Towards Approximate Model Transformations

Javier Troya, Manuel Wimmer, Loli Burgueño, Antonio Vallecillo

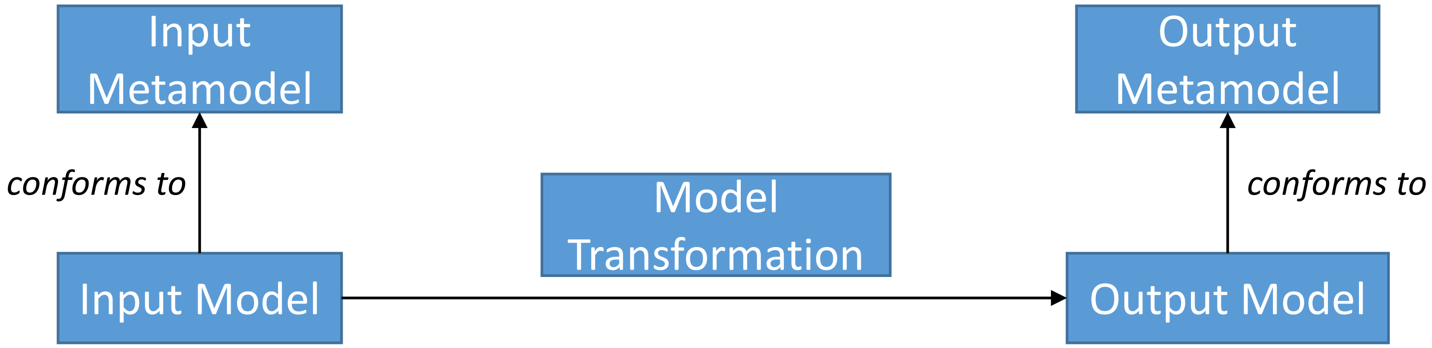
Third Workshop on the Analysis of Model Transformations
AMT'14 Valencia, Spain 29.09.14

Business Informatics Group
Institute of Software Technology and
Interactive Systems
Vienna University of Technology

MODELS 2014
VALENCIA, SPAIN

GISUM/Atenea Research Group
ETS Ingeniería Informática
Unviersidad de Málaga

Model Transformations



- Main focus is on the correct implementation of model transformations

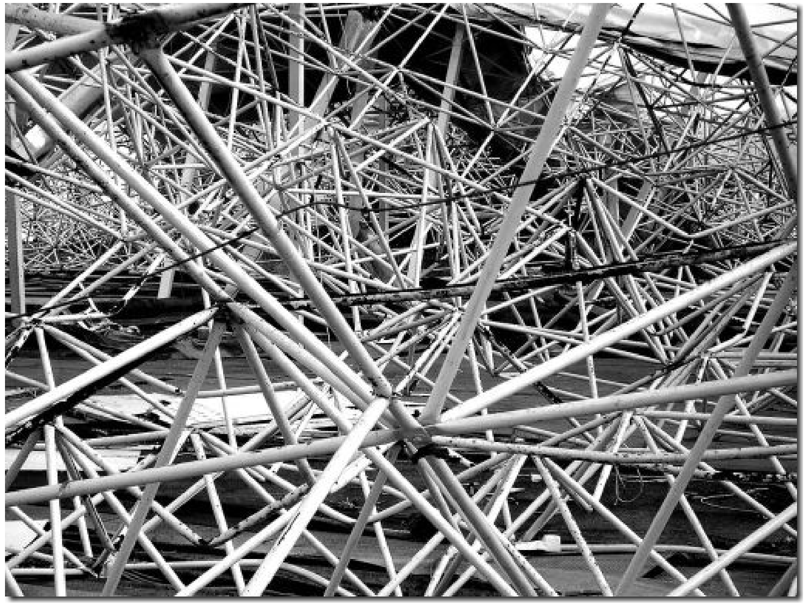
- We need to consider other (non-functional) aspects
  - Performance
  - Scalability
  - Usability
  - Maintainability
  - …

Model Transformations

- In practice, we need to
  - Handle models with millions of instances
  - Transform these models in reasonable time




- Furthermore, although specified at a high level of abstraction, model transformations are becoming very complex
  - Complex relationships among elements



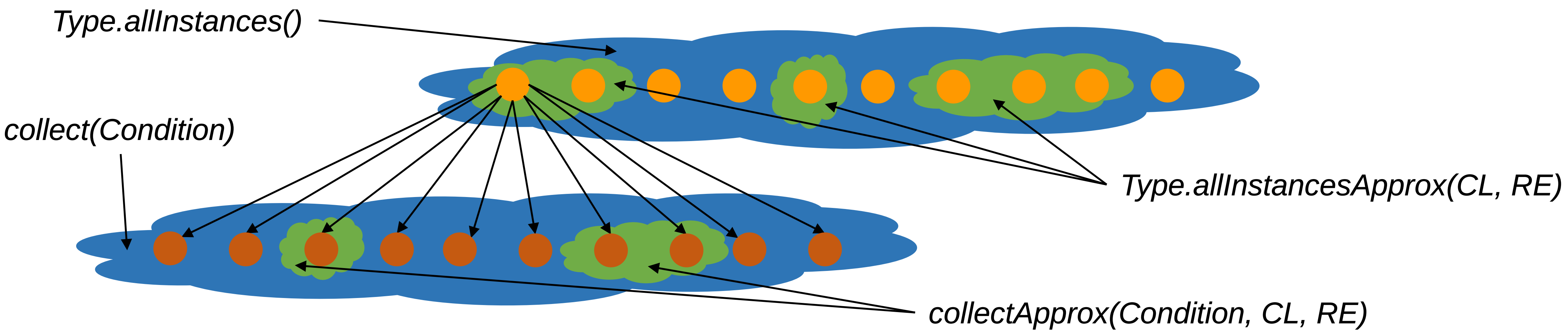
Performance improvement in MTs

- Our primary goal: **improve** the **performance** in MTs
- Current approaches for improving performance
  - Incremental execution *Jouault, F., Tisi, M.: **Towards Incremental Execution of ATL Transformations**. In: Proc. of ICMT. pp. 123–137 (2010)*
  - Lazy execution *Tisi, M., Perez, S.M., Jouault, F., Cabot, J.: **Lazy execution of model-to-model transformations**. In: Proc. of MODELS. pp. 32–46 (2011)*
  - Parallel execution *Burgueño, L., Troya, J., Wimmer, M., Vallecillo, A.: **On the Concurrent Execution of Model Transformations with Linda**. In: Proc. of Big MDE @ STAF 2013*
- Our approach for improving performance: **Approximate Model Transformations** (AMTs)
  - t1 approximates another transformation t2 if t1 is equivalent to t2 up to a certain error margin **
  - Weaken the need to produce *exact* models
  - Trade accuracy for performance
  - Produce *approximate* models
    - Accurate enough to provide meaningful and useful results
    - Alleviating the need for the MT to generate fully correct models
    - They do not rely on previous transformation runs
    - May be combined with other orthogonal approaches

