf{Tim}}} \overset{[}{]}_{\langle 1:=\rangle}\!\longrightarrow B_{\mathsf{Bth}}$, it is nothing more than the attributewise FD $A_{\mathsf{Plc}}\underline{A_{\mathsf{Tim}}} \rightarrow B_{\mathsf{Bth}}$, modulo coarsening. If all values for B_{Bth} are of the same granularity, it is exactly the attributewise FD. For the case of subsumption; e.g. $A_{\mathsf{Plc}}\underline{A_{\mathsf{Tim}}} \overset{[}{]}_{\langle 1:\sqsubseteq\rangle}\!\longrightarrow B_{\mathsf{Bth}}$, it is the order dependency $A_{\mathsf{Plc}}\underline{A_{\mathsf{Tim}}} \overset{\leqslant}{\rightarrow} B_{\mathsf{Bth}}$, again modulo coarsening. For $A_{\mathsf{Plc}}\underline{A_{\mathsf{Tim}}} \overset{[}{]}_{\langle 1:\sqsupseteq\rangle}\!\longrightarrow B_{\mathsf{Bth}}$, it is, modulo coarsening, $A_{\mathsf{Plc}}\underline{A_{\mathsf{Tim}}} \overset{\geqslant}{\rightarrow} B_{\mathsf{Bth}}$. Thus, the ordinary FDs and ODs which underlie the more complex TMCDs are representable in this framework.

Although inference of TMCDs is not a focus of this paper, it is nevertheless worthwhile to point out a few of the simplest ones.

5.8. *Implications of TMCDs*

Under the notation of 5.3, the following implications hold, with $\models$ denoting semantic entailment on TMCDS.

$$\underline{A_1}\,\underline{A_2}\ldots\underline{A_{i-1}}A_i\underline{A_{i+1}}\ldots\underline{A_k} \overset{[}{]}_{\langle \sqcup : \eta\rangle}\!\longrightarrow \langle B, \oplus\rangle$$

$$\models \underline{A_1}\,\underline{A_2}\ldots\underline{A_{i-1}}A_i\underline{A_{i+1}}\ldots\underline{A_k} \overset{[}{]}_{\langle \uplus : \eta\rangle}\!\longrightarrow \langle B, \oplus\rangle$$

$$\models \underline{A_1}\,\underline{A_2}\ldots\underline{A_{i-1}}A_i\underline{A_{i+1}}\ldots\underline{A_k} \overset{[}{]}_{\langle 1 : \eta\rangle}\!\longrightarrow \langle B, \oplus\rangle,$$

$$\{\underline{A_1}\,\underline{A_2}\ldots\underline{A_{i-1}}A_i\underline{A_{i+1}}\ldots\underline{A_k} \overset{[}{]}_{\langle \lfloor ? \rfloor : \sqsubseteq\rangle}\!\longrightarrow \langle B, \oplus\rangle,$$

$$\underline{A}_1\underline{A}_2 \ldots \underline{A}_{i-1} A_i \underline{A}_{i+1} \ldots \underline{A}_k \overset{[}{]}_{\langle \bigsqcup^{?} : \sqsupseteq\rangle}\!\!\longrightarrow \langle B, \oplus\rangle\}$$
$$\models \underline{A}_1\underline{A}_2 \ldots \underline{A}_{i-1} A_i \underline{A}_{i+1} \ldots \underline{A}_k \overset{[}{]}_{\langle \bigsqcup^{?} : =\rangle}\!\!\longrightarrow \langle B, \oplus\rangle,$$

$$\underline{A}_1\underline{A}_2 \ldots \underline{A}_{i-1} A_i \underline{A}_{i+1} \ldots \underline{A}_k \overset{[}{]}_{\langle 1 : \sqsubseteq\rangle}\!\!\longrightarrow \langle B, \oplus\rangle$$
$$\models \underline{A}_1\underline{A}_2 \ldots \underline{A}_{i-1} A_i \underline{A}_{i+1} \ldots \underline{A}_k \overset{[}{]}_{\langle 1 : =\rangle}\!\!\longrightarrow \langle B, \oplus\rangle.$$

Proof. All are very basic verifications. □

5.9. *TMCDs and Properties of the Aggregation Operator*

Given a TMCD $\underline{A}_1\underline{A}_2 \ldots \underline{A}_{i-1} A_i \underline{A}_{i+1} \ldots \underline{A}_k \overset{[}{]}_{\langle \beta : \eta\rangle}\!\!\longrightarrow \langle B, \oplus\rangle$, consider the following requirements.

