# A Formal Theory of Granularity

*Toward enhancing life science information systems with granularity*

**Catharina Maria Keet**

Advisor: Prof. Enrico Franconi
Co-Advisor: Dr. Alessandro Artale

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

*PhD Thesis Defense, 28 April 2008*

# Outline

## Background

- Information overload, (non-)interoperability of software tools [Keet05a]
  - Need for *vertical* integration for linking databases and ontologies. e.g. the "virtual human", metagenomics, health-GIS, pharmacoinformatics [Keet&Roos&Marshall07]
  - User should be able to select section(s) of interest and retrieve just that
- Too many one-off bioinformatics tools
  - Need for new foundational methodologies to push implementations to the next phase of *in silico* biology and biomedicine

# Granularity

- **How to (computationally) manage coarser- and finer-grained (less/more detailed) data, information, and knowledge?**
- Granular computing at present
  - distinction "structured thinking" & "structured problem solving" [Y05]; a.o. fuzzy logic, (rough) mereology, data mining, rough sets
  - quantitative & a bit (mostly informal) qualitative
- KR: set-theoretic or mereology
- Subject domains: mainly GIS, biology & biomedicine, spatial, temporal
- Theories & languages, a.o.: TGP [BS03], $\mathcal{GMD}$ [FK03] and MultiDimER [MZ06], MADS [PSZ06]

# Granularity

- **How to (computationally) manage coarser- and finer-grained (less/more detailed) data, information, and knowledge?**
- Granular computing at present
  - distinction "structured thinking" & "structured problem solving" [Y05]; a.o. fuzzy logic, (rough) mereology, data mining, rough sets
  - quantitative & a bit (mostly informal) qualitative
- KR: set-theoretic or mereology
- Subject domains: mainly GIS, biology & biomedicine, spatial, temporal
- Theories & languages, a.o.: TGP [BS03], $\mathcal{GMD}$ [FK03] and MultiDimER [MZ06], MADS [PSZ06]

# Problems with granularity

- **Engineering**: are not reusable in the current format beyond the difficult to scale-up software application each one is designed for that are not-interoperable.
- **Informal**: are not usable or reusable for computation and reasoning, are ontologically inconsistent, and underspecified.
- **Formal**: are more or less compatible partial data-centric theories that neither address what granularity is nor what its components are nor how to use it.

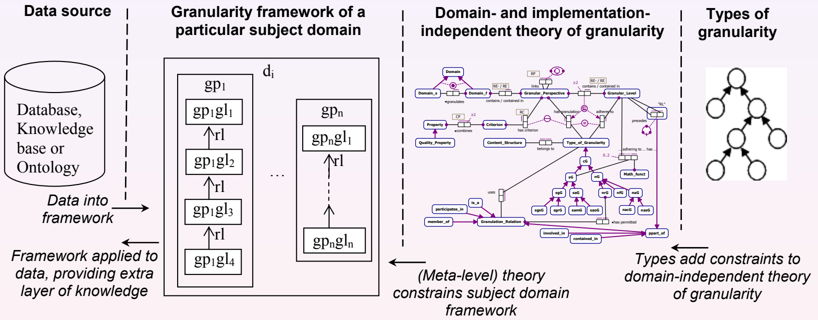
# Research question and Aims

- The main **research question**: *Why, how, and where will usage of granularity improve knowledge representation and knowledge management?*
- The **aims**:
    - Develop a formal, domain- and implementation- independent theory of granularity (TOG) that can be used for computational reasoning.
    - This TOG is integrated with data- and knowledge bases and ontologies to enhance data, information and knowledge management, including querying and reasoning across levels of granularity.
    - The TOG will be sufficiently comprehensive to be useful in the subject domain of biology.

## Research question and Aims

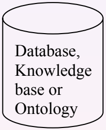
- The main **research question**: *Why, how, and where will usage of granularity improve knowledge representation and knowledge management?*
- The **aims**:
  - Develop a formal, domain- and implementation- independent theory of granularity (TOG) that can be used for computational reasoning.
  - This TOG is integrated with data- and knowledge bases and ontologies to enhance data, information and knowledge management, including querying and reasoning across levels of granularity.
  - The TOG will be sufficiently comprehensive to be useful in the subject domain of biology.

# Proposed solution



# Proposed solution

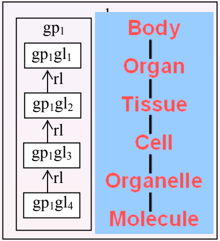
**Data source**



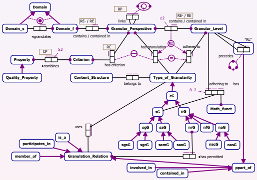

*Data into framework*

*Framework applied to data, providing extra layer of knowledge*

**Granularity framework of a particular subject domain**

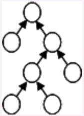



**Domain- and implementation-independent theory of granularity**



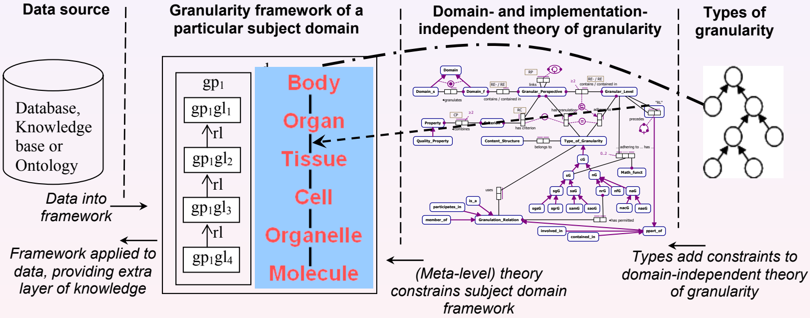
*(Meta-level) theory constrains subject domain framework*