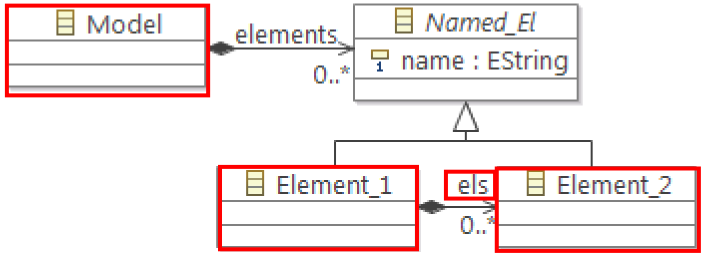
***Lúcio, L., Amrani, M., Dingel, J., Lambers, L., Salay, R., Selim, G., Syriani, E., Wimmer, M.: **Model Transformation Intents and Their Properties**. SoSyM pp. 1–35 (2014)*

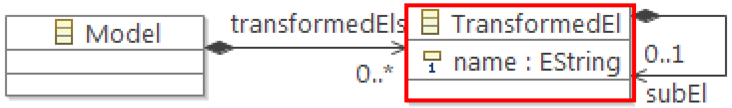
Initial Ideas for AMTs

- Redefine common operators OCL uses for collections
  - *allInstances*, *collect*, *select*… are now *allInstancesApprox*, *collectApprox*, *selectApprox*
  - We add two additional input parameters: confidence level (CL) and relative error (RE)



  - Applicable to declarative rules (e.g. matched rules in ATL apply an implicit *allInstances*)
  - Applicable in the imperative part
    - Loops such as *for (element in Type.property)*



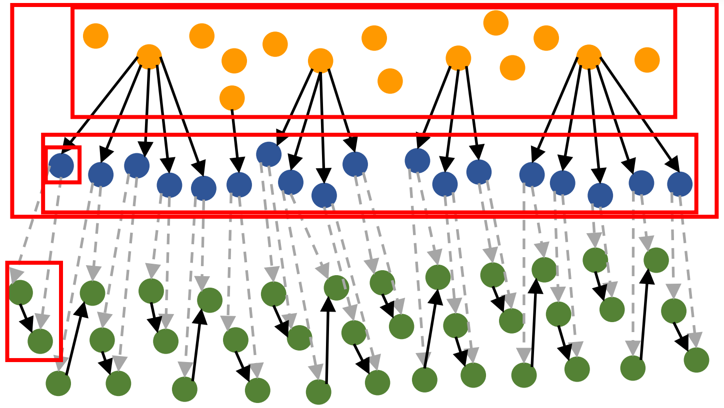


Initial Ideas

```
rule Model2Model{
   from inM : MM!Model
   to    outM : MM1!Model(
      transformedEls <- inM.elements -> select (e | e.oclIsTypeOf(MM!Element_1)) ->
         collect(el | c.els) -> flatten()
           -> collect(e | thisModule.Element2TransformedElement(e)) ) }

lazy rule Element2TransformedElement{
   from el : MM!Element_2
   to    transEl1 : MM
      name <- 'T
      transEl2 : MM1 ! TransformedEl(name <- 'Sub_' + transEl1.name) }
```

```
transformedEls <- MM!Element_2.allInstances() ->
   collect(e | thisModule.Element2TransformedElement(e)
```

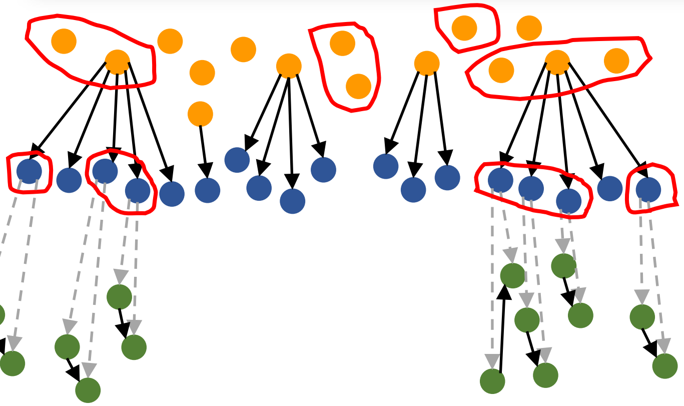


- Element_1
- Element_2
- TransformedEl

```
rule Model2Model{
   from inM : MM!Model
   to    outM : MM1!Model(
      transformedEls <- inM.elements -> selectApprox(CL, ER, e | e.oclIsTypeOf(MM!Element_1)) ->
         collectApprox(CL, ER, el | c.els) -> flatten()
           -> collect(e | thisModule.Element2TransformedElement(e)) ) }

lazy rule Element2TransformedElement{
   from el : MM! Element_2
   to    transEl1 : MM
      name <- 'T
      transEl2 : MM1 ! TransformedEl(name <- 'Sub_' + transEl1.name) }
```