(cond-mn) For $\eta \in \{\sqsubseteq, \sqsupseteq\}$, $\oplus$ must be monotonic.

(cond-di) For $\beta = \bigsqcup$, $\oplus$ must be duplicate invariant.

In a practical sense, these statements are always true. However, it is possible to construct pathological examples in which they fail. A full treatment of this subject is beyond the scope and space limitations of this paper. Therefore, the above statements are to be taken only as real-world design guidelines, not mathematically established results.

This is not a limitation of the framework, since in practice it is clear which aggregation operators apply in a given framework.

5.10. *Discarding Attributewise Specification*

In the case that the same thematic order and aggregation operator is used with respect to all attributes on the LHS of a TMCD, it is tempting to consider discarding the attributewise specification, and combine all into one big dependency, which might be represented as $A_1 A_2 \ldots A_k \overset{[}{]}_{\langle \beta : \eta\rangle}\!\!\longrightarrow B$, for a multigranular relation scheme $R\langle\alpha\rangle$ whose attribute vector α contains at least $\{A_1, A_2, \ldots, A_k, B\}$.

$$(\forall t_1 \in \mathsf{Tuples}\langle\alpha\rangle)(\forall T_2 \subseteq_f \mathsf{Tuples}\langle\alpha\rangle)$$
$$(\forall G \in \mathsf{CoarsenSetMUB}_{\mathfrak{G}_B}\langle\{t_1.B\} \cup \{t.B \mid t \in T_2\}\rangle)$$
$$\left(\left(R\langle t_1\rangle \wedge \left(\bigwedge_{t_2 \in T_2} R\langle t_2\rangle\right) \wedge \left(\bigwedge_{i\in[1,k]}\left(t_1.A_i \begin{bmatrix} \sqsubseteq_{\mathfrak{S}_{A_i}} \\ = \\ \sqsupseteq_{\mathfrak{S}_{A_i}} \end{bmatrix} \bigsqcup^{?}_{t_2\in T_2}{}_{\mathfrak{S}_{A_i}} t_2.A_i\right)\right)\right)\right)$$
$$\Rightarrow \left(\mathsf{Coarsen}_{\mathfrak{G}_B}\langle \mathsf{Glty}_{\mathfrak{G}_B}(t_2.B) \lhd G\rangle(t_2.B) \begin{bmatrix} \leqslant_{\mathfrak{G}_B} \\ = \\ \geqslant_{\mathfrak{G}_B} \end{bmatrix}\right.$$
$$\mathsf{Coarsen}_{\mathfrak{G}_B}\langle \mathsf{BaseGlty}\langle\mathfrak{G}_\mathsf{B}\rangle \lhd G\rangle$$
$$\left.\left(\bigoplus_{t_1\in T_1} \mathsf{Refine}_{\mathfrak{G}_B}\langle \mathsf{Glty}_{\mathfrak{G}_B}(t_2.B) \rhd \mathsf{BaseGlty}\langle\mathfrak{G}_\mathsf{B}\rangle\rangle(t_1.B)\right)\right)\Bigg)\Bigg).$$

From a theoretical point of view, this definition is fine. However, without suitable adaptation, it does not recapture what would normally be expected of such a dependency. To illustrate, work within the context of $R_{\mathsf{sumb}}\langle A_{\mathsf{Plc}}, A_{\mathsf{Tim}}, B_{\mathsf{Bth}}\rangle$, with the rules $p = \bigsqcup_{\mathfrak{S}_{A_{\mathsf{Plc}}}}\{p_1, p_2\} = \bigsqcup_{\mathfrak{S}_{A_{\mathsf{Plc}}}}\{p_3, p_4\}$ holding in $\mathsf{Constr}\langle\, \mathfrak{S}_{A_{\mathsf{Plc}}}\rangle$ and the rules $t = \bigsqcup_{A_2}\{s_1, s_2\} = \bigsqcup_{A_2}\{s_3, s_4\}$ holding in $\mathsf{Constr}\langle \mathfrak{S}_{A_{\mathsf{Tim}}}\rangle$. Now, suppose that $T_1 = \{\langle p_1, s_1, b_1\rangle, \langle p_2, s_2, b_2\rangle\}$, and $t_2 = \langle p, s, b\rangle$ in the above formula. Assume further that all values for attribute B_{Bth} are at the same granularity G, so no coarsening is necessary. Then the above rule mandates that $b_1 + b_2 = b$. However, this is not realistic modelling. b_1 is the number of births in region p_1 during time s_1, while b_2 is the number of births in region p_2 during time interval s_2. To get the total number of births in region p during time interval t, it would be necessary to find and add tuples of the form $\langle p_1, s_2, b_3\rangle$ and $\langle p_2, s_1, b_4\rangle$. Then, and only then, would $b_1 + b_2 + b_3 + b_4 = b$ hold. In other words, there must be a tuple which captures every (place, time) point of an appropriate "rectangle" in order to get the correct total number of births.

