On the unification of conceptual data modelling languages

C. Maria Keet[1]

Department of Computer Science, University of Cape Town, South Africa,
mkeet@cs.uct.ac.za

Seminar at the Poznan University of Technology, Poland, June 16, 2015

[1]Joint work with Pablo Rubén Fillottrani, Universidad Nacional del Sur, Bahía Blanca, Argentina

Outline

1. Motivation
2. Unification approach
   - Metamodel
   - Transformations and intermodel assertions
3. Quantitative analysis
4. Conclusions

Outline

1. Motivation
2. Unification approach
   - Metamodel
   - Transformations and intermodel assertions
3. Quantitative analysis
4. Conclusions

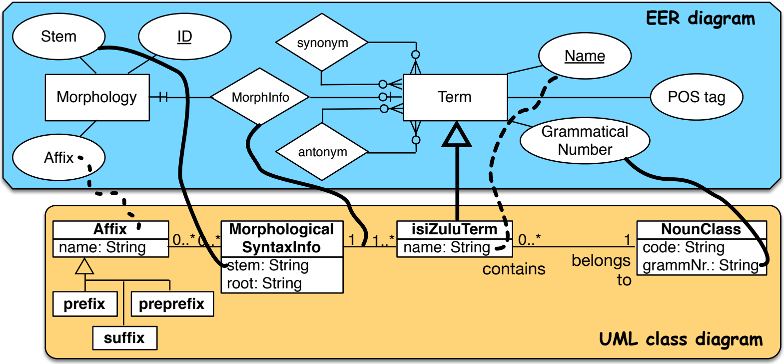
Context

- Bilateral project "ontology-driven unification of conceptual data modelling languages" (mid 2012 - mid 2015)[2], funded by SA Dept. of Sci & Tech and AR's MINCyT

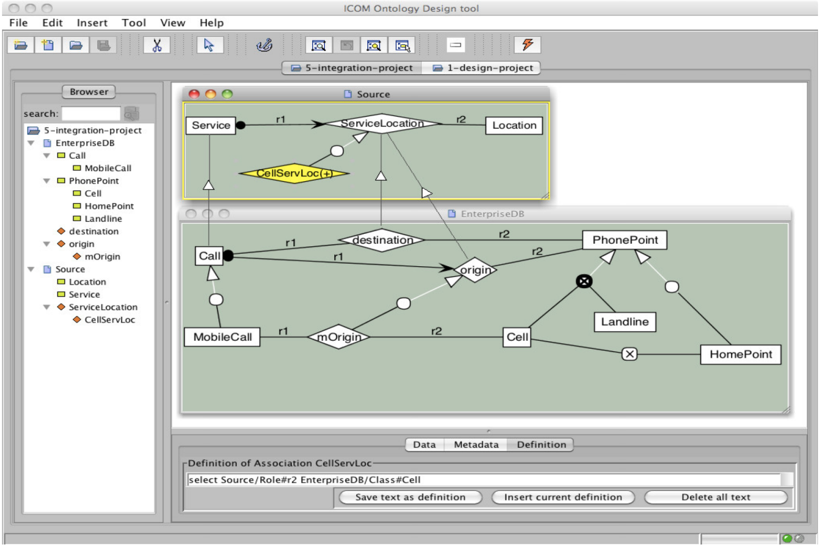
- Conceptual data modelling for complex system development and information integration
- Languages for conceptual modelling: UML Class Diagram, ER and EER, ORM and ORM2
- Develop formal basis for model linking and integration, tools and techniques

[2] Project page: http://www.meteck.org/SAAR.html

Example: isiZulu termbank (simplified)



Example: ICOM (Franconi and others)



Previous work

- Inter-model assertions between models in the same language [Atzeni et al.(2008), Fillottrani et al.(2012)]
- Inter-model assertions between models in different languages, but subset only
  [Atzeni et al.(2012), Boyd and McBrien(2005), Venable and Grundy(1995), Zhu et al.(2004)]
- Limited model transformations
  [Atzeni et al.(2012), Boyd and McBrien(2005)]
- Limited or no automated reasoning, verification
  [Calvanese et al.(1999), Fillottrani et al.(2012), Keet(2009)]