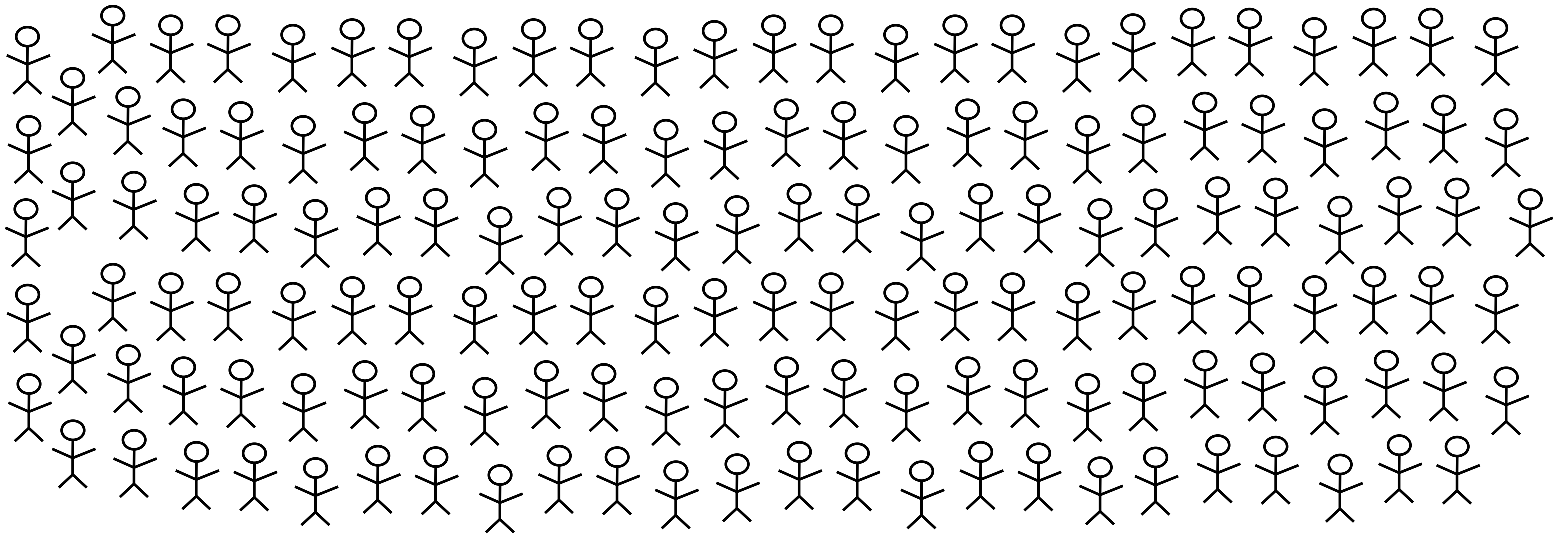
```
transformedEls <- MM!Element_2.allInstancesApprox(CL, ER) ->
   collect(e | thisModule.Element2TransformedElement(e)
```



- Element_1
- Element_2
- TransformedEl

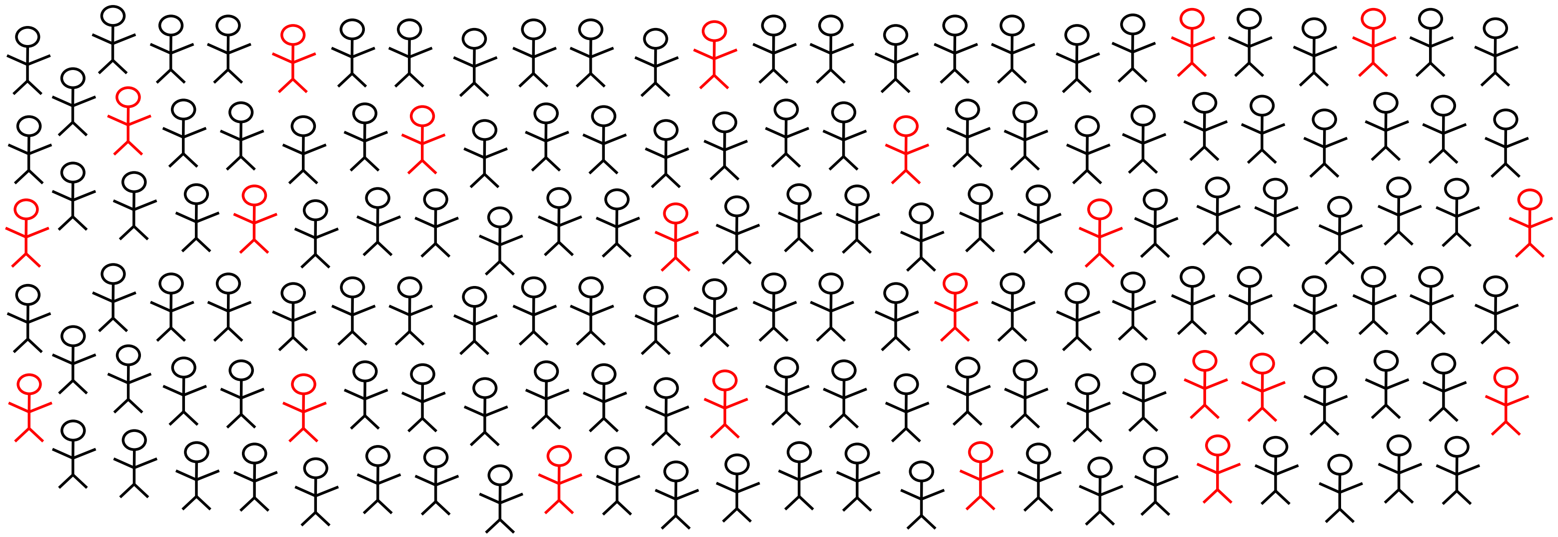
Initial Ideas

- How to select a subset of elements?
  - Leveraging sampling for Approximate Model Transformations



