Structuring GIS information with types of granularity: a case study

Estructuración de información SIG con tipos de granularidad: un estudio de caso

C. Maria Keet

Faculty of Computer Science, Free University of Bozen-Bolzano, Italy
keet@inf.unibz.it

VI Conference on Geomatics, 10-13 February 2009, la Habana, Cuba

1. Scope and Problems
   - Motivation
   - Problem analysis
2. Foundations of granularity
   - Types of granularity
   - Lean theory of granularity
3. Problems revisited
4. Conclusions

Setting

- Granularity is a essential dimension in the subject domains of GISs
- Long-term goal is to manage in one system the granulated instance data, type-level information and knowledge, scale-based granularity and non-scale-based granularity
- Such a system has to be usable, *reusable*, *interoperable*, and *scalable*

Related works on granularity

- Data-centric focus (e.g., Bittner, Rigaux, Stell, Zhou [3, 11, 12, 13, 14]), OGC, GML
- Minor adornments in conceptual data modelling languages (Oracle Cartridge, MADS [9], DISTIL [10], MultiDimER [8]); ‘semantic granularity’ noted but not widely investigated
- Reduction in resolution and ‘hiding’ attributes or whole objects, set theory and mereology, ‘horizontal and ‘vertical’ components [3, 4, 11, 12]
- Known problems with choosing the wrong level of granularity; e.g., spatial-temporal niche partitioning of grassland ants [1]

Example

USDA ecological units vs. Köppen: mixing types and instances, area-based vs. *combination of properties* independent of a particular area and time:

Köppen		**USDA's "Ecoregion equivalents"**
Warm temperate climate (C)	$\equiv$	Humid Temperate Domain (200)
$\uparrow$		$\uparrow$
Warm temperate climate, dry summer (Cs)	$\equiv$	Mediterranean Division (260)
$\uparrow$		$\uparrow$
no equivalent given in [2] (Warm temperate climate, (dry, hot summer (Csa) [7])	$\not\equiv$	California coastal steppe, Mixed forest, and Redwood forest Province (263)

Problems

- Implementation-focus makes it difficult to reuse in a similar setting
- There is little, and lack of consensus, on representing and using granularity at the semantic layer
- Cumbersome to compute granular levels, not as ‘first-class citizen’ available for modelling and operation
- There is no systematic approach to and mechanism to devise perspectives / views / contexts other than that there are granulation hierarchies
- Interplay between quantitative and qualitative aspects of granularity, linking levels, hierarchies

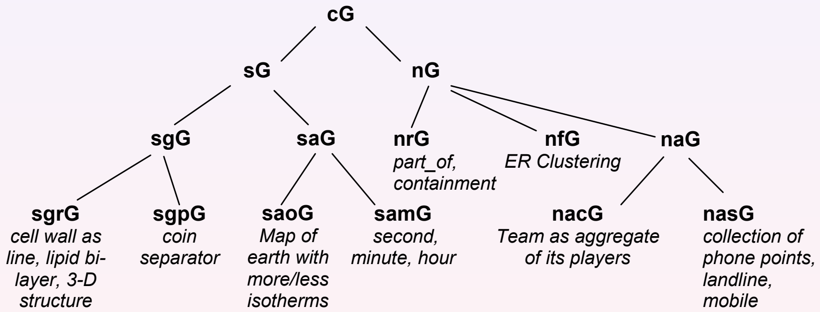
Foundational semantics of granularity

- Extract ‘patterns’ of granulation, i.e, types of granulation hierarchies and ways how levels are identified
- Thus, identifying **mechanisms of granulation**
- Each mechanism is *subject domain-independent* and *implementation-independent* because the focus is on foundational semantics, hence, reusable and facilitating interoperability

Foundational semantics of granularity

- Extract ‘patterns’ of granulation, i.e, types of granulation hierarchies and ways how levels are identified
- Thus, identifying **mechanisms of granulation**
- Each mechanism is *subject domain-independent* and *implementation-independent* because the focus is on foundational semantics, hence, reusable and facilitating interoperability

Foundational semantics of granularity



Basic TOG

- Fragment of the TOG [6] that is a logical theory in FOL with model-theoretic semantics
- Advantages of the ontological motivations for the definitions, derivations of constraints, yet smaller so that it may be easier to implement in real systems
- With the features—such as level, perspective, criterion of granulation, relation between levels, perspectives, and linkage to types of granularity—we can address the problems outlined and demonstrate the modelling approach