Outline

Overview

- All static, structural elements of main CDM languages
- First ontological, then logical, finally implement
- Develop *unifying* and *ontology-driven* metamodel, then formalise it
- Mechanism for inter-model assertions and transformations
- Quantitative evaluation to prioritise rule specification
  - Language profile specification (tractable languages?)
- Implement, and look at modularisation (ongoing)

Overview

- All static, structural elements of main CDM languages
- First ontological, then logical, finally implement
- Develop *unifying* and *ontology-driven* metamodel, then formalise it
- Mechanism for inter-model assertions and transformations
- Quantitative evaluation to prioritise rule specification
  - Language profile specification (tractable languages!)
- Implement, and look at modularisation (ongoing)

Metamodel: overview

- Captures all structural elements in the selected languages[3] [4]
- Captures also their relations and constraints
- Describes the rules in which they may be combined
- The metamodel is designed in UML Class Diagram, and formalized in FOL (precision) and OWL (practical usability)[5]

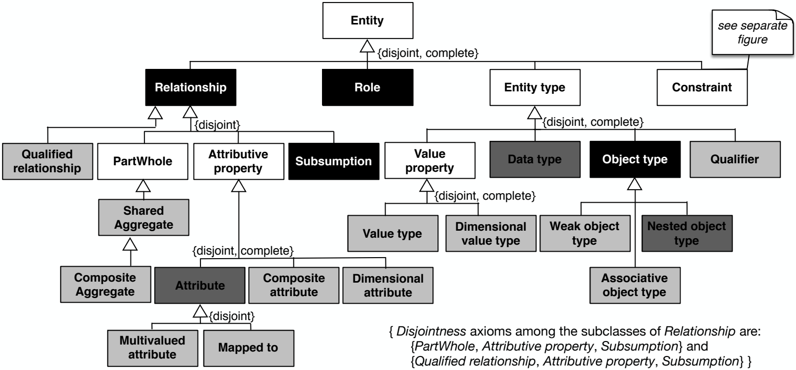
---

[3] Keet, C.M., Fillottrani, P.R. Toward an ontology-driven unifying metamodel for UML Class Diagrams, EER, and ORM2. ER'13. W. Ng, V.C. Storey, and J. Trujillo (Eds.). Springer LNCS vol. 8217, 313-326.

[4] Keet, C.M., Fillottrani, P.R. Structural entities of an ontology-driven unifying metamodel for UML, EER, and ORM2. MEDI'13. A. Cuzzocrea and S. Maabout (Eds.). Springer LNCS vol. 8216, 188-199.

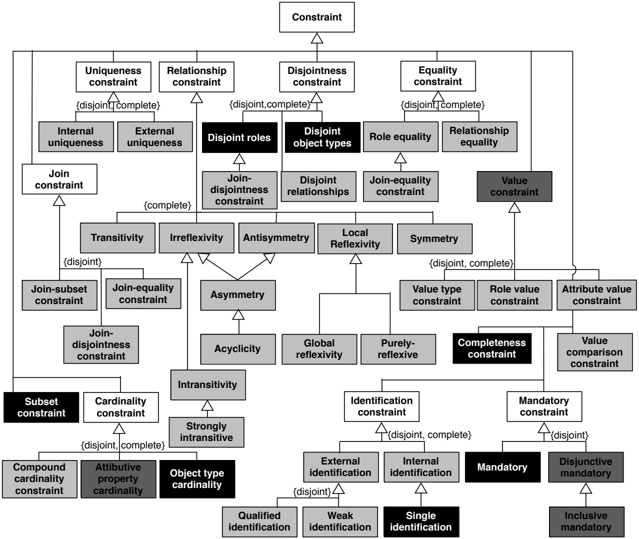
[5] Fillottrani, P.R., Keet, C.M.. *KF metamodel formalization*. Technical Report, Arxiv.org `http://arxiv.org/abs/1412.6545`. Dec 19, 2014. 26p.