Unfortunately, things can become even more complex. Suppose instead that $t_1 = \langle p, s, b\rangle$ and $T_2 = \{\langle p_1, s_1, b_1\rangle, \langle p_2, s_1, b_2\rangle, \langle p_3, s_2, b_3\rangle, \langle p_4, s_2, b_4\rangle\}$. It is easy to see that $b_1 + b_2 + b_3 + b_4 = b$ must hold here as well. In other words, different decompositions of p may be used for different corresponding values of attribute A_{Tim}. From a formal point of view, the most elegant solution is to regard $A_1A_2 \ldots A_k$ as a combined domain, and replace $(t_1.A_i \;=\; \bigwedge_{i=1}^{k} \bigsqcup_{\substack{\mathfrak{S}_{A_i}\\ t_2 \in T_2}} t_2.A_i)$ with something of the form $(t_1.A_1A_2 \ldots A_k = \bigsqcup_{\substack{\mathfrak{S}_{A_i}\\ t_2 \in T_2}} t_2.A_1A_2 \ldots A_k)$.

However, it seems that to implement something so complex efficiently would be almost impossible. Thus, it seems that attributewise specification is a necessity.

5.11. *Comparison to CFDs*

In contrast to the CFDs (conditional functional dependencies) of Bravo and Rodríguez (2014), the TMCDs developed here are specifically oriented towards data integration. CFDs are designed to recapture dependencies which hold only for certain granularities, with no support for aggregation or tolerance. TMCDs, on the other hand, are designed to support these latter two concepts. The overlap of CFDs and TCMDs is therefore minimal; they address complementary issues in the context of constraints for multigranular schemata.

5.12. *Tolerant Agreement*

If the data of different granularities come from different sources, there may be discrepancies which cannot be accounted for entirely from rounding. Rather than having a TMCD reject such data as failing to satisfy the basic integration constraint, a more tolerant approach might be employed. In Hegner and Rodríguez (2016; 2.13), a *coarsening tolerance*, in the spirit of a *tolerance relation* (Zeeman, 1962; Arbib, 1967; Peters and Wasilewski, 2012) is part of every TMCD. There may, however, be other forms of tolerant agreement which may be more appropriate in a given situation, particularly ones which provide a degree, rather than an absolute, measure. Examples include approaches using statistical or fuzzy tools. Consequently, in this work, no fixed notion of tolerant agreement is formulated. Rather, the intent is that the TMCD framework may be expanded to include such features, as necessary.

6. Relationship to Other Work

The notion of granularity appears in different contexts of data management. Granularity defines the units that quantitatively measure data with respect to the dimensions of the domain they represent. In Bettini *et al.* (1997), temporal granularity is formalized as a mapping function from a domain of indexes to the time domain. Each portion of the time domain corresponding to the mapping is referred as a temporal granule, which cannot overlap with any other granule of the same granularity. In a similar way, Wang and Liu (2004) define spatial granularity as a mapping function from a domain of indexes to portions of a space, called spatial granules. Later, the works in Camossi *et al.* (2006) defined a spatio-temporal granule as a tuple (s, t), meaning that at the time index t, the spatial index s is valid. In a similar way, the work in Belussi *et al.* (2009) assigns to each spatio-temporal granule a sequence of spatial granules, one for each granule in the temporal granularities. Based on the concept of spatial and temporal granules and granularities, the work in Camossi *et al.* (2006), Bertino *et al.* (2005) proposes a multigranular object-oriented framework that supports conversion operators between granules related by inclusion and provides a language where users can specify a particular conversion for moving from one to another granularity.