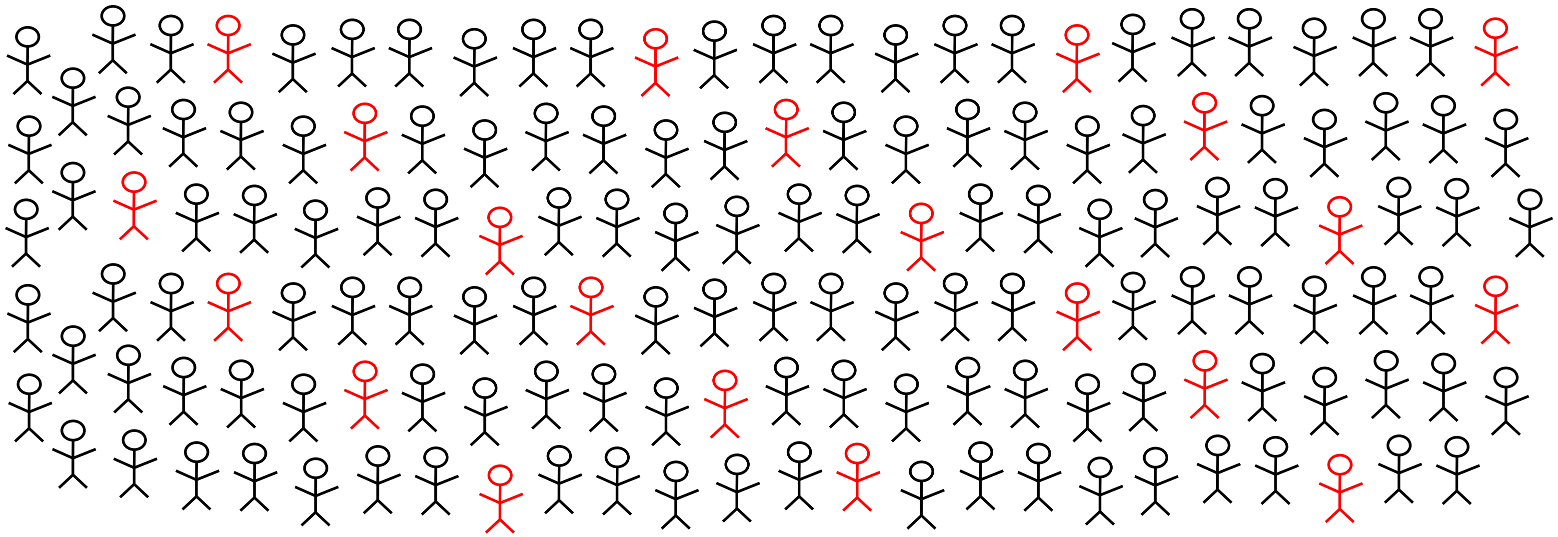
Initial Ideas

- How to select a subset of elements?
  - Leveraging sampling for Approximate Model Transformations



Random Sampling

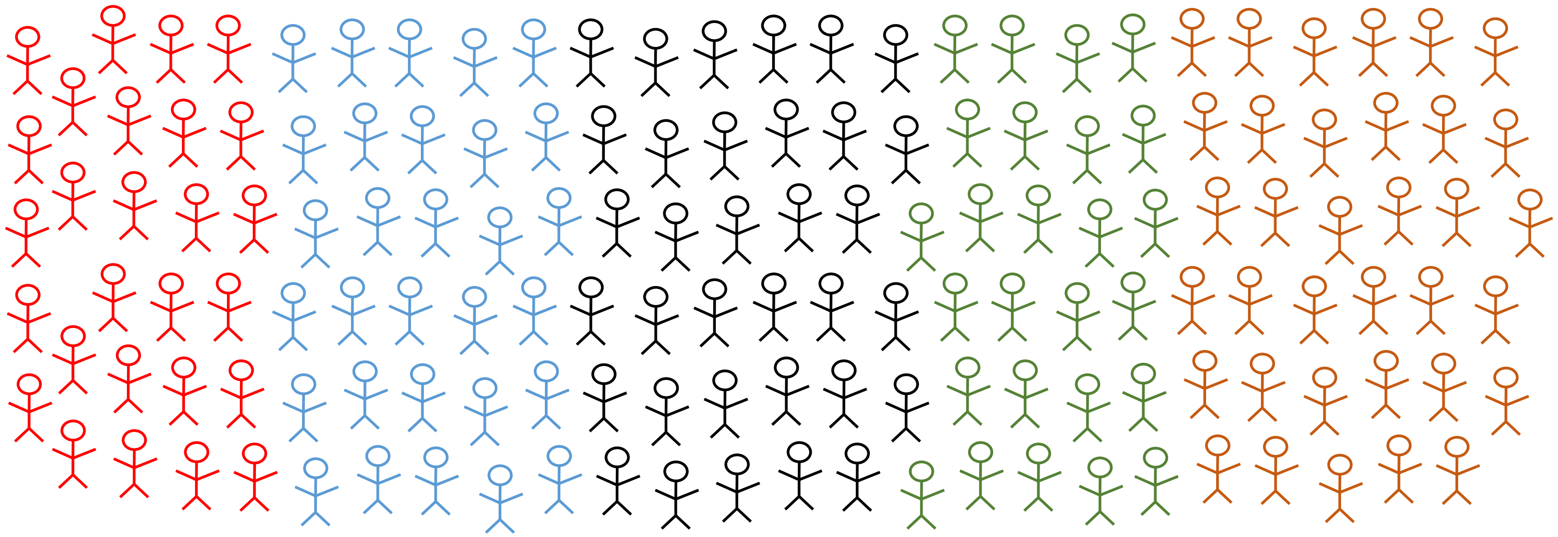
- How to select a subset of elements?
  - Leveraging sampling for Approximate Model Transformations



Systematic Sampling

Initial Ideas

- How to select a subset of elements?
  - Leveraging sampling for Approximate Model Transformations