Static entities

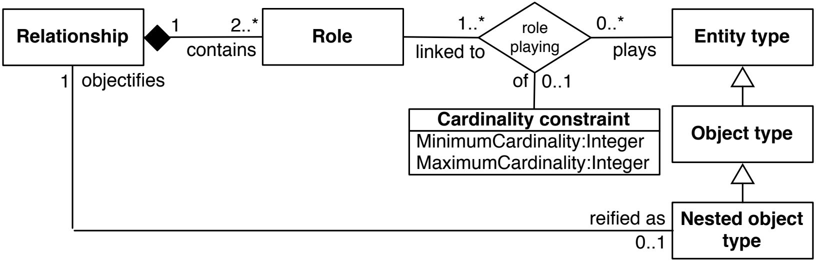


{ *Disjointness* axioms among the subclasses of *Relationship* are:
{*PartWhole*, *Attributive property*, *Subsumption*} and
{*Qualified relationship*, *Attributive property*, *Subsumption*} }

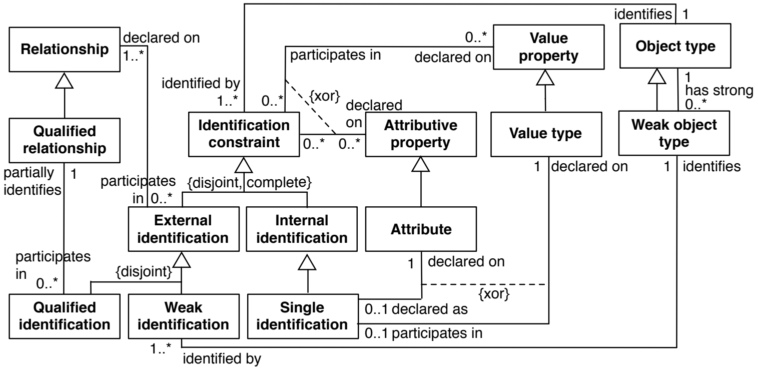
Constraints



Selection of constraints between them (1/2)



Selection of constraints between them (2/2)



{ A *Weak identification* is a combination of one or more *Attributive property* of the *Weak object type* it *identifies* together with the *Identification constraint* of the *Object type* it has a *Relationship* with and this *Object type* is disjoint with the *Weak object type*. }

{ The *Single identification* has a *Mandatory* constraint on the participating *Role* and the *Relationship* that *Role* is *contained in* has a 1:1 *Cardinality constraint* declared on it. }

{ *Qualified identification* and *External identification* are *declared on* only *Attributive property*. }

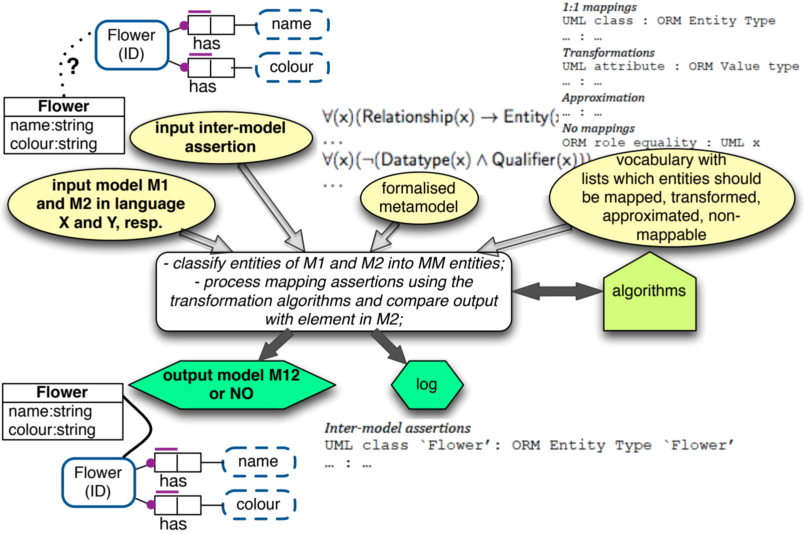
{ A *Qualified relationship participates in* a *Qualified identification* only if the *Cardinality constraint* is 1. }

Transformation Rules and Inter-model assertions[6]

- Process for linking and translating models
- Based on different kinds of rules: mappings, transformations, approximations
- Together with the (formalised) metamodel, it can be used to verify inter-model assertions

[6] Fillottrani, P.R., Keet, C.M. Conceptual Model Interoperability: a Metamodel-driven Approach. RuleML'14, A. Bikakis et al. (Eds.). Springer LNCS vol. 8620, 52-66.

