

CONTENTS

1	INTRODUCTION	1
1.1	Context	1
1.1.1	Environmental Computational Models	1
1.1.2	Transparency in Environmental Modeling	2
1.1.3	Natural science spreadsheets	3
1.2	Making knowledge explicit	4
1.2.1	The knowledge model in CommonKADS	4
1.2.2	Ontologies and the Semantic Web	6
1.2.3	Making knowledge explicit in scientific spreadsheets	6
1.3	Research questions	8
1.4	Approach	9
1.5	Project context	11
1.6	Thesis outline and guide to the reader	11
1.7	Publications	12
2	CONTEXT AND MOTIVATION	15
2.1	Introduction	15
2.2	Methodology	18
2.2.1	Assessing regime effectiveness: building on international regime literature	18
2.2.2	Constructing a conceptual framework	20
2.2.3	Review of modeling approaches	25
2.2.4	Our approach: fuzzy logic	27
2.2.5	Constructing a fuzzy model	28
2.3	Evaluation of the model	34
2.3.1	Collection of model input	34
2.3.2	Model calculation	37
2.3.3	Model results	38
2.3.4	Reflection on model results	42
2.4	General discussion and conclusions	44
3	PROBLEM ANALYSIS: TRANSPARENCY IN ENVIRONMENTAL MODELS	47
3.1	Introduction	47
3.2	Material and methods	49
3.3	Results	50

3.3.1	Documentation	50	
3.3.2	Conceptual model	51	
3.3.3	Calculation process	51	
3.3.4	Data	52	
3.3.5	Source code	52	
3.3.6	Reproducibility and transparency	52	
3.4	Discussion	54	
3.4.1	Model size and complexity	55	
3.4.2	Lack of incentives	55	
3.4.3	Use of computers	56	
3.4.4	Change of attitude	57	
3.5	Conclusion	58	
4	PILOT STUDIES	59	
4.1	Introduction	59	
4.2	Pilot 1: Ontologies in peer reviews	61	
4.2.1	Related work	61	
4.2.2	Method	62	
4.2.3	Results	64	
4.2.4	Discussion	66	
4.3	Pilot study 2:Reconstructing Semantics	68	
4.3.1	Related Work	68	
4.3.2	Method	69	
4.3.3	Ontology reconstruction	70	
4.3.4	Characteristics of the analyzed spreadsheets	73	
4.3.5	Interview with Model Developers	74	
4.3.6	Discussion	75	
4.4	Pilot 3: Knowledge Representation in Scientific Models	77	
4.4.1	Related work	77	
4.4.2	Method	77	
4.4.3	Results	79	
4.4.4	Discussion	83	
4.5	Discussion and Conclusion	83	
5	SPREADSHEET ANNOTATION	85	
5.1	Introduction	85	
5.2	Related work	87	
5.3	Method	90	
5.3.1	Basic principles	90	
5.3.2	Property 1: blocks within a table	93	
5.3.3	Property 2: units of measure	93	

5.3.4	Property 3: quantities	94
5.3.5	Property 4: block typology	94
5.3.6	Property 5: grouping of similar domain concepts	95
5.4	Case study	97
5.4.1	Data set	98
5.4.2	Vocabularies	100
5.4.3	Ground truth annotation	100
5.4.4	Baseline annotation	101
5.5	Results	102
5.5.1	Annotation of unit terms: approach 2	102
5.5.2	Annotation of quantity terms: approach 3	105
5.5.3	Block typology	106
5.5.4	Grouping of similar domain concepts: approach 5a	108
5.6	Discussion	108
5.6.1	Discussion of the annotation approach per property	108
5.6.2	General discussion	111
5.7	Conclusions and future work	113
6	CONSTRUCTING THE CALCULATION MODEL	115
6.1	Introduction	115
6.2	Related work	117
6.3	Methodology	119
6.3.1	Basic principles	119
6.3.2	Construction and aggregation of the cell dependency graph	120
6.3.3	Development and application of heuristics	121
6.4	Case study: set up	122
6.4.1	Raw Data	122
6.4.2	Construction of the ground truth calculation workflow	123
6.4.3	Results of the ground truth calculation workflow	124
6.4.4	Comparison with the ground truth calculation workflow	126
6.5	Case study: results	126
6.5.1	Automatic aggregation	126
6.5.2	Development and application of heuristics	127
6.5.3	Comparison with the ground truth workflow	128
6.6	Discussion	130

6.7	Conclusion and future work	133
7	SYNTHESIS	135
7.1	Introduction	135
7.2	Related Work	137
7.3	Reconstruction of the domain model	138
7.3.1	Two layers within the domain model	138
7.3.2	Requirements for spreadsheet tables	140
7.3.3	Reconstruction of the descriptive layer	140
7.3.4	Reconstruction of the computational layer	145
7.4	Analysis of table design	146
7.4.1	Method	147
7.4.2	Results	148
7.5	Heuristics	154
7.5.1	Recognizing and annotating blocks	156
7.5.2	Collecting missing information using vocabularies	157
7.5.3	Deducing missing information from the table context	158
7.5.4	Formula analysis	158
7.6	Repairability and implications of deviations	159
7.6.1	Deviant basic structure	159
7.6.2	Grouped phenomena	160
7.6.3	Deviant unit notations	160
7.6.4	Incomplete quantity notations	161
7.6.5	Unclear phenomenon notations	161
7.6.6	Insufficient c.e. regions	161
7.7	Conclusion and discussion	162
8	CONCLUSION AND DISCUSSION	165
8.1	Research Questions Revisited	165
8.1.1	How can the domain model behind a natural science spreadsheet be adequately described?	165
8.1.2	What are guidelines for the process of reconstructing the domain model behind a natural science spreadsheet?	168
8.1.3	General conclusion	171
8.2	Discussion and Further Research	171
A	APPENDICES	177
A.1	Appendix Chapter 2	177

A.1.1	Supplementary Information	177
A.1.2	Knowledge rules	177
A.2	Appendix Chapter 4: Heuristics	182
A.3	Appendix Chapter 5: Heuristics and rules	183
BIBLIOGRAPHY		197