**Types of granularity**

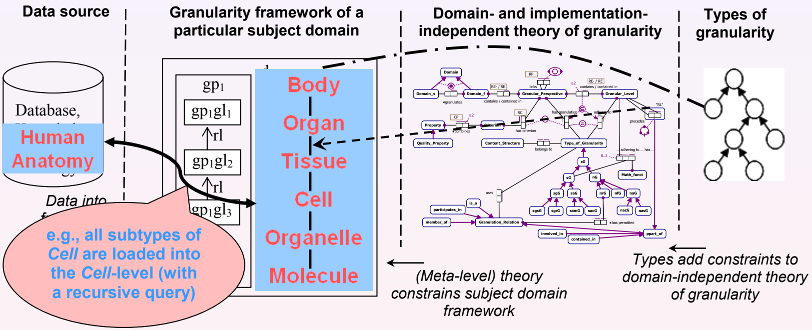


*Types add constraints to domain-independent theory of granularity*

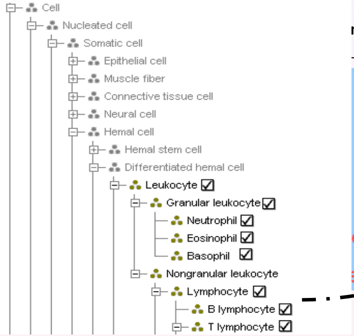
# Proposed solution



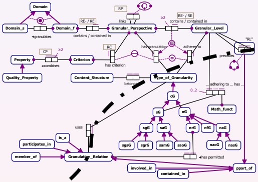
# Proposed solution



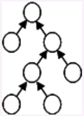
# Proposed solution




**Domain- and implementation-independent theory of granularity**



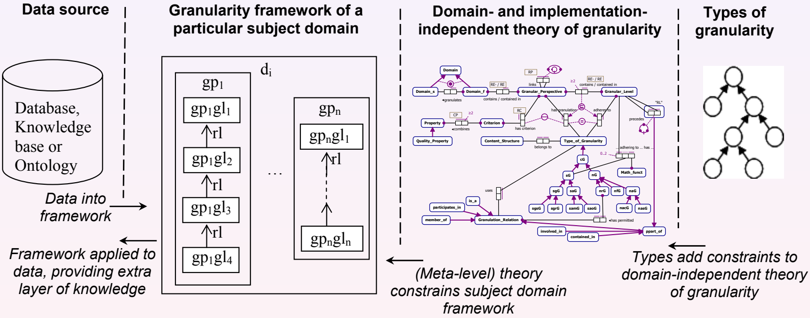

**Types of granularity**



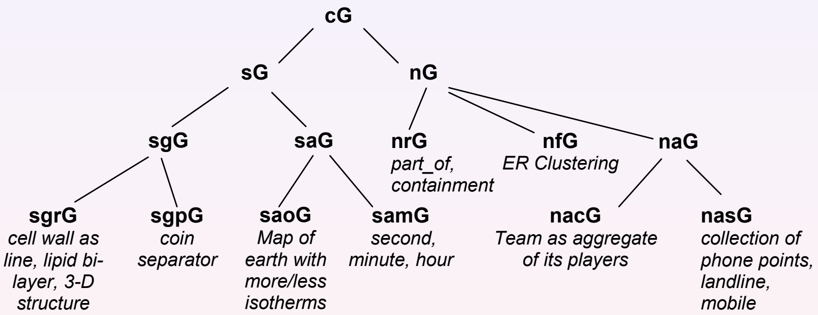
*(Meta-level) theory constrains subject domain framework*

*Types add constraints to domain-independent theory of granularity*

# Proposed solution



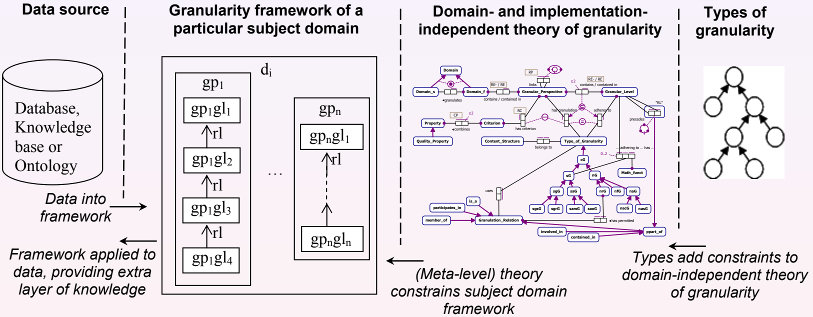
# Foundational semantics of granularity




branching points

key requirements

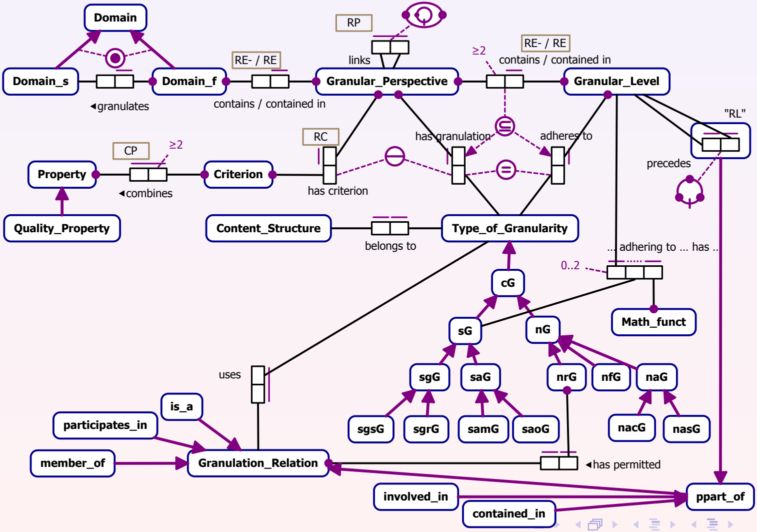
# Proposed solution



# Main characteristics

- Ontologically-motivated, e.g., taxonomy of part-whole relations view [Keet06b, Keet&Artale07], characterisation similarity & indistinguishability [Keet07a], DOLCE
- FOL, model theoretic semantics
- 27 definitions, 19 propositions, 19 lemmas, 7 theorems, and 6 corollaries
- Tested with Mace4 model searcher