Approach (inter-model assertions)



1:1 mapping rules and the metamodel (selection)

(R1) $\texttt{Association} \xrightarrow{\text{UML to MM}} \text{Relationship}$

in:
$\texttt{Association}(\texttt{AssociationEnd} : \texttt{Class}, \texttt{AssociationEnd} : \texttt{Class})$

out: $\texttt{AssociationEnd} \rightarrow$ Role // *i.e., using (Ro1)*

out: $\texttt{Association} \rightarrow$ Relationship

out: $\texttt{Class} \rightarrow$ Object Type // *i.e., using (O1)*

out: Relationship(Role:Object type, Role:Object Type)

(1R) $\text{Relationship} \xrightarrow{\text{MM to UML}} \texttt{Association}$

in: Relationship(Role:Object type, Role:Object Type)

out: Role $\rightarrow \texttt{AssociationEnd}$ // *i.e., using (1Ro)*

out: Relationship $\rightarrow \texttt{Association}$

out: Object Type $\rightarrow \texttt{Class}$ // *i.e., using (1O)*

out:
$\texttt{Association}(\texttt{AssociationEnd} : \texttt{Class}, \texttt{AssociationEnd} : \texttt{Class})$

Generating and mapping

GenOT Class $\xrightarrow{\text{UML to ORM}}$ Entity type

in: $\mathtt{C}$

out: (O1)

out: (2O) // *i.e., an ORM* $\mathtt{EntityType}$ *named* $\mathtt{C}$

MapR Association $\xrightarrow{\text{UML to ER}}$ Relationship

in: $\mathtt{A}(\mathtt{ae}_1 : \mathtt{C}_1, \mathtt{ae}_2 : \mathtt{C}_2)$

out: (R1)

out: (3R)

out: match pattern out(3R) with $\mathtt{R}(\mathtt{rc}_1 : \mathtt{E}_1, \mathtt{rc}_2 : \mathtt{E}_2)$

Formalised metamodel (section), highlighted for step 2

$\forall(x, y)(\mathtt{Contains}(x, y) \rightarrow \mathtt{Relationship}(x) \land \mathtt{Role}(y))$

$\forall(x)\exists^{\geq 2}y(\mathtt{Contains}(x, y))$

$\forall(x)(\mathtt{Role}(x) \rightarrow \exists(y)(\mathtt{Contains}(y, x)))$

$\forall(x, y, z)(\mathtt{Contains}(x, y) \land \mathtt{Contains}(z, y) \rightarrow (x = z))$

$\forall(x, y, z)(\mathtt{RolePlaying}(x, y, z) \rightarrow \mathtt{Role}(x) \land \mathtt{CardinalityConstraint}(y) \land \mathtt{EntityType}(z))$

$\forall(x)(\mathtt{Role}(x) \rightarrow \exists(y, z)(\mathtt{RolePlaying}(x, y, z)))$

$\forall(x, y, z, v, w)(\mathtt{RolePlaying}(x, y, z) \land \mathtt{RolePlaying}(x, v, w) \rightarrow (y = v) \land (z = w))$

$\forall(x, y, z, v, w)(\mathtt{RolePlaying}(x, y, z) \land \mathtt{RolePlaying}(v, y, w) \rightarrow (x = v) \land (z = w))$

$\forall(x)(\mathtt{CardinalityConstraint}(x) \rightarrow \exists(y)(\mathtt{MinimumCardinality}(x, y) \land \mathtt{Integer}(y)))$

$\forall(x)(\mathtt{CardinalityConstraint}(x) \rightarrow \exists(y)(\mathtt{MaximumCardinality}(x, y) \land \mathtt{Integer}(y)))$

$\forall(x, y)(\mathtt{Identifies}(x, y) \rightarrow (\mathtt{IdentificationConstraint}(x) \land \mathtt{ObjectType}(y)))$

$\forall(x)(\mathtt{IdentificationConstraint}(x) \rightarrow \exists(y)(\mathtt{Identifies}(x, y)))$

$\forall(x, y, z)((\mathtt{Identifies}(x, y) \land \mathtt{Identifies}(x, z)) \rightarrow (y = z))$