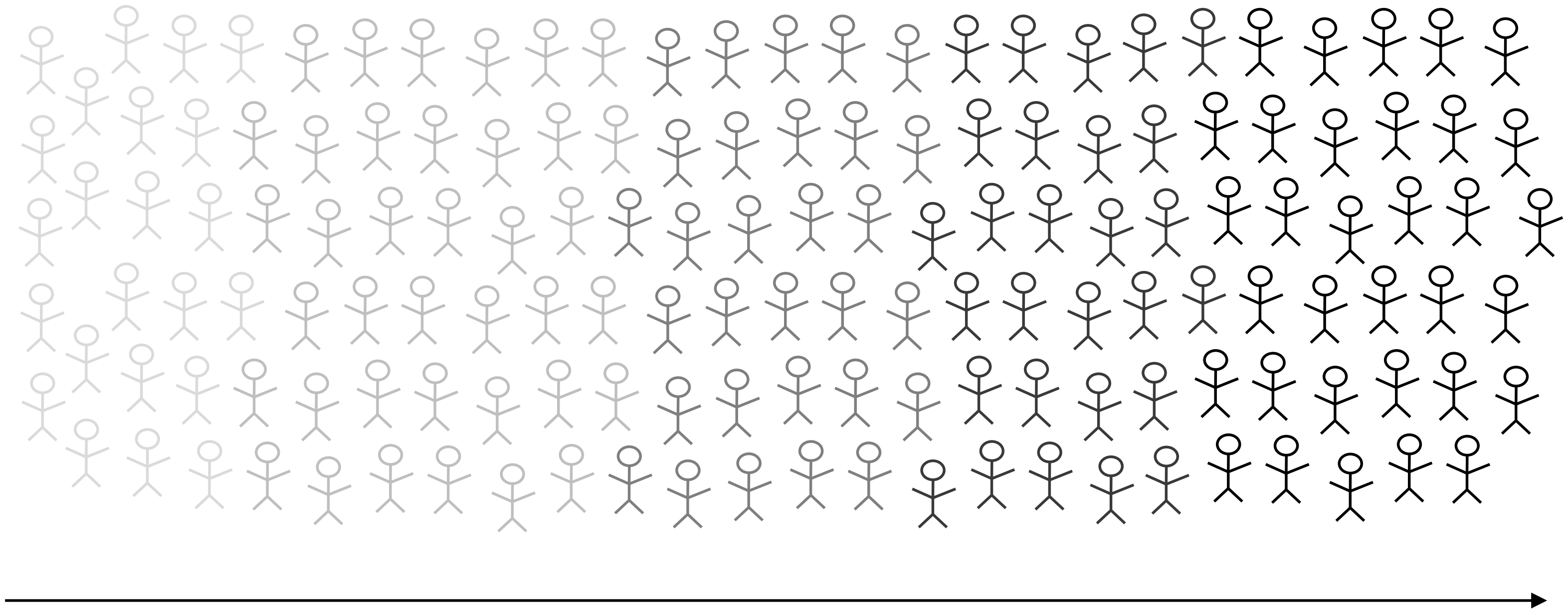
Stratified Sampling

Each strata is treated independently

Homogeneous subgroups

Initial Ideas

- How to select a subset of elements?
  - Leveraging sampling for Approximate Model Transformations



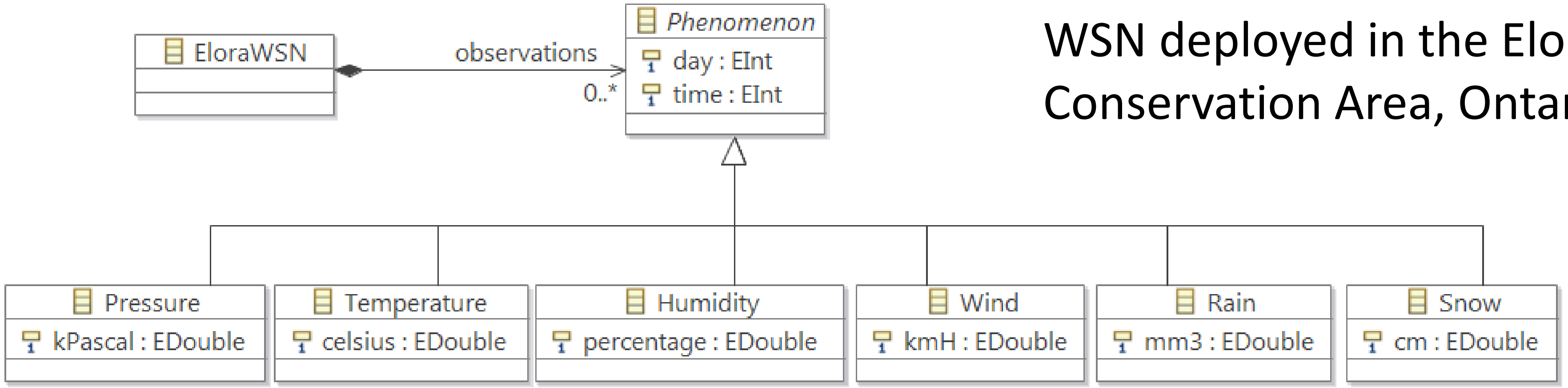
Cluster Sampling

Sampling clustered by geography

Sampling clustered by time periods

Motivating Example

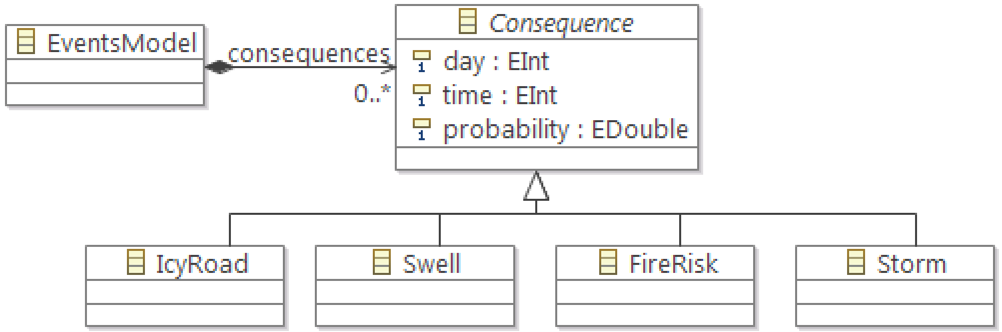
- Transformation Scenario



WSN deployed in the Elora George Conservation Area, Ontario, Canada,

Rule-based reasoning

- If it has not rained for weeks, the humidity is low and we have very high temperatures, the risk of having a fire increases



Sheth, A., Henson, C., Sahoo, S.S.: ***Semantic Sensor Web****. IEEE Internet Computing 12(4), 78–83 (2008)*