# Graphical rendering of main components



# 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 $D^f(w)$.*

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

# Excerpt (definitions)

**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 $\varphi$-indistinguishable with respect to its adjacent coarser-grained level.*

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

# Excerpt (constraints)

- Part-whole relations for *RE*, *RL*, and *GR*.
- Why some entity (/type) resides in a level, with similarity, indistinguishability, and equivalence [Keet07a], 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 [Keet07a], 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

# Functions

- Querying the TOG components [Keet08a]
  - **19 functions** for: level assignment; selection and retrieval of the domain, perspective, and level; retrieval of contents of a level and intersection of contents from different levels; and selection of one or more entities (/types) and subsequent retrieval of the one or more levels it (/they) reside in.
    $\Rightarrow$ One can define functions to retrieve each component of the TOG.

# Functions

- Querying the TOG components [Keet08a]
- Abstraction & expansion functions [Keet05b, Keet07b]
  - 3 Main modes of abstraction
  - 27 entity (/type)-focussed abstraction and expansion functions
  - Using, a.o., the *granulation relations* for abstractions
  - Use of *types of granularity* to achieve correct behaviour of expansions
- Reasoning over role hierarchies with the *RBox compatibility service* [Keet&Artale07]

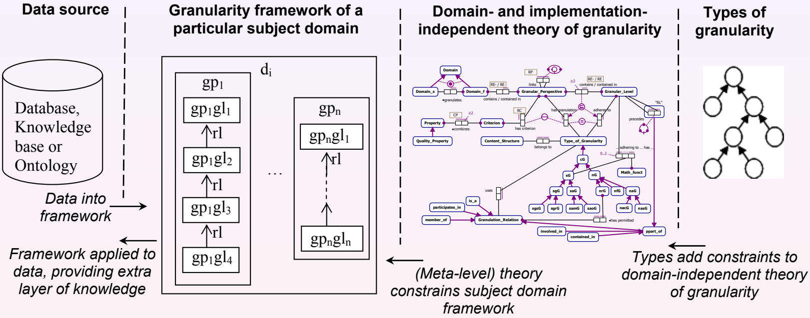
# Functions

- Querying the TOG components [Keet08a]
- Abstraction & expansion functions [Keet05b, Keet07b]
  - 3 Main modes of abstraction
  - 27 entity (/type)-focussed abstraction and expansion functions
  - Using, a.o., the *granulation relations* for abstractions
  - Use of *types of granularity* to achieve correct behaviour of expansions
- Reasoning over role hierarchies with the *RBox compatibility service* [Keet&Artale07]

# Basic experiments and examples

- TOG verified with Mace4 model searcher data; computed proofs with Prover9
- Investigation from FOL to DL, complexity of language required to represent the TOG (at least ExpTime) analysis

# Proposed solution



# Basic experiments and examples

- Infectious diseases [Keet&Kumar05, Keet06e]
- Other (manual) examples, with, a.o., Second messenger system, OBO Foundry [Keet08a] view
- Querying the GO database and FMA through OQAFMA [Keet06d]

more scenarios

# Conclusions

- **Foundational semantics of granularity** is disambiguated and structured in a taxonomy of types of granularity [Keet06a];
- The **static components of granularity**—such as levels, indistinguishability, how finer- and coarser-grained levels and entities relate, and granulation criteria—were subjected to an ontological analysis and formalised in a consistent and satisfiable logical theory (TOG) with model-theoretic semantics;
- An extensible set of domain- and implementation-independent **functions** were defined for both the TOG elements to enable granular querying and reasoning over the theory and for moving between entities residing in different levels through abstractions and expansions. [Keet07b]

# Conclusions

- **Foundational semantics of granularity** is disambiguated and structured in a taxonomy of types of granularity [Keet06a];
- The **static components of granularity**—such as levels, indistinguishability, how finer- and coarser-grained levels and entities relate, and granulation criteria—were subjected to an ontological analysis and formalised in a consistent and satisfiable logical theory (TOG) with model-theoretic semantics;
- An extensible set of domain- and implementation-independent **functions** were defined for both the TOG elements to enable granular querying and reasoning over the theory and for moving between entities residing in different levels through abstractions and expansions. [Keet07b]

# Conclusions

- **Foundational semantics of granularity** is disambiguated and structured in a taxonomy of types of granularity [Keet06a];
- The **static components of granularity**—such as levels, indistinguishability, how finer- and coarser-grained levels and entities relate, and granulation criteria—were subjected to an ontological analysis and formalised in a consistent and satisfiable logical theory (TOG) with model-theoretic semantics;
- An extensible set of domain- and implementation-independent **functions** were defined for both the TOG elements to enable granular querying and reasoning over the theory and for moving between entities residing in different levels through abstractions and expansions. [Keet07b]

# Future work

- Relation between mereology and granularity (w.r.t. mereogeometry [BM07] and aggregates), rough/fuzzy extensions for level specification, integrating rules for conditional level
- DL languages ($\mathcal{DLR}x$ and/or OWL 1.1), granularity-enhanced conceptual data modelling, vertical integration of ontologies or conceptual data models
- Types of granularity as extension to DOLCE
- Address Ontology Comprehension Problem with granularity and abstraction and visualisation of metabolic pathways.
- Ontology linking and cross-granular querying: experiment with the OBO Foundry [Keet08a, Keet&Roos&Marshall07, 3]

summary

# Main Research Question

*Why, how, and where will usage of granularity improve knowledge representation and knowledge management?*

# Why

Granularity is a novel way of structured knowledge representation that is orthogonal to extant modelling methods. It enables a *structured and consistent way for carving up the subject domain* through granularity type selection and criterion-selection for a particular granular perspective and at different levels of detail so that a user can avoid unnecessary detail or too coarse generalisations, yet have the full domain at one's disposal for reasoning across levels of granularity, and also be deployed for users with different foci.