$\forall(x)(\mathtt{ObjectType}(x) \rightarrow \exists(y)(\mathtt{Identifies}(y, x)))$

$\forall(x, y, z)((\mathtt{DeclaredOn}(x, y) \land \mathtt{DeclaredOn}(x, z) \land \mathtt{IdentificationConstraint}(x) \land (\neg(y =$
$(\mathtt{ValueProperty}(y) \leftrightarrow \neg\mathtt{AttributiveProperty}(z)))$

$\forall(x)(\mathtt{IdentificationConstraint}(x) \rightarrow \exists(y)(\mathtt{DeclaredOn}(x, y)))$

$\forall(x, y)((\mathtt{DeclaredOn}(x, y) \land \mathtt{SingleIdentification}(x)) \rightarrow (\mathtt{Attribute}(y) \lor \mathtt{ValueType}(y))$

$\forall(x)(\mathtt{SingleIdentification}(x) \rightarrow \exists(y)(\mathtt{DeclaredOn}(x, y))$

$\forall(x, y, z)((\mathtt{SingleIdentification}(x) \land \mathtt{DeclaredOn}(x, y) \land \mathtt{DeclaredOn}(x, z)) \rightarrow (y = z))$

Highlighted section for step 3

$\forall(x,y)(\texttt{Contains}(x,y) \rightarrow \texttt{Relationship}(x) \wedge \texttt{Role}(y))$
$\forall(x)\exists^{\geq 2}y(\texttt{Contains}(x,y))$
$\forall(x)(\texttt{Role}(x) \rightarrow \exists(y)(\texttt{Contains}(y,x)))$
$\forall(x,y,z)(\texttt{Contains}(x,y) \wedge \texttt{Contains}(z,y) \rightarrow (x = z))$
$\forall(x,y,z)(\texttt{RolePlaying}(x,y,z) \rightarrow \texttt{Role}(x) \wedge \texttt{CardinalityConstraint}(y) \wedge \texttt{EntityType}(z))$
$\forall(x)(\texttt{Role}(x) \rightarrow \exists(y,z)(\texttt{RolePlaying}(x,y,z)))$
$\forall(x,y,z,v,w)(\texttt{RolePlaying}(x,y,z) \wedge \texttt{RolePlaying}(x,v,w) \rightarrow (y = v) \wedge (z = w))$
$\forall(x,y,z,v,w)(\texttt{RolePlaying}(x,y,z) \wedge \texttt{RolePlaying}(v,y,w) \rightarrow (x = v) \wedge (z = w))$
$\forall(x)(\texttt{CardinalityConstraint}(x) \rightarrow \exists(y)(\texttt{MinimumCardinality}(x,y) \wedge \texttt{Integer}(y)))$
$\forall(x)(\texttt{CardinalityConstraint}(x) \rightarrow \exists(y)(\texttt{MaximumCardinality}(x,y) \wedge \texttt{Integer}(y)))$
$\forall(x,y)(\texttt{Identifies}(x,y) \rightarrow (\texttt{IdentificationConstraint}(x) \wedge \texttt{ObjectType}(y)))$
$\forall(x)(\texttt{IdentificationConstraint}(x) \rightarrow \exists(y)(\texttt{Identifies}(x,y)))$
$\forall(x,y,z)((\texttt{Identifies}(x,y) \wedge \texttt{Identifies}(x,z)) \rightarrow (y = z))$
$\forall(x)(\texttt{ObjectType}(x) \rightarrow \exists(y)(\texttt{Identifies}(y,x)))$
$\forall(x,y,z)((\texttt{DeclaredOn}(x,y) \wedge \texttt{DeclaredOn}(x,z) \wedge \texttt{IdentificationConstraint}(x) \wedge (\neg(y =$
$(\texttt{ValueProperty}(y) \leftrightarrow \neg\texttt{AttributiveProperty}(z)))$
$\forall(x)(\texttt{IdentificationConstraint}(x) \rightarrow \exists(y)(\texttt{DeclaredOn}(x,y)))$
$\forall(x,y)((\texttt{DeclaredOn}(x,y) \wedge \texttt{SingleIdentification}(x)) \rightarrow (\texttt{Attribute}(y) \vee \texttt{ValueType}(y))$
$\forall(x)(\texttt{SingleIdentification}(x) \rightarrow \exists(y)(\texttt{DeclaredOn}(x,y))$
$\forall(x,y,z)((\texttt{SingleIdentification}(x) \wedge \texttt{DeclaredOn}(x,y) \wedge \texttt{DeclaredOn}(x,z)) \rightarrow (y = z))$