Basic TOG

- Fragment of the TOG [6] that is a logical theory in FOL with model-theoretic semantics
- Advantages of the ontological motivations for the definitions, derivations of constraints, yet smaller so that it may be easier to implement in real systems
- With the features—such as level, perspective, criterion of granulation, relation between levels, perspectives, and linkage to types of granularity—we can address the problems outlined and demonstrate the modelling approach

Excerpt (definitions)

Definition 1 (Granular perspective)

$\forall x \exists ! w, y, z, \phi$ such that $GP(x)$ is a concept $CN(x)$, has a definition $DF(x, y)$, relates to its criterion $C(z)$ through the relation $RC(x, z)$, has_granulation type $TG(\phi)$ and is contained in $D^f(w)$.

$$\forall x (GP(x) \triangleq \exists w, y, z, \phi (DF(x, y) \wedge RC(x, z) \wedge C(z) \wedge RE(x, w) \wedge has\_granulation(x, \phi))) \quad (1)$$

DEFINITION 2 (Granular level)

*$\forall x \exists! v, w, y, z \exists p$ such that $GL(x)$ is a concept $CN(x)$, has a definition $DF(x, y)$, is related to $GP(w)$ with $RE(x, w)$ and uses criterion $C(z)$ with $RC(w, z)$ and has_value(z, v) where the value is in region $V(v)$ for any $GL(x)$ that adheres_to **sG**, $GL^s(x)$, and z's label for any $GL(x)$ that adheres_to type **nG**, $GL^n(x)$. Entities residing in $GL^s(x)$ are similar to each other with respect to (the value z of) $V(v)$, entities residing in $GL^n(x)$ are similar to each other with respect to (the label of the universal of) Prop(p) of $C(z)$, and both are φ-indistinguishable with respect to its adjacent coarser-grained level.*

$$\forall x(GL(x) \triangleq \exists! v, w, y, z(DF(x, y) \wedge GP(w) \wedge RE(x, w) \wedge C(z) \wedge RC(w, z) \wedge R(v) \wedge has\_value(z, v))) \quad (2)$$

Excerpt (constraints)

- Part-whole relations for *RE*, *RL*, and *GR*.
- Why some entity (/type) resides in a level, with similarity, indistinguishability, and equivalence [5], resulting in:

THEOREM 1 (3.2)

A granular perspective GP must contain at least two granular levels GL: $\forall x(GP(x) \rightarrow \exists^{\geq 2} y(RE^{-}(x,y) \wedge GL(y)))$

THEOREM 2 (3.1)

The combination of some $C(y)$ *with a* $TG(\phi)$ *determines uniqueness of each* $GP(x)$.

more theorems

Excerpt (constraints)

- Part-whole relations for *RE*, *RL*, and *GR*.
- Why some entity (/type) resides in a level, with similarity, indistinguishability, and equivalence [5], resulting in:

THEOREM 1 (3.2)

A granular perspective GP must contain at least two granular levels GL: $\forall x(GP(x) \rightarrow \exists^{\geq 2} y(RE^{-}(x, y) \wedge GL(y)))$

THEOREM 2 (3.1)

The combination of some $C(y)$ *with a* $TG(\phi)$ *determines uniqueness of each* $GP(x)$.

more theorems

Toward implementations

- Need for three principal components:
  - the types of granularity that link to the basic TOG
  - an instantiation (model) of this theory for a specific subject domain
  - a data source to be granulated
- E.g., perspective Biogeography (π_i) with level Biotope (λ_i) and at least one other level (e.g., Bioregion, λ_j), criterion scale-delimited biogeography (υ_i), and granulation type samG (θ_i)

Toward implementations

- Need for three principal components:
  - the types of granularity that link to the basic TOG
  - an instantiation (model) of this theory for a specific subject domain
  - a data source to be granulated
- E.g., perspective Biogeography (π_i) with level Biotope (λ_i) and at least one other level (e.g., Bioregion, λ_j), criterion scale-delimited biogeography (υ_i), and granulation type samG (θ_i)

Sample granular perspectives

Spatial data representation		Conditional perspectives		
Shape (π_1)	**Raster (π_2)** (Size in m)	**Admin (π_3)**		**Hydro (π_4)** (river with flow $\geq$)
Point	1000	Country	$\Leftrightarrow$	100 000 litres/min
$\uparrow$	$\uparrow$	$\uparrow$		$\uparrow$
Line	100	Province	$\Leftrightarrow$	10 000 litres/min
$\uparrow$	$\uparrow$	$\uparrow$		$\uparrow$
Polygon	10	Region	$\Leftrightarrow$	2500 litres/min
$\uparrow$	$\uparrow$	$\uparrow$		$\uparrow$
Polyhedron	1	Municipality	$\Leftrightarrow$	1000 litres/min
		$\uparrow$		$\uparrow$
		Municipality district	$\Leftrightarrow$	250 litres/min