# How

The formal, hence unambiguous, Theory of Granularity, TOG, can be added to ontologies, knowledge bases, database, and conceptual data models with a formal foundation. That is, the hitherto data-centric solutions for dealing with granularity are lifted to the ontological and logical layers and the informal subject domain and philosophical contributions are precisiated in a formal theory, thereby *providing easily reusable modelling constructs for granularity components* such as level, perspective, type of granularity, and granulation criterion. These static components are augmented with *clear, extensible functions for granular querying and reasoning* that are reusable across implementation scenarios.

# Where

The structured and consistent representation of granularity enabled by the TOG contributes to *good modelling methodology and practice, hence, to the overall quality of the data source*. The genericity of the TOG can guarantee *interoperability between granulated data sources*, such as GISs and bio-ontologies, because its FOL representation is implementation- and scenario-independent. Further, it offers an additional method for *linking data sources that contain contents at different levels of granularity*, and, subsequently, *reasoning across those levels of granularity*.

# Summary

- General problems (vertical integration, visualisation, foundational methodolgies) and specific for granularity (data-centric, informal or partial formal theories, non-reusable engineering solutions)
- Overview of architecture for a granulated information system
- Main characteristics of the TOG
  - From a data-centric treatment of granularity to the ontological and logical layers
  - Elements of granularity have become ontologically-motivated modelling constructs proper
  - Representation of granularity is now domain- and implementation independent
  - Reusability across implementations is ensured, which in turn facilitates interoperability among information systems
- Toward implementation, with exploration into DLs and bio-experiments

# Publications by thesis chapter

- **Chapter 1—Introduction and methodology**:

  **1.** [Keet06f] Keet, C.M. Enhancing biological information systems with granularity. *KnowledgeWeb PhD Symposium (KWEPSY06)*, Budva, Montenegro, 17 June 2006.

  **2.** [Keet05a] Keet, C.M. Current characteristics and historical perspective of Computer Science and IT with/for biology. In: *CSBio reader: extended abstracts of the 'CS&IT with/for biology' Seminar Series*, Free University of Bozen-Bolzano. Keet, C.M., Franconi E. (eds.). 2005. pp1-8.

  **3.** [Keet&Kumar05] Keet, C.M., Kumar, A. (2005). Applying partitions to infectious diseases. *XIX International Congress of the European Federation for Medical Informatics (MIE2005)*, 28-31 August 2005, Geneva, Switzerland. 2005. In: Connecting Medical Informatics and bio-informatics, Engelbrecht, R., Geissbuhler, A., Lovis, C. Mihalas, G. (eds.). Amsterdam: IOS Press. pp1236-1241.

# Publications by thesis chapter

- **Chapter 1—Intro and methodology** (continued):
  **4.** [Keet05c] Keet, C.M. Factors affecting ontology development in ecology. *Data Integration in the Life Sciences (DILS2005)*. Ludäscher, B., Raschid, L. (eds.). San Diego, USA, 20-22 July 2005. Springer LNBI 3615, 46-62.
- **Chapter 2—Foundational Semantics**:
  **5.** [Keet06a] Keet, C.M. A taxonomy of types of granularity. *IEEE Conference in Granular Computing (GrC2006)*, 10-12 May 2006, Atlanta, USA. IEEE Computer Society, 106-111.
- **Chapter 3—The** TOG:
  **6.** [Keet07a] Keet, C.M. Granulation with indistinguishability, equivalence or similarity. *IEEE Int'l Conf. on Granular Computing (GrC2007)*, San Francisco, Nov 2-4, 2007. IEEE Computer Society, 11-16.

# Publications by thesis chapter

- **Chapter 3—The** TOG (continued):
  **7.** [Keet06c] Keet, C.M. *Introduction to part-whole relations: mereology, conceptual modelling and mathematical aspects*. KRDB Research Centre TR KRDB06-3, Faculty of CS, FUB, Italy. Oct 2006.
  **8.** [Keet06b] Keet, C.M. Part-whole relations in Object-Role Models. *2nd Int'l WS on Object-Role Modelling (ORM 2006)*, Montpellier, France, Nov 2-3, 2006. In: OTM Workshops 2006. Meersman, R., Tari, Z., Herrero., P. *et al.* (Eds.), Springer, LNCS 4278, 1116-1127.
  **9.** [Keet&Artale07] Keet, C.M., Artale, A. Representing and Reasoning over a Taxonomy of Part-Whole Relations. *Applied Ontology – S.I. on Ontological Foundations for Conceptual Models*, 2007.
- Keet, C.M. A formal theory to represent granularity. submitted to KR'08.

# Publications by thesis chapter

- **Chapter 3—The** TOG (continued):
  book chapter invitation: Artale, A., Keet, C.M. Modelling essential and shared parts in conceptual models. In: *Innovations in Information Systems modeling: Methods and Best Practices*. IGI Global, Advances in Database Research Series, Halpin, T.A., Proper, H.A., Krogstie, J.(Eds.).
- **Chapter 4—Functions**:
  **10.** [Keet08a] Keet, C.M. Toward cross-granular querying over modularized ontologies. *International Workshop on Ontologies: Reasoning and Modularity (WORM08)*. 2 June 2008, Tenerife, Spain. CEUR-WS Vol-xx.
  **11.** [Keet05b] Keet, C.M. Using abstractions to facilitate management of large ORM models and ontologies. *Int'l WS on Object-Role Modeling (ORM'05)*. In: OTM Workshops 2005. Halpin, T., Meersman, R. (eds.), Springer, LNCS 3762, 603-612

# Publications by thesis chapter

- **Chapter 4—Functions** (continued):
  **12.** [Keet07b] Keet, C.M. Enhancing comprehension of ontologies and conceptual models through abstractions. *10th Congress of the Italian Association for Artificial Intelligence (AI*IA 2007)*, Rome, September 10-13, 2007. Basili, R., Pazienza, M.T. (Eds.), Springer, LNAI 4733, 814-822.