Formalised metamodel (section), highlighted for step 5

$$\forall(x,y)(\texttt{Contains}(x,y) \rightarrow \texttt{Relationship}(x) \wedge \texttt{Role}(y))$$
$$\forall(x)\exists^{\geq 2}y(\texttt{Contains}(x,y))$$
$$\forall(x)(\texttt{Role}(x) \rightarrow \exists(y)(\texttt{Contains}(y,x)))$$
$$\forall(x,y,z)(\texttt{Contains}(x,y) \wedge \texttt{Contains}(z,y) \rightarrow (x = z))$$
$$\forall(x,y,z)(\texttt{RolePlaying}(x,y,z) \rightarrow \texttt{Role}(x) \wedge \texttt{CardinalityConstraint}(y) \wedge \texttt{EntityType}(z))$$
$$\forall(x)(\texttt{Role}(x) \rightarrow \exists(y,z)(\texttt{RolePlaying}(x,y,z)))$$
$$\forall(x,y,z,v,w)(\texttt{RolePlaying}(x,y,z) \wedge \texttt{RolePlaying}(x,v,w) \rightarrow (y = v) \wedge (z = w))$$
$$\forall(x,y,z,v,w)(\texttt{RolePlaying}(x,y,z) \wedge \texttt{RolePlaying}(v,y,w) \rightarrow (x = v) \wedge (z = w))$$
$$\forall(x)(\texttt{CardinalityConstraint}(x) \rightarrow \exists(y)(\texttt{MinimumCardinality}(x,y) \wedge \texttt{Integer}(y)))$$
$$\forall(x)(\texttt{CardinalityConstraint}(x) \rightarrow \exists(y)(\texttt{MaximumCardinality}(x,y) \wedge \texttt{Integer}(y)))$$
$$\forall(x,y)(\texttt{Identifies}(x,y) \rightarrow (\texttt{IdentificationConstraint}(x) \wedge \texttt{ObjectType}(y)))$$
$$\forall(x)(\texttt{IdentificationConstraint}(x) \rightarrow \exists(y)(\texttt{Identifies}(x,y)))$$
$$\forall(x,y,z)((\texttt{Identifies}(x,y) \wedge \texttt{Identifies}(x,z)) \rightarrow (y = z))$$
$$\forall(x)(\texttt{ObjectType}(x) \rightarrow \exists(y)(\texttt{Identifies}(y,x)))$$
$$\forall(x,y,z)((\texttt{DeclaredOn}(x,y) \wedge \texttt{DeclaredOn}(x,z) \wedge \texttt{IdentificationConstraint}(x) \wedge (\neg(y =$$
$$(\texttt{ValueProperty}(y) \leftrightarrow \neg\texttt{AttributiveProperty}(z)))$$
$$\forall(x)(\texttt{IdentificationConstraint}(x) \rightarrow \exists(y)(\texttt{DeclaredOn}(x,y)))$$
$$\forall(x,y)((\texttt{DeclaredOn}(x,y) \wedge \texttt{SingleIdentification}(x)) \rightarrow (\texttt{Attribute}(y) \vee \texttt{ValueType}(y))$$
$$\forall(x)(\texttt{SingleIdentification}(x) \rightarrow \exists(y)(\texttt{DeclaredOn}(x,y))$$
$$\forall(x,y,z)((\texttt{SingleIdentification}(x) \wedge \texttt{DeclaredOn}(x,y) \wedge \texttt{DeclaredOn}(x,z)) \rightarrow (y = z))$$

Outline

1. Motivation
2. Unification approach
   - Metamodel
   - Transformations and intermodel assertions
3. Quantitative analysis
4. Conclusions

Conceptual modelling in practice – an analysis[7]

- Few elements belong to all three language families

⇒ Is it worth trying to link or integrate or translate their models?