Data warehouses (DWs) and OLAP cubes are multigranular systems where conversion operators are fixed along a dimension. DWs support large datasets in query processing because they store pre-computed sub-aggregate measures associated with granules. OLAP cubes allow navigation through different levels of aggregate information of a data warehouse. In this context, Lenz and Thalheim (2005) provide a formal and functional definition of data cube that specifies a hierarchically ordered dimension that forms a lattice. In their work, they assume well-defined hierarchic dimensions, which impose that granules, what they call groupings, form a partition, that is, they are pairwise disjoint and form a cover.

Classical data warehouses assume data stored at the finest level of detail and where each value (granule) at this level can be mapped onto a value at a coarser level of a dimension. However, recent works highlight the need of storing data at different granularities (Iftikhar, 2012; Iftikhar and Pedersen, 2010) and handling complex data objects (Boukraâ *et al.*, 2010). A multigranular model was introduced in Bravo and Rodríguez (2014) and then refined in Hegner and Rodríguez (2016). The model in Hegner and Rodríguez (2016) is a formalization in terms of a partial set structure enhanced with rules to express conditions on the underline domain. In particular, it defines join rules that may or may not represent partitions of the space. In this work, integrity constraints define valid states over aggregation operators. These constraints relate, but differ from previous types of constraints applied in the context of data cubes. In particular, in Wijsen and Ng (1999), *roll-up dependencies* assert that certain thematic values (such as tax rates) must be invariant under roll-up. However, these constraints do not address thematic values which vary with granularity, or which involve aggregation.

7. Conclusions and Further Directions

A comprehensive model of multigranular attributes, as well as their use in defining multigranular relational schemata, has been developed. This includes not only basic attributes (typically spatio-temporal), but also thematic attributes and how they integrate with aggregation operators. In contrast to previous work, this approach recaptures not only the order structure of granules at distinct granularities, but also lattice-like operations on them. This supports the definition of integrity constraints, called TMCDs, which can express constraints which require the thematic value of one tuple be the same as the aggregation of the thematic values of several other tuples.

There are several avenues for further study.

EFFECTIVE IMPLEMENTATION: The ideas developed in this paper will only prove useful if they can be implemented effectively. An immediate task is to develop a prototype implementation for granular attributes, and relations, and then to apply them to testing satisfaction of TMCDs.

ALGORITHMS FOR TESTING RULE SATISFIABILITY: Although it has been argued in Section 3.14 that rules which arise from natural modelling situations will be satisfiable, it would nevertheless be advantageous to have effective algorithms for testing a set of rules for consistency. Work is currently underway to develop and implement a satisfaction algorithm for the basic rules of Section 3.10.

QUERY LANGUAGE: The work here proposes only constraints. An accompanying query language which takes into account the special needs of the multigranular framework must also be developed.

Acknowledgement. The work of M. Andrea Rodríguez, as well as two visits of Stephen J. Hegner to Concepción, during which many of the ideas reported here were developed, were partly funded by Fondecyt-Conicyt grant number 1140428. M. Andrea Rodríguez is also funded by Complex Engineering Systems Institute (CONICYT: FBO 16) and Fondecyt 1170497.

References

Arbib, M.A. (1967). Tolerance automata. *Kybernetika*, 3(3), 223–233.

Armstrong, W.W. (1974). Dependency structures of data base relationships. In: Rosenfeld, J.L. (Ed.), *Information Processing 74, Proceedings of IFIP Congress 74*. North Holland, Amsterdam, pp. 580–583.

Belussi, A., Combi, C., Pozzani, G. (2009). Formal and conceptual modeling of spatio-temporal granularities. In: Desai, B.C., Saccà, D., Greco, S. (Eds.), *International Database Engineering and Applications Symposium (IDEAS, 2009), September 16–18, 2009, Cetraro, Calabria, Italy*. *ACM International Conference Proceeding Series*. ACM, pp. 275–283.

Bertino, E., Cuadra, D., Martínez, P. (2005). An object-relational approach to the representation of multi-granular spatio-temporal data. In: Pastor, O., Falcão e Cunha, J. (Eds.), *Advanced Information Systems Engineering, 17th International Conference, CAiSE 2005, Porto, Portugal, June 13–17, 2005, Proceedings*, *Lecture Notes in Computer Science*, Vol. 3520. Springer, pp. 119–134.