  [Keet&Artale07] Keet, C.M., Artale, A. Representing and Reasoning over a Taxonomy of Part-Whole Relations. *Applied Ontology – Special Issue on Ontological Foundations for Conceptual Models*, 2007.
- **Chapter 5—Related works and toward implementation**:
  **13.** [Keet&Rodríguez07] Keet, C.M., Rodríguez, M. Toward using biomedical ontologies: trade-offs between ontology languages. *AAAI 2007 WS on Semantic e-Science (SeS07)*, 23 July 2007, Vancouver, Canada. AAAI 2007 TR WS-07-11, 65-68.

# Publications by thesis chapter

- **Chapter 5—Toward implementation** (continued):

**14.** [Keet&Roos&Marshall07] Keet, C.M., Roos, M., Marshall, M.S. A survey of requirements for automated reasoning services for bio-ontologies in OWL. *3rd Int'l WS OWL: Experiences and Directions (OWLED 2007)*, 6-7 June 2007, Innsbruck, Austria. CEUR-WS Vol-258.

**15.** [Keet07c] Keet, C.M. Prospects for and issues with mapping the Object-Role Modeling language into $\mathcal{DLR}_{ifd}$. *20th Int'l WS on Description Logics (DL'07)*, 8-10 June 2007, Bressanone, Italy. CEUR-WS, Vol-250, 331-338.

**16.** [Tziviskou&Keet07] Tziviskou, C., Keet, C.M. A Meta-Model for Ontologies with ORM2. *3rd Int'l WS on Object-Role Modelling (ORM 2007)*, Algarve, Portugal, Nov 28-30, 2007. In: *OTM Workshops 2007*. Meersman, R., Tari, Z, Herrero, P. et al (Eds.), Springer, LNCS 4805, 624-633.

# Publications by thesis chapter

- **Chapter 5—Toward implementation** (continued):
  **17.** [Keet08b] Keet, C.M. A formal comparison of conceptual data modeling languages. *13th International Workshop on Exploring Modeling Methods in Systems Analysis and Design (EMMSAD'08)*. 16-17 June 2008, Montpellier, France. CEUR-WS (ISSN 1613-0073), Vol-xx. 15p.
- **Experimentation**:
  **18.** [Keet06d] Keet, C.M. *Granular information retrieval from the Gene Ontology and from the Foundational Model of Anatomy with OQAFMA*. Technical Report KRDB06-1, Faculty of Computer Science, Free University of Bozen-Bolzano, 2006. Extended abstract from *Bioinformatics, Ontologies and databases (SBIOLBD 2006)*, 13-2-2006, EPFL, Lausanne, Switzerland.
  **19.** [Keet06e] Keet, C.M. *A formal approach for using granularity with the infectious diseases domain*. Technical Report KRDB06-2, Faculty of Computer Science, FUB, 2006.

# Publications by thesis chapter

- **Experimentation** (continued):

  [Keet&Kumar05] Keet, C.M., Kumar, A. (2005). Applying partitions to infectious diseases. *XIX International Congress of the European Federation for Medical Informatics (MIE2005)*, 28-31 August 2005, Geneva, Switzerland. 2005. In: Connecting Medical Informatics and bio-informatics, Engelbrecht, R., Geissbuhler, A., Lovis, C. Mihalas, G. (eds.). Amsterdam: IOS Press. pp1236-1241.

# References

[BS03] Bittner, T., Smith, B. A Theory of Granular Partitions. In: *Foundations of Geographic Information Science*, Duckham, M, Goodchild, MF, Worboys, MF (eds.), London: Taylor & Francis Books, 2003, pp117-151.

[BM07] Borgo, S., Masolo, C. Full mereogeometries. *Journal of Philosophical Logic*, 2007. (to appear).

[CGL98] Calvanese, C., De Giacomo, G., Lenzerini, M. On the decidability of query containment under constraints. In: *Proceedings of the 17th ACM SIGACT SIGMOD SIGART Sym. on Principles of Database Systems (PODS'98)*, 149-158.

[CGL99] Calvanese, D., De Giacomo, G., Lenzerini, M. Reasoning in expressive description logics with fixpoints based on automata on infinite trees. In: *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI'99)*, 84-89.

[CGL01] Calvanese, D., De Giacomo, G., Lenzerini, M. Identification constraints and functional dependencies in Description Logics. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI 2001)*, 2001, 155-160.

[CDGLR03] Calvanese, D., Damaggio, E., De Giacomo, G., Lenzerini, M., Rosati, R. Semantic data integration in P2P systems. *International Workshop on Databases, Information Systems and Peer-to-Peer Computing*, 2003.

[FK03] Franconi, E., Kamble, A. The GMD Data Model for Multidimensional Information: a brief introduction. *5th International Conference on Data Warehousing and Knowledge Discovery (DaWaK'03)*, Prague, Czech Republic, 2003.

[Keet05a] Keet, C.M. Current characteristics and historical perspective of Computer Science and IT with/for biology. In: *CSBio reader: extended abstracts of the 'CS&IT with/for biology' Seminar Series*, Free University of Bozen-Bolzano. Keet, C.M., Franconi E. (eds.). 2005. pp1-8.

# References

[Keet05b] Keet, C.M. Using abstractions to facilitate management of large ORM models and ontologies. *International Workshop on Object-Role Modeling (ORM'05)*. In: OTM Workshops 2005. Halpin, T., Meersman, R. (eds.), Berlin: Springer-Verlag, Lecture Notes in Computer Science **3762**, 603-612.

[Keet05c] Keet, C.M. Factors affecting ontology development in ecology. *Data Integration in the Life Sciences (DILS2005)*. Ludäscher, B., Raschid, L. (eds.). San Diego, USA, 20-22 July 2005. Springer Verlag, Lecture Notes in Bioinformatics **3615**, 46-62.

[Keet06a] Keet, C.M. A taxonomy of types of granularity. *IEEE Conference in Granular Computing (GrC2006)*, 10-12 May 2006, Atlanta, USA. IEEE Computer Society, 106-111.