Π	Θ	Υ criterion	Γ gran.rel.	**Comments**
π_1	θ_1 = sgrG	υ_1 = GIS vector-based spatial data representation	γ_1 = *has_ppart*	relation to the granulated entity, relation to resolution and how to convert between these resolutions
π_2	θ_2 = saoG	υ_2 = GIS raster-based spatial data representation	γ_2 = *ppart_of*	additional conversion function to aggregate the squares into the next coarser level, relation to the granulated entity
π_3	θ_3 = nrG	υ_3 = Administrative region	γ_3 = *contained_in*	
π_4	θ_4 = sgpG	υ_4 = River water throughput	–	
π_5	θ_5 = saoG	υ_5 = July isotherm, average		optional aggregation function to move from finer-to coarser-grained level, linked to an administrative region entity
π_6	θ_6 = saoG	υ_6 = Yearly precipitation, average		optional aggregation function to move from finer-to coarser-grained level, linked to an administrative region entity

Conditional selections

- "*if one makes a map with granularity at the Province-level then only rivers with a flow* $\geq$ *10 000 litres/min should be included in the map*"
- With $\mathcal{G}$ and two functions to select a level (*selectL* : $\mathcal{L} \mapsto \mathcal{L}$, with $\mathcal{L}$ the set of all levels $\lambda_1 ... \lambda_n$) and retrieve the contents of a level (*getC* : $\mathcal{L} \mapsto \mathcal{E}$, and $\mathcal{E}$ the collection of universals or particulars residing in a level λ_i), we can generalise this into a constraint pattern for conditional selection and retrieval (where $i \neq j$):
- ***if*** *selectL*(λ_i) ***and*** *getC*(λ_i) ***where*** $r_e(\lambda_i, \pi_i)$***, then*** *selectL*(λ_j) ***and*** *getC*(λ_j)***, where*** $r_e(\lambda_j, \pi_j)$***, as well***

Reassessing extant hierarchies

Table: Varying scales at different levels of regions as well as within-scale variations (values populating the levels are taken from maps in the Dutch “Grote Bos Atlas”).

	Avg. July temperature (π_5) (°C)	**Avg. Yearly Precipitation (π_6)** (in mm)
λ_1 World	0 – 10 – 20 – 30	$<$250 – 250-500 – 500-1000 – 1000-2000 – $\geq$2000
	↑	↑
λ_2 Europe (EU)	$<$10 – 10-15 – 15-17.5 – 17.5-20 – 20-25 – $\geq$25	$<$200 – 200-400 – 400-600 – 600-800 – 800-1200 – 1200-2000 – $\geq$2000
	↑	↑
λ_3 Netherlands (country)	16 – 16.5 – 17 – 17.5	$<$750 – 750-800 – 800-850 – 850-900 – $\geq$900

Conclusions and current work

- Types of granularity and a basic framework for modelling granularity
- Lifting it up to a higher level of abstraction independent of design and implementation $\longrightarrow$ declaring explicitly the levels, perspectives, criteria for granulation, mechanism of granulation
- Illustrated the simplifications for modelling granulation hierarchies in GIS and GIS-enabled ecology transparently, consistently, and in a reusable manner
- Thus facilitating flexibility, reusability, transparency, interoperability of implementations

Thank you for your attention

M. Albrecht and N. J. Gotelli, "Spatial and temporal niche partitioning in grassland ants," *Oecologia*, vol. 126, pp. 134–141, 2001.

R. G. Bailey, "Description of the ecoregions of the United States," USDA Forest Service," Technical Report, March 1995, http://www.fs.fed.us/land/ecosysmgmt/ecoreg1_home.html, Date last modified: 8-17-2001 (Date accessed 20-8-2007).

T. Bittner and B. Smith, *Foundations of Geographic Information Science*. London: Taylor & Francis Books, 2003, ch. A Theory of Granular Partitions, pp. 117–151.

F. Fonseca, M. Egenhofer, C. Davis, and G. Camara, "Semantic granularity in ontology-driven Geographic Information Systems," *Annals of Mathematics and Artificial Intelligence*, vol. 36, no. 1-2, pp. 121–151, 2002.