Bettini, C., Dyreson, C.E., Evans, W.S., Snodgrass, R.T., Wang, X.S. (1997). *A glossary of time granularity concepts*. In: *Temporal Databases, Dagstuhl*, pp. 406–413.

Bonham-Carter, G.F. (1995). *Geographic Information Systems for Geoscientists: Modelling with GIS*. Pergamon, Oxford.

Boukraâ, D., Boussaïd, O., Bentayeb, F. (2010). OLAP operators for complex object data cubes. In: Catania, B., Ivanović, M., Thalheim, B. (Eds.), *Advances in Databases and Information Systems, 14th East European Conference, ADBIS 2010, Novi Sad, Serbia, September 20–24, 2010, Proceedings*, *Lecture Notes in Computer Science*, Vol. 6295, pp. 103–116.

Bravo, L., Rodríguez, M.A. (2014). A multi-granular database model. In: Beierle, C., Meghini, C. (Eds.), *Foundations of Information and Knowledge Systems – 8th International Symposium, FoIKS 2014, Bordeaux, France, March 3–7, 2014. Proceedings*, *Lecture Notes in Computer Science*, Vol. 8367. Springer, pp. 344–360.

Calvo, T., Kolesárová, A., Komorníková, M., Mesiar, R. (2002). Aggregation operators: properties, classes and construction methods. In: Calvo, T., Mayor, G., Mesiar, R. (Eds.), *Aggregation operators: New Trends and Applications*, Studies in Fuzziness and Soft Computing, Vol. 97. Springer, pp. 3–104.

Camossi, E., Bertolotto, M., Bertino, E. (2006). A multigranular object-oriented framework supporting spatio-temporal granularity conversions. *International Journal of Geographical Information Science*, 20(5), 511–534.

Clark, K.L. (1978). Negation as failure. In: Gallaire, H., Minker, J. (Eds.), *Logic and Data Bases*. Plenum Press, New York, pp. 293–322.

Cosmadakis, S.S., Kanellakis, P.C., Spyratos, N. (1986). Partition semantics for relations. *Journal of Computer and System Sciences*, 33(2), 203–233.

Davey, B.A., Priestly, H.A. (2002). *Introduction to Lattices and Order*, 2nd ed. Cambridge University Press, Cambridge.

Egenhofer, M.J., Franzosa, R.D. (1991). Point set topological relations. *International Journal of Geographical Information Systems*, 5(2), 161–174.

Fagin, R. (1982a). *Armstrong Databases*. Research Report RJ3440, IBM Research Laboratory, San Jose, California, also appeared in 7th IBM Symposium on Mathematical Foundations of Computer Science, Kanagawa, Japan, May 1982.

Fagin, R. (1982b). Horn clauses and database dependencies. *Journal of the ACM*, 29(4), 952–985.

Ginsburg, S., Hull, R. (1983). Order dependency in the relational model. *Theoretical Computer Science*, 26, 149–195.

Girish, K.P., John, S.J. (2009). Relations and functions in multiset context. *Information Sciences*, 179(6), 758–768.

Hegner, S.J. (1994). Distributivity in incompletely specified type hierarchies: theory and computational complexity. In: Dörre, J. (Ed.), *Computational Aspects of Constraint-Based Linguistic Description II, DYANA-2, ESPRIT Basic Research Project 6852, Deliverable R1.2B*. DYANA, pp. 29–120.

Hegner, S.J., Rodríguez, M.A. (2016). Integration integrity for multigranular data. In: Pokorny, J., Ivanovic, M., Thalheim, B., Saloun, P. (Eds.), *Advances in Databases and Information Systems, 20th East European Conference, ADBIS 2016, Prague, Czech Republic, August 29–31, 2016, Proceedings*, *Lecture Notes in Computer Science*, Vol. 9809. Springer, pp. 226–242.

Iftikhar, N. (2012). MMDW: a multi-dimensional and multi-granular schema for data warehousing. In: Graña, M., Toro, C., Posada, J., Howlett, R.J., Jain, L.C. (Eds.), *Advances in Knowledge-Based and Intelligent Information and Engineering Systems – 16th Annual KES Conference, San Sebastian, Spain, 10–12 September 2012*, *Frontiers in Artificial Intelligence and Applications*, Vol. 243. IOS Press, pp. 1211–1220.