[Keet06b] Keet, C.M. Part-whole relations in Object-Role Models. *2nd International Workshop on Object-Role Modelling (ORM 2006)*, Montpellier, France, Nov 2-3, 2006. In: OTM Workshops 2006. Meersman, R., Tari, Z., Herrero., P. *et al.* (Eds.), Berlin: Springer-Verlag, LNCS **4278**, 1116-1127.

[Keet06c] Keet, C.M. *Introduction to part-whole relations: mereology, conceptual modelling and mathematical aspects*. KRDB Research Centre Technical Report KRDB06-3, Faculty of Computer Science, Free University of Bozen-Bolzano, Italy. 2 October 2006. http://www.inf.unibz.it/krdb/pub/.

[Keet06d] Keet, C.M. *Granular information retrieval from the Gene Ontology and from the Foundational Model of Anatomy with OQAFMA*. Technical Report KRDB06-1, Faculty of Computer Science, Free University of Bozen-Bolzano, 2006. http://www.inf.unibz.it/krdb/pub/. Extended abstract: Keet, C.M. Using bio-ontologies in RDBMSs for querying granular information. *Bioinformatics, Ontologies and databases (SBIOLBD 2006)*, 13-2-2006, EPFL, Lausanne, Switzerland. (http://www.meteck.org/files/SBIOLBD06_extabs.pdf).

[Keet06e] Keet, C.M. *A formal approach for using granularity with the infectious diseases domain*. Technical Report KRDB06-2, Faculty of Computer Science, Free University of Bozen-Bolzano, 2006. http://www.inf.unibz.it/krdb/pub/.

# References


[Keet06f] Keet, C.M. Enhancing biological information systems with granularity. *KnowledgeWeb PhD Symposium (KWEPSY06)*, Budva, Montenegro, 17 June 2006.

[Keet07a] Keet, C.M. Granulation with indistinguishability, equivalence or similarity. *IEEE International Conference on Granular Computing (GrC2007)*, San Francisco, November 2-4, 2007. IEEE Computer Society, 11-16.

[Keet07b] Keet, C.M. Enhancing comprehension of ontologies and conceptual models through abstractions. *10th Congress of the Italian Association for Artificial Intelligence (AI*IA 2007)*, Rome, September 10-13, 2007. Basili, R., Pazienza, M.T. (Eds.), Springer-Verlag, Lecture Notes in Artificial Intelligence **4733**, 814-822.

[Keet07c] Keet, C.M. Prospects for and issues with mapping the Object-Role Modeling language into $\mathcal{DLR}_{ifd}$. *20th International Workshop on Description Logics (DL'07)*, 8-10 June 2007, Bressanone, Italy. CEUR-WS, Vol-250, 331-338.

[Keet08a] Keet, C.M. Toward cross-granular querying over modularized ontologies. *International Workshop on Ontologies: Reasoning and Modularity (WORM08)*. 2 June 2008, Tenerife, Spain. CEUR-WS Vol-xx.

[Keet08b] Keet, C.M. A formal comparison of conceptual data modeling languages. *13th International Workshop on Exploring Modeling Methods in Systems Analysis and Design (EMMSAD'08)*. 16-17 June 2008, Montpellier, France. CEUR-WS (ISSN 1613-0073), Vol-xx. 15p.

[Keet&Artale07] Keet, C.M., Artale, A. Representing and Reasoning over a Taxonomy of Part-Whole Relations. *Applied Ontology – Special Issue on Ontological Foundations for Conceptual Models*, 2007, in print.

# References

[Keet&Kumar05] Keet, C.M., Kumar, A. (2005). Applying partitions to infectious diseases. *XIX International Congress of the European Federation for Medical Informatics (MIE2005)*, 28-31 August 2005, Geneva, Switzerland. 2005. In: Connecting Medical Informatics and bio-informatics, Engelbrecht, R., Geissbuhler, A., Lovis, C. Mihalas, G. (eds.). Amsterdam: IOS Press. pp1236-1241.

[Keet&Rodríguez07] Keet, C.M., Rodríguez, M. Toward using biomedical ontologies: trade-offs between ontology languages. *AAAI 2007 Workshop on Semantic e-Science (SeS07)*, 23 July 2007, Vancouver, Canada. AAAI 2007 Technical Report WS-07-11, 65-68.

[Keet&Roos&Marshall07] Keet, C.M., Roos, M., Marshall, M.S. A survey of requirements for automated reasoning services for bio-ontologies in OWL. *Third International Workshop OWL: Experiences and Directions (OWLED 2007)*, 6-7 June 2007, Innsbruck, Austria. CEUR-WS (ISSN 1613-0073), **Vol-258**. 10p.

[MZ06] Malinowski, E., Zimányi, E. Hierarchies in a multidimensional model: From conceptual modeling to logical representation. *Data & Knowledge Engineering*, 2006, **59**(2): 348-377.

[MBR03] Mork, P., Brinkley, J.F., Rosse, C. OQAFMA Query Agent for the Foundational Model of Anatomy: a prototype for providing flexible and efficient access to large semantic networks. *Journal of Biomedical Informatics*, 2003, 36: 501-517.

[PSZ06] Parent, C., Spaccapietra, S., Zimányi, E. Conceptual modeling for traditional and spatio-temporal applications—the MADS approach. Berlin Heidelberg: Springer Verlag. 2006.

[RM03] Rosse, C., Mejino, J.L.V.: A reference ontology for biomedical informatics: the foundational model of anatomy. *Journal of Biomedical Informatics*, 2003, **36**: 478-500.

# References

[Setal07] Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., The OBI Consortium, Leontis, N., Rocca-Serra, A.B., Ruttenberg, A., Sansone, S-A., Shah, M., Whetzel, P.L., Lewis, S. The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration. *Nature Biotechnology*, 2007, **25**(11): 1251-1255.