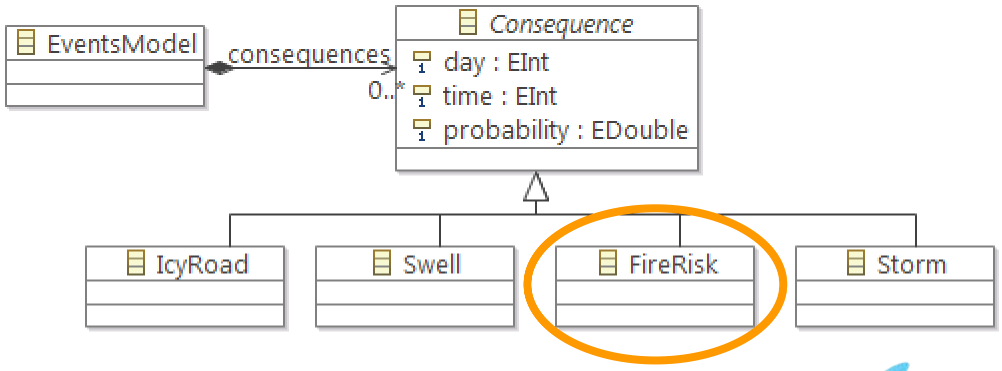
Implementation

- We focus on calculating the risk of having a fire
  - MacArthur FFDI index:

$$FFDI = 2 * (0.987 * log(\underbrace{D}_{\text{Drought}}) - 0.45 + 0.0338 * \underbrace{T}_{\text{Temperature}} + 0.0234 * \underbrace{V}_{\text{Wind}} - 0.0345 * \underbrace{H}_{\text{Humidity}})^{e}$$

Noble, I.R., Gill, A.M., Bary, G.A.V.: ***Mcarthur's fire-danger meters expressed as equations****. Australian Journal of Ecology 5(2), 201–203 (1980)*

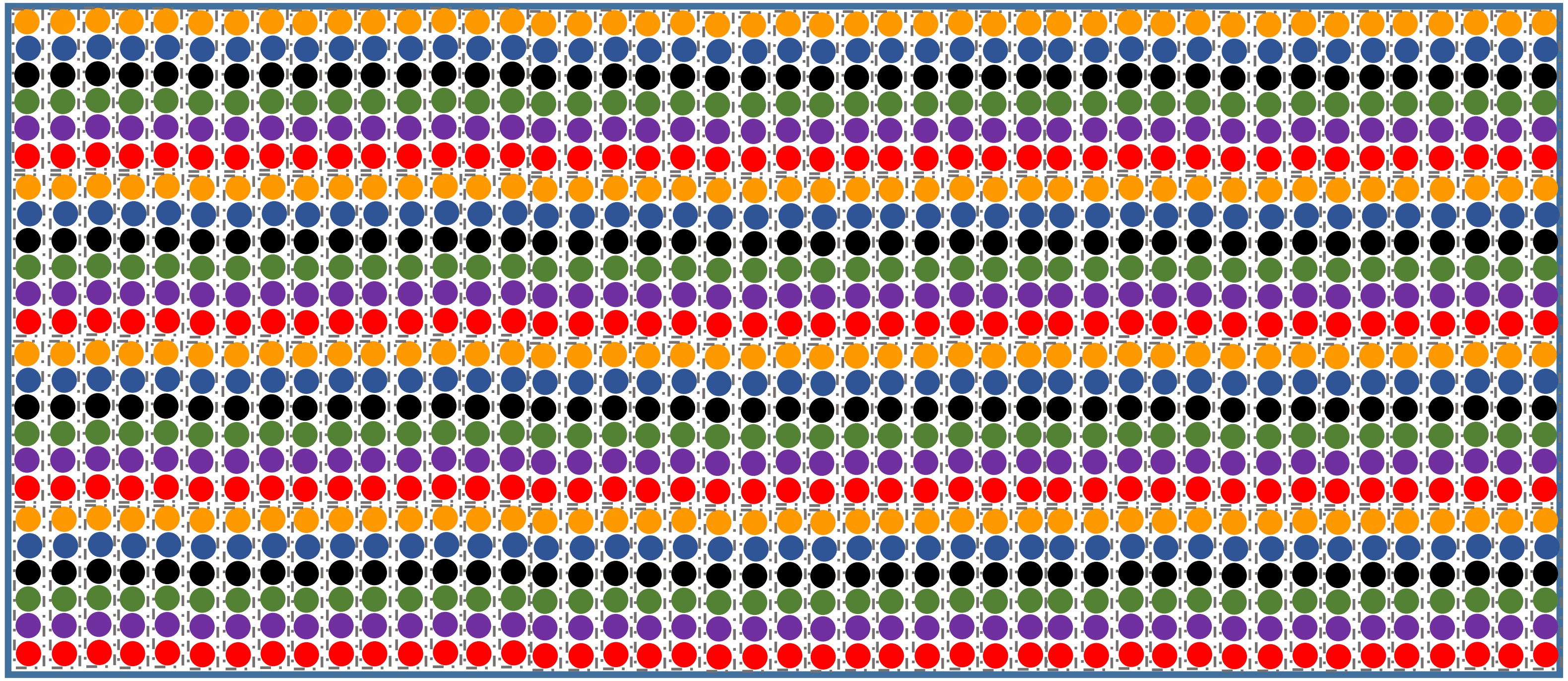
  - Risk of having a fire every month
    - The data from the previous month is used