- Collected available models on each language, and studied the usage of metamodel elements on them (approx. 35 on each language)
  - Only 64% of the entities are the kind of entities that appear in all three language families
  - When more features are available in a language, they are used in the models (though some very few times)
  - Specification of a feature-based 'characteristic profile' for each family

[7] Keet, C.M., Fillottrani, P.R. An analysis and characterisation of publicly available conceptual models. ER'15. Springer LNCS. (accepted)

Conceptual modelling in practice – an analysis[7]

- Few elements belong to all three language families

⇒ Is it worth trying to link or integrate or translate their models?

- Collected available models on each language, and studied the usage of metamodel elements on them (approx. 35 on each language)
  - Only 64% of the entities are the kind of entities that appear in all three language families
  - When more features are available in a language, they are used in the models (though some very few times)
  - Specification of a feature-based 'characteristic profile' for each family

[7] Keet, C.M., Fillottrani, P.R. An analysis and characterisation of publicly available conceptual models. ER'15. Springer LNCS. (accepted)

Table: Prevalence of particular entity in the models, as percent of total number of entities for that family, aggregated by model family and rounded off to one decimal. OT: Object type; VT: Value type; Rel.: Relationship; Int. Unique.: Internal uniqueness constraint; ID: Identifier.

Top-5		
UML CD	ORM/2	(E)ER
Attribute (31.2%)	OT cardinality (29.0%)	Attribute (39.5%)
OT (21.2%)	OT (14.5%)	OT cardinality (22.1%)
OT cardinality (17.5%)	2-ary Rel. (14.4%)	2-ary Rel. (11.6%)
2-ary Rel. (12.4%)	Int. unique. (13.1%)	OT (11.5%)
OT subsumption (9.6%)	VT (10.4%)	single ID (7.7%)

Ratios of entities aggregated by family and combined

Ratio	UML	ORM/2	(E)ER	comb.
model size:total entities	0.8	0.5	0.7	0.6
Attribute or Value type:Object type	1.5	0.7	3.5	1.7
binaries:n-aries	180.5	12.4	20.9	20.4
Subsumption(class):Object type	0.5	0.3	0.2	0.3
Relationship (non isa):Object type	0.8	1.1	1.1	1.0
Object type cardinality: other constraint	7.4	1.2	2.2	1.8
Single identification:other ID	–	17.3	5.4	8.4
role:relationship naming	4.3	(readings, mostly)	0.1	N/A

Logic foundation for profiles

- Common features: Object type, Relationship, Object type cardinality, Subsumption (object type), Single identification, Disjoint and Complete object types.

⇒ Seems to fit some tractable language; which one(s)?

- Avail of Description Logic languages to gain insight in language and computational complexity
- Common core that covers ±87%; language-specific profiles[8]
- There is no DL that matches precisely, but a PTIME language is feasible—$\mathcal{ALNI}$ for the Core Profile
- Good match is $\mathcal{CFDI}_{nc}^{\forall-}$ (PTIME), with n-aries, identifiers[9]

[8] Fillottrani, P.R., Keet, C.M. Evidence-based Languages for Conceptual Data Modelling Profiles. ADBIS'15. Springer LNCS. Poitiers, France, Sept 8-11, 2015. (accepted)

[9] Fillottrani, P.R., Keet, C.M., Toman, D. Polynomial encoding of ORM conceptual models in $\mathcal{CFDI}_{nc}^{\forall-}$. DL'15, CEUR-WS vol. 1350, 401-414.

Logic foundation for profiles

- Common features: Object type, Relationship, Object type cardinality, Subsumption (object type), Single identification, Disjoint and Complete object types.

$\Rightarrow$ Seems to fit some tractable language; which one(s)?

- Avail of Description Logic languages to gain insight in language and computational complexity
- Common core that covers $\pm$87%; language-specific profiles[8]
- There is no DL that matches precisely, but a PTIME language is feasible—$\mathcal{ALNI}$ for the Core Profile
- Good match is $\mathcal{CFDI}_{nc}^{\forall -}$ (PTIME), with n-aries, identifiers[9]

---

[8] Fillottrani, P.R., Keet, C.M. Evidence-based Languages for Conceptual Data Modelling Profiles. ADBIS'15. Springer LNCS. Poitiers, France, Sept 8-11, 2015. (accepted)

[9] Fillottrani, P.R., Keet, C.M., Toman, D. Polynomial encoding of ORM conceptual models in $\mathcal{CFDI}_{nc}^{\forall -}$. DL'15, CEUR-WS vol. 1350, 401-414.