[Tziviskou&Keet07] Tziviskou, C., Keet, C.M. (2007). A Meta-Model for Ontologies with ORM2. *Third International Workshop on Object-Role Modelling (ORM 2007)*, Algarve, Portugal, Nov 28-30, 2007. In: *OTM Workshops 2007*. Meersman, R., Tari, Z, Herrero, P. et al (Eds.), Berlin: Springer-Verlag, Lecture Notes in Computer Science **4805**, 624-633.

[Y05] Yao, Y.Y. (2005a). Perspectives of Granular Computing. *IEEE Conference on Granular Computing (GrC2005)*, **1**: 85-90.

[1] Descriptive Ontology for Linguistic and Cognitive Engineering. http://www.loa-cnr.it/DOLCE.html.

[2] Foundational Model of Anatomy (FMA). http://fme.biostr.washington.edu:8089/FME/index.html.

[3] OBO Foundry. http://obofoundry.org.

# The TOG and OBO Foundry's approach to granulation

**A**

| | Continuant | | | | Occurrent |
|---|---|---|---|---|---|
| | Independent | | Dependent | | |
| Granularity | | | | | |
| Organ and organism | Organism (NCBI taxonomy or similar) | Anatomical entity (FMA, CARO) | Organ function (Physiology ontology, to be determined) | Phenotypic quality (PATO) | Organism-level process (GO) |
| Cell and cellular component | Cell (CL, FMA) | Cellular component (FMA,GO) | Cellular function (GO) | | Cellular process (GO) |
| Molecule | Molecule (ChEBI, SO, RnaO, PRO) | | Molecular function (GO) | | Molecular process (GO) |

**B**

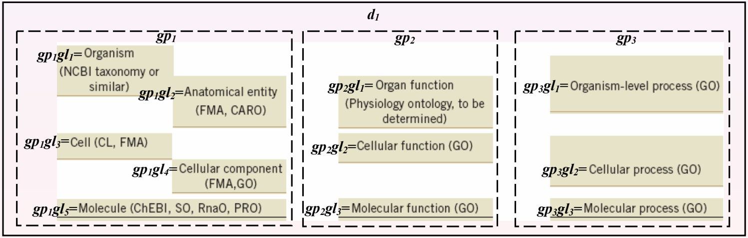



Figure: A: [Setal07] B: preliminary manual granulation with the TOG [Keet08a].

# Example

- The Foundational Model of Anatomy [RM03, 2] with some 70k types and 1.9mln relations (subsumption, parthood, containment hierarchies etc.).
- Built *manually* with Protégé, reengineered for query optimization in OQAFMA [MBR03]
- Sample queries:
  - What are the cellular components of blood?
    - Blood resides in another level as the cells it has as parts.
  - Which hormones are located in the kidney, and where in the kidney?
    - Two different perspectives, structural for kidney with its parts and functional for hormone.

# Example

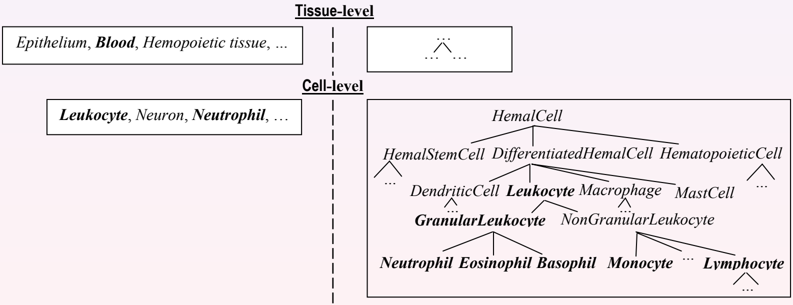
- The Foundational Model of Anatomy [RM03, 2] with some 70k types and 1.9mln relations (subsumption, parthood, containment hierarchies etc.).
- Built *manually* with Protégé, reengineered for query optimization in OQAFMA [MBR03]
- Sample queries:
  - What are the cellular components of blood?
    - Blood resides in another level as the cells it has as parts.
  - Which hormones are located in the kidney, and where in the kidney?
    - Two different perspectives, structural for kidney with its parts and functional for hormone.

# Example

- The Foundational Model of Anatomy [RM03, 2] with some 70k types and 1.9mln relations (subsumption, parthood, containment hierarchies etc.).
- Built *manually* with Protégé, reengineered for query optimization in OQAFMA [MBR03]
- Sample queries:
  - What are the cellular components of blood?
    - Blood resides in another level as the cells it has as parts.
  - Which hormones are located in the kidney, and where in the kidney?
    - Two different perspectives, structural for kidney with its parts and functional for hormone.

# Example (cont'd)

- Blood has a list of 13 direct parts. Which ones are cells, which are not? Is it a complete list of cell-parts?
- $\Rightarrow$ Combine the taxonomy & partonomy in one query.




**A. Unordered sets fill each level.**

**B. The structure of the source data is preserved.**

# Summary of the general requirements for any theory of granularity

- Use with contents at the instance level and in a format for defining a domain granularity framework at the type level;
- The logic-based representation has to permit both set theoretical and mereological granularity;
- Accommodate both the quantitative and qualitative aspects of granularity (or: arbitrary scale and non-scale-dependent granularity);
- There have to be at least two levels within a perspective, else there is no granularity;

# Summary of the general requirements for any theory of granularity

- A higher level simplifies, makes indistinguishable, the finer-grained details that are indistinguishable at that higher level;
- The entities (/types) in a granular level have at least one aspect in common, which is a criterion by which to granulate the data, information, or knowledge;

back to main

# Foundational semantics of granularity

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

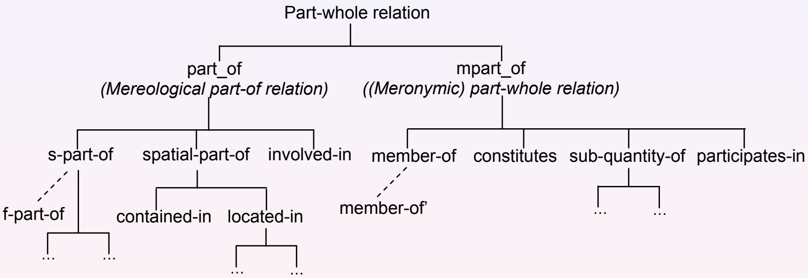
| Branching point | Distinguishing feature |
|---|---|
| sG – nG | scale – non-scale (or, roughly: quantitative – qualitative) |
| sgG – saG | grain size – aggregation (or: scale *on* entity – scale *of* entity) |
| sgrG – sgpG | resolution – size of the entity |
| saoG – samG | overlay aggregated – 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 |