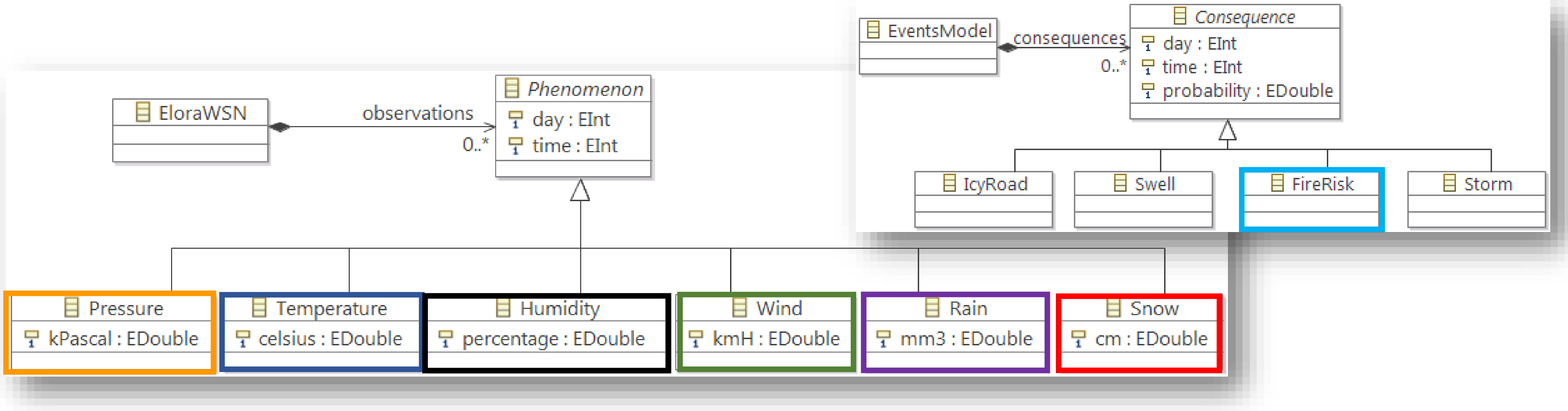
Implementation

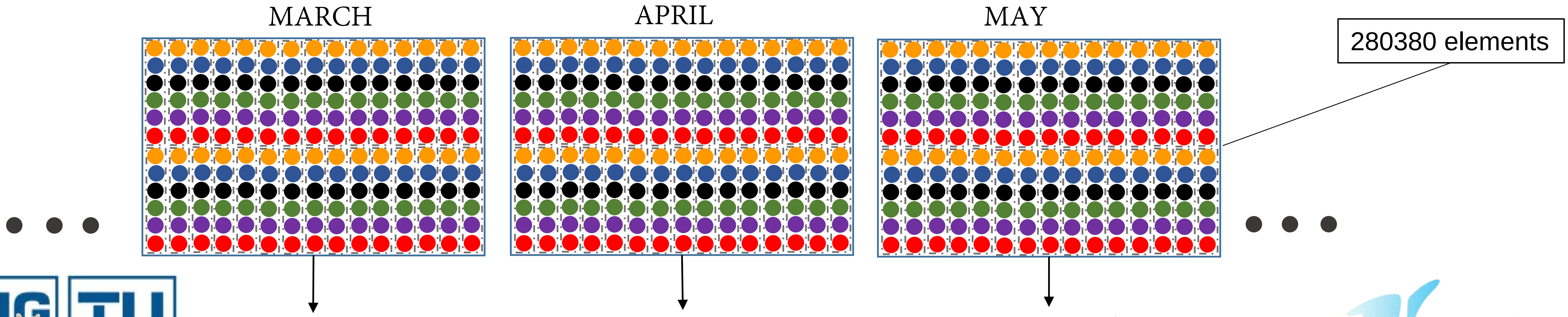
- Input data gathered in Elora during 2013 *(http://dataverse.scholarsportal.info/dvn/dv/ugardr)*
  - 8760 points in time (365 days, 24 measurements per day)
  - Data extrapolated
    - 6 types of data, 560,768 different measurements through time
    - Total of 3,364,608 objects in our input model
    - File of size 306MB

Day of Year	TIME	Pressure	Temperature	Humidity	wind speed	Rain	Snow
1	100	96,71	-4,091	71,3	34,92872	0	15,84
1	200	96,78	-4,48	69,95	36,61404	0	16
1	300	96,85	-5,065	72,1	36,61404	0	15,55
1	400	96,84	-5,564	69,52	37,81784	0	15,82
1	500	96,84	-6,304	76,6	29,28012	0	14,85



Implementation





April Fire Risk
Probability: X

May Fire Risk
Probability: Y

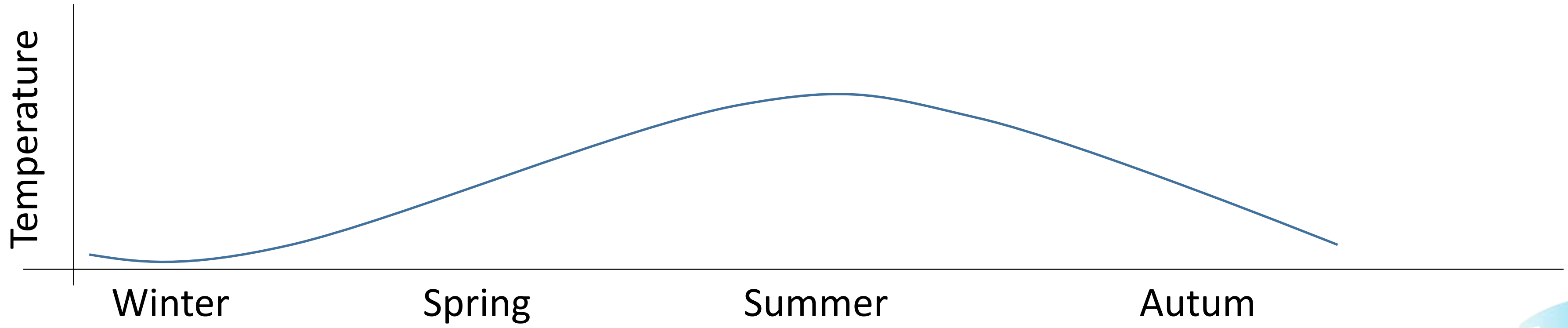
June Fire Risk
Probability: Z

Business Informatics Group

Strategy Selected

- Sampling Strategy
  - Systematic Sampling
    - Random starting point and a fixed periodic interval
    - Sample size needed for calculating the interval
- Election of Sample Size **
  - According to confidence level and relative error
  - Appropriate for normal distributions

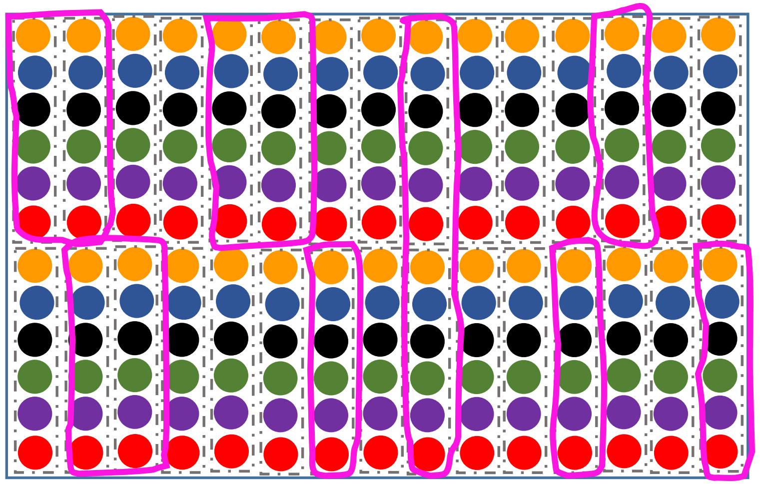
***Israel, G.D.: **Determining Sample Size**, University of Florida, Institute of Food and Agriculture Sciences (1992)*