Outline

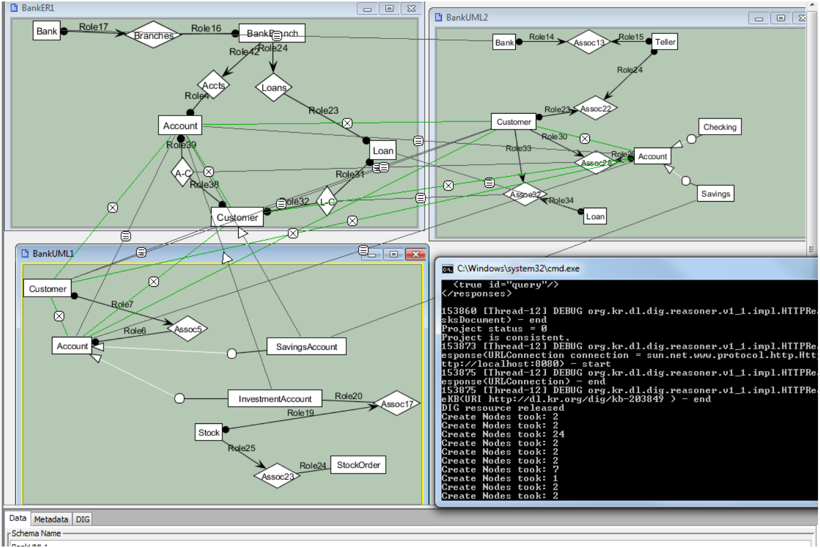
Conclusions

- Unifying ontology-driven metamodel
- Inter-model assertions and model transformation approaches with basic set of rules (1:1, transformations, and approximations)
- Quantitative analysis on feature usages
- Profile characterisation

Ongoing and future work

- Integrate these results into design tools
- ‘Scalability’ of graphical representation and inferences?
- Modularisation

Example: ICOM



References I

Paolo Atzeni, Paolo Cappellari, Riccardo Torlone, Philip A. Bernstein, and Giorgio Gianforme.
Model-independent schema translation.
VLDB Journal, 17(6):1347–1370, 2008.

Paolo Atzeni, Giorgio Gianforme, and Paolo Cappellari.
Data model descriptions and translation signatures in a multi-model framework.
AMAI 63:1–29, 2012.

Michael Boyd and Peter McBrien.
Comparing and transforming between data models via an intermediate hypergraph data model.
Journal on Data Semantics, IV:69–109, 2005.

D. Calvanese, M. Lenzerini, and D. Nardi.
Unifying class-based representation formalisms.
Journal of Artificial Intelligence Research, 11:199–240, 1999.

Pablo R. Fillottrani, Enrico Franconi, and Sergio Tessaris.
The ICOM 3.0 intelligent conceptual modelling tool and methodology.
Semantic Web Journal, 3(3):293–306, 2012.

C. Maria Keet.
Positionalism of relations and its consequences for fact-oriented modelling.
In R. Meersman, P. Herrero, and Dillon T., editors, OTM Workshops, International Workshop on Fact-Oriented Modeling (ORM'09), volume 5872 of LNCS, pages 735–744. Springer, 2009. Vilamoura, Portugal, November 4-6, 2009.

References II

J.R. Venable and J.C. Grundy.
Integrating and supporting Entity Relationship and Object Role Models.
In Mike P. Papazoglou, editor, Proceedings of the 14th International Conference on Object-Oriented and Entity-Relationship Modelling (ER'95), volume 1021 of LNCS, pages 318–328. Springer, 1995.
Gold Coast, Australia, December 12-15, 1995.

N. Zhu, J.C. Grundy, and J.G. Hosking.
Pounamu: a metatool for multi-view visual language environment construction.
In IEEE Conf. on Visual Languages and Human-Centric Computing 2004, 2004.

Thank you!