# Part-whole relations

- Ground Mereology [Keet06c]
  - with *part_of*, which is reflexive, antisymmetric, and transitive
  - *ppart_of*, irreflexive, asymmetric, transitive
  - other relations one can define with it, such as *overlap* and *overcross*
- Other non-mereological part-whole relations that can be used to granulate
- Part-whole relations formalised and structured in a taxonomy

[Keet06b, Keet&Artale07]

# Part-whole relations




$$\forall x, y(\mathit{involved\_in}(x, y) \triangleq \mathit{part\_of}(x, y) \wedge PD(x) \wedge PD(y))$$

back to main

# 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)**

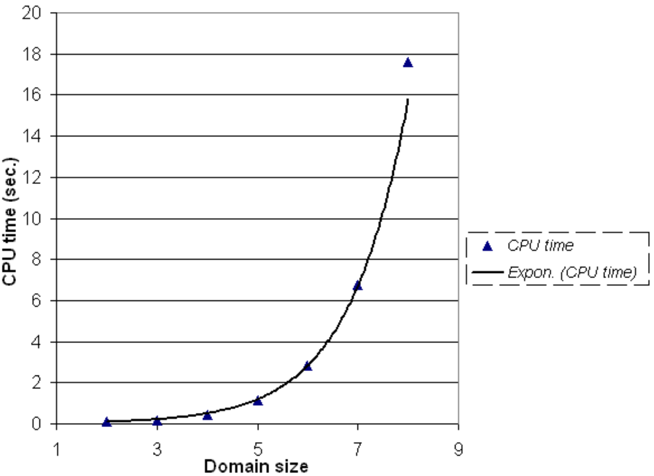
*If two levels in different perspectives overcross, then their perspectives overcross:*
$\forall x_1, x_2, y_1, y_2(overcross(x_1, x_2) \wedge GL(x_1) \wedge GL(x_2) \wedge GP(y_1) \wedge$
$GP(y_2) \wedge R_E(x_1, y_1) \wedge R_E(x_2, y_2) \rightarrow overcross(y_1, y_2))$.

back to main

# Tested with Mace4 model searcher

- *Known* to be satisfiable if indistinguishability is made a primitive relation, acyclicity hard-coded, no quantification over universals
- With that input, *known* that there are models up to at least domain size 8
- So, is the TOG within the decidable fragment of FOL, and what is the complexity of the language needed to represent it?

# Mace4 model searcher statistics



# "Interesting" constructors used in the TOG

- *RL* is acyclic. Complexity of DL languages: the only DL language that has a constructor for it, $\mathcal{DLR}_\mu$ [CGL99], is ExpTime like its base language $\mathcal{DLR}$ and sister languages $\mathcal{DLR}_{ifd}$ and $\mathcal{DLR}_{reg}$ [CGL98, CDGLR03].
- *GP* is uniquely identified by the combination of its *C* and *TG*, which means it is, in ER terminology, a weak entity type: **id** constructor of $\mathcal{DLR}_{ifd}$ [CGL01].
- Others: qualified number restrictions ($RE^-$, *CP*, and *conv*), a ternary relation (*conv*), and transitivity of relations (*RL*, *RE*, and mereological parthood).

back to main

# A comparison

| **Language** $\Rightarrow$ **Feature** $\Downarrow$ | OWL **Lite** | OWL **DL** | OWL **v1.1** | *DL-Lite* $\mathcal{F}$ | *DL-Lite* $\mathcal{R}$ | *DL-Lite* $\mathcal{A}$ | $\mathcal{DLR}$ *ifd* | $\mathcal{DLR}$ $\mu$ | $\mathcal{DLR}$ *reg* |
|---|---|---|---|---|---|---|---|---|---|
| Role hierarchy | + | + | + | - | + | + | + | + | + |
| N-ary roles (where $n \geq 2$) | - | - | - | ± | ± | ± | + | + | + |
| Role concatenation | - | - | + | - | - | - | - | - | + |
| Role acyclicity | - | - | - | - | - | - | - | + | - |
| Symmetry | + | + | + | - | + | + | - | - | - |
| Role values | - | - | - | - | - | + | - | - | - |
| Qualified number restrictions | - | - | + | - | - | - | + | + | + |
| One-of, enumerated classes | - | + | + | - | - | - | - | - | - |
| Functional dependency | + | + | + | + | - | + | + | - | + |
| Covering constraint over concepts | - | + | + | - | - | - | + | + | + |
| Complement of concepts | - | + | + | + | + | + | + | + | + |
| Complement of roles | - | - | + | + | + | + | + | + | + |
| Concept identification | - | - | - | - | - | - | + | - | - |
| Range typing | - | + | + | - | + | + | + | + | + |
| Reflexivity * | - | - | + | - | - | - | - | + | + |
| Antisymmetry * | - | - | - | - | - | - | - | - | - |
| Transitivity * ‡ | + | + | + | - | - | - | - | + | + |
| Asymmetry ‡ | + | + | + | - | + | + | - | ± | - |
| Irreflexivity ‡ | - | - | + | - | - | - | - | + | - |

Table: Differences between DL-based ontology and conceptual modelling languages (after [Keet&Rodríguez07]).

# What to do with it?

- For its own sake, ‘paper exercise’ for developing perspectives and levels, stay in FOL and then alike a global view with LAV
- Orthogonally positioned to conceptual data model or integrated in a conceptual data model
- As meta-layer for conceptual data modelling languages or ontology languages
- Advanced querying & reasoning
- ...

back to main