C. M. Keet, "Granulation with indistinguishability, equivalence or similarity," in *IEEE International Conference on Granular Computing (GrC2007)*, vol. 2. IEEE Computer Society, 2007, pp. 11–16, san Francisco, November 2-4, 2007.

——, "A formal theory of granularity," PhD Thesis, KRDB Research Centre, Faculty of Computer Science, Free University of Bozen-Bolzano, Italy, April 2008.

M. Kottek, J. Grieser, C. Beck, B. Rudolf, and F. Rubel, "World map of the Köppen-Geiger climate classification updated," *Meteorologische Zeitschrift*, vol. 15, no. 3, pp. 259–263, 2006.

E. Malinowski and E. Zimányi, "Logical representation of a conceptual model for spatial data warehouses," *Geoinformatica*, vol. 11, pp. 431–457, 2007.

C. Parent, S. Spaccapietra, and E. Zimányi, *Conceptual modeling for traditional and spatio-temporal applications—the MADS approach*. Berlin Heidelberg: Springer Verlag, 2006.

S. Ram, R. Snodgrass, V. Khatri, and Y. Hwang, "DISTIL: a design support environment for conceptual modeling of spatio-temporal requirements," in *Proceedings of the 20th International Conference on Conceptual Modeling (ER'01)*, ser. LNCS, vol. 2224. Springer, Berlin, 2001, pp. 70–83.

P. Rigaux and M. Scholl, "Multi-scale partitions: applications to spatial and statistical databases," in *Proceedings of the 4th International Symposium on Advances in Spatial Databases (SSD'95)*, ser. LNCS, vol. 951. Springer, Berlin, 1995, pp. 170–183.

J. Stell and M. Worboys, "Stratified map spaces: a formal basis for multi-resolution spatial databases," in *Proceedings of the 8th International Symposium on Spatial Data Handling (SDH'98)*. International Geographical Union, 1998, pp. 180–189.

S. Zhou and C. Jones, "A multi-representation spatial data model," in *Proceedings of the 8th International Symposium on Advances in Spatial and Temporal Databases (SSTD 2003)*, ser. LNCS, vol. 2750. Springer, Berlin, 2003, pp. 394–411.

X. Zhou, S. Prasher, S. Sun, and K. Xu, "Multiresolution spatial databases: making web-based spatial applications faster," in *Proceedings of the 6th Asia-Pacific Web Conference (APWeb 2004)*, ser. LNCS, vol. 3007. Springer, Berlin, 2004, pp. 36–47.

Foundational semantics of granularity

Table: Distinguishing characteristics at the branching points in the taxonomy of types of granularity.

Branching point	**Distinguishing feature**
sG – nG	scale – non-scale (or, roughly: quantitative – qualitative)
sgG – saG	grain size – aggregation (or: scale *on* entity – scale *of* entity)
sgrG – sgpG	resolution – size of the entity
saoG – samG	overlay aggregated – entities aggregated according to scale
naG – nrG – nfG	semantic aggregation – one type of relation between entities in different levels – different type of relation between entities in levels and relations among entities in level
nacG – nasG	parent-child not taxonomic and relative independence of contents of higher/lower level – parent-child with taxonomic inheritance

Excerpt (constraints)

THEOREM 3 (3.5)

RL is of the same type, s_ppart_of, not only within some particular instance of GP, but it is of the same type between granular levels in all granular perspectives.

THEOREM 4 (3.6)

The multiplicity (cardinality) of RL and RL^- is 1:1, i.e. $\forall x \exists! y (RL(x, y))$ *and* $\forall x \exists! y (RL^-(x, y))$.

Excerpt (constraints)

LEMMA 1 (3.19)

Two levels in different perspectives can overcross:
$\forall x, y(overcross(x, y) \wedge GL(x) \wedge GL(y) \wedge \neg(x = y) \rightarrow$
$\exists v, w(R_E(x, v) \wedge R_E(y, w) \wedge \neg(v = w)))$.

THEOREM 5 (3.7)

If two levels in different perspectives overcross, then their perspectives overcross:
$\forall x_1, x_2, y_1, y_2(overcross(x_1, x_2) \wedge GL(x_1) \wedge GL(x_2) \wedge GP(y_1) \wedge$
$GP(y_2) \wedge R_E(x_1, y_1) \wedge R_E(x_2, y_2) \rightarrow overcross(y_1, y_2))$.

back to main