Iftikhar, N., Pedersen, T.B. (2010). Schema Design Alternatives for Multi-granular Data Warehousing. In: Bringas, P.G., Hameurlain, A., Quirchmayr, G. (Eds.), *Database and Expert Systems Applications, 21th*

International Conference, DEXA 2010, Bilbao, Spain, August 30–September 3, 2010, Proceedings, Part II, *Lecture Notes in Computer Science*, Vol. 6262. Springer, pp. 111–125.

Lenz, H., Thalheim, B. (2005). OLAP schemata for correct applications. In: Draheim, D., Weber, G. (Eds.), *Trends in Enterprise Application Architecture, VLDB Workshop, TEAA 2005, Trondheim, Norway, August 28, 2005, Revised Selected Papers*, *Lecture Notes in Computer Science*, Vol. 3888. Springer, pp. 99–113.

Lenz, H., Thalheim, B. (2009). A formal framework of aggregation for the OLAP-OLTP model. *Journal of Universal Computer Science*, 15(1), 273–303.

Lenzerini, M. (2002). Data integration: a theoretical perspective. In: Popa, L., Abiteboul, S., Kolaitis, P.G. (Eds.), *Proceedings of the Twenty-first ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 3–5, Madison, Wisconsin, USA*. ACM, pp. 233–246.

Maier, D. (1983). *The Theory of Relational Databases*. Computer Science Press, Rockville.

Molnár, A. (2007). *A General Partition Data Model and a Contribution to the Theory of Functional Dependencies*. PhD thesis, Eötvös Loránd University, Hungary, Faculty of Informatics.

Monk, J.D. (1976). *Mathematical Logic*. Springer-Verlag, Berlin.

Ng, W. (2001). An extension of the relational data model to incorporate ordered domains. *ACM Transactions on Database Systems*, 26(3), 344–383.

Peters, J.F., Wasilewski, P. (2012). Tolerance spaces: origins, theoretical aspects and applications. *Information Sciences*, 195, 211–225.

Randell, D.A., Cui, Z., Cohn, A.G. (1992). A spatial logic based on regions and connection. In: Nebel, B., Rich, C., Swartout, W.R. (Eds.), *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR'92). Cambridge, MA, October 25–29, 1992*. Morgan Kaufmann, pp. 165–176.

Reiter, R. (1978). On closed world data bases. In: Gallaire, H., Minker, J. (Eds.), *Logic and Data Bases*. Plenum Press, New York, pp. 55–76.

Rodríguez, M.A., Bravo, L. (2012). Multi-granular schemas for data integration. In: Freire, J., Suciu, D. (Eds.), *Proceedings of the 6th Alberto Mendelzon International Workshop on Foundations of Data Management, Ouro Preto, Brazil, June 27–30, 2012*, *CEUR Workshop Proceedings*, Vol. 866. CEUR-WS.org, pp. 142–153.

Spyratos, N. (1987). The partition model: a deductive database model. *ACM Transactions on Database Systems*, 12(1), 1–37.

Szlichta, J., Godfrey, P., Gryz, J. (2012). Fundamentals of order dependencies. *Proceedings of the VLDB Endowment*, 5(11), 1220–1231.

Wang, S., Liu, D. (2004). Spatio-temporal database with multi-granularities. In: Li, Q., Wang, G., Feng, L. (Eds.), *Advances in Web-Age Information Management: 5th International Conference, WAIM 2004, Dalian, China, July 15–17, 2004*, *Lecture Notes in Computer Science*, Vol. 3129. Springer, pp. 137–146.

Wijsen, J., Ng, R.T. (1999). Temporal dependencies generalized for spatial and other dimensions. In: Böhlen, M.H., Jensen, C.S., Scholl, M. (Eds.), *Spatio-Temporal Database Management, International Workshop STDBM'99, Edinburgh, Scotland, September 10–11, 1999, Proceedings*, *Lecture Notes in Computer Science*, Vol. 1678. Springer, pp. 189–203.

Zeeman, E.C. (1962). The topology of the brain and visual perception. In: Fort, M.K. (Ed.), *Topology of 3-Manifolds*. Prentice-Hall, Englewood Cliffs, pp. 240–256.

S.J. Hegner was a senior lecturer in the Department of Computing Science at Umeå University, Sweden. He is now retired. His main area of research interest is database systems.

M.A. Rodríguez is a full-time professor in the Department of Computer Science at the University of Concepción, Chile. Her research interests include spatial and spatio-temporal databases, data integration, and information retrieval.