Evaluation

- AMT1: Confidence Level: 95%, Error Rate 3%
- AMT2: Confidence Level: 99%, Error Rate 3%

# Elements	EMT	AMT1	AMT2
Of each type per month	46,730	1,043	1,764
In total per month	280,380	6,258	10,584
In total per year	3,364,608	37,548	63,504



- Implementation realized in Java/EMF as proof-of-concept
  - API generated for the metamodels in Java
  - EMT considers all elements
  - AMTs consider only data selected by Systematic Sampling

Implementation and Results

Month	FFDI in EMT	95% CL and 3% RE FFDI in AMT	95% CL and 3% RE Error	99% CL and 3% RE FFDI in AMT	99% CL and 3% RE Error
January	0.36845	0.36778	0.00181	0.36445	0.01097
February	0.37482	0.40547	0.07559	0.36478	0.02751
March	0.36216	0.37434	0.03252	0.35931	0.00795
April	0.40994	0.43950	0.06727	0.38833	0.05565
May	0.59438	0.59420	0.00031	0.59081	0.00604
June	0.69891	0.76448	0.08577	0.69739	0.00217
July	0.73598	0.78531	0.06281	0.71403	0.03073
August	0.80324	0.83444	0.03739	0.80616	0.00362
September	0.76696	0.71483	0.07939	0.83423	0.08063
October	0.80842	0.73761	0.09599	0.79740	0.01381
November	0.77053	0.79436	0.02999	0.75324	0.02296
December	0.48258	0.49907	0.03304	0.47241	0.02154
Exec Time	0.25651	0.005154	–	0.011153	–

- Performance gain
  - AMT1 49 times faster than EMT
  - AMT2 23 times faster than EMT
- Approximation of results
  - AMT1 -> relative error 6.287%
  - AMT2 -> relative error 2.85%

Conclusions, Open Questions, Future Directions

- We have come up with the concept of Approximate Model Transformation
  - We do not define a new language
  - This opens an extensive line of research
  - First results show the feasibility to keep studying AMTs

What’s next?

- How accurate should the models obtained by an AMT be?
  - How can we measure accuracy?
- How fast should an AMT be?

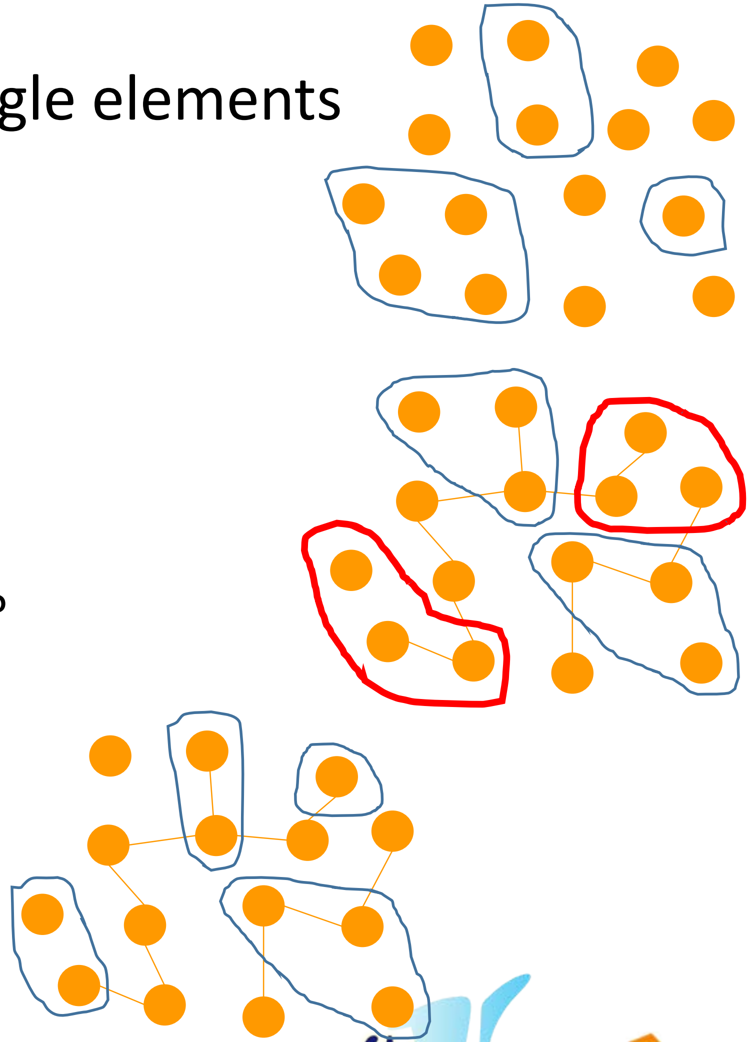
Open questions in the design of AMTs

- How many and which elements should be considered from the input model?
  - How big should the sample be?
  - What elements should be included in the sample?
- How complete should the models produced by an AMT be?
- Provide formal and precise specifications for our approximate operators
  - Integrate our operators in OCL, ATL

Conclusions and Future Directions

- In our scenario, the match has been performed with single elements
  - What if we want to match sub graphs
    - This is expensive in databases: joining of tables
    - It could be improved with graph query/transformation languages
  - Maybe a subset of the sub graph would be enough

- In our scenario, the data followed a normal distribution
  - The sampling mechanism was appropriate for it
    - How to select the sampling mechanism depending on the problem?
    - How much has to be known about the models?



Type.allInstancesApprox(maxTime)
Type.allInstancesApprox(maxNumElements)

Business Informatics Group

Atenea Modeling Software Systems

Towards Approximate Model Transformations

Javier Troya, Manuel Wimmer, Loli Burgueño, Antonio Vallecillo

THANKS!!

Third Workshop on the Analysis of Model Transformations
AMT'14. Valencia, Spain 29.09.14

MODELS 2014
VALENCIA, SPAIN

Business Informatics Group
Institute of Software Technology and Interactive Systems
Vienna University of Technology

GISUM/Atenea Research Group
ETS Ingeniería Informática
Unviersidad de Málaga