

THE UNIVERSITY OF SOUTH ALABAMA
COLLEGE OF ENGINEERING

TOWARDS THE EVOLUTION OF INFORMATION IN DIGITAL ECOSYSTEMS

BY

Kari J. Lippert

A Dissertation





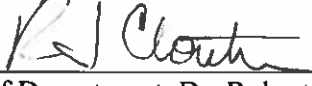


Submitted to the Graduate Faculty of the
University of South Alabama
in partial fulfillment of the
requirements for the degree of

Doctor of Science
in
Systems Engineering

May 2018

Approved:

Date:

 Chair of Dissertation Committee: Dr. Robert J. Cloutier	<u>3/9/18</u>
 Committee Member: Dr. Michael V. Doran	<u>3/21/18</u>
 Committee Member: Dr. Samuel Russ	<u>3/21/18</u>
 Committee Member: Dr. Thomas G. Thomas, Jr.	<u>3/21/18</u>
 Chair of Department: Dr. Robert J. Cloutier	<u>3/22/18</u>
 Director of Graduate Studies: Dr. Robert J. Cloutier	<u>3/22/18</u>
 Dean of the Graduate School: Dr. J. Harold Pardue	<u>4/9/18</u>

PREVIEW

TOWARDS THE EVOLUTION OF INFORMATION IN DIGITAL ECOSYSTEMS

A Dissertation

Submitted to the Graduate Faculty of the
University of South Alabama
in partial fulfillment of the
requirements for the degree of

Doctor of Science

in

Systems Engineering

by

Kari J. Lippert

B. S., University of Toledo, 1980

M.M.C., University of Toledo, 1991

M.S., Johns Hopkins University, 2002

December 2017

Copyright ©2018 by Kari J. Lippert
All rights reserved.

ProQuest Number: 10790760

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10790760

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ACKNOWLEDGMENTS

To all my friends and family, I thank you for coming with me on this journey and providing me with support and the drive to persevere. To my advisor, Dr. Robert Cloutier, thank you for being the lighthouse on this journey and understanding the many times that life got in the way.

PREVIEW

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES	ix
ABSTRACT	xiv
CHAPTER I OVERVIEW	1
1.1 Problem Statement	3
1.2 Background	5
1.3 Purpose of this Research	12
1.3.1 Hypotheses	14
1.3.2 Objectives	14
1.3.3 Scope	16
1.3.4 Contributions of this Research	17
1.4 Approach and Structure	20
1.4.1 Objective 1. Demonstration of the Evolution of Data.	21
1.4.2 Objective 2. Discussion of the Innovation with TRIZ.	21
1.4.3 Objective 3. Development of the Concept of the Digital Ecosystem of Cyberspace.	21
1.4.4 Objective 4. Coping with Evolution in a Digital Ecosystem.	22
1.4.5 Objective 5. Visualization and Simulation of the Digital Ecosystem.	23
CHAPTER II BACKGROUND AND RELEVANT LITERATURE	24
2.1 Models	25
2.1.1 Mental Models	26
2.1.1.1 The Structure of Mind	32

2.1.1.2 Language.....	34
2.1.1.3 Perception and Viewpoint.....	39
2.1.1.4 Analogy and Metaphor.....	42
2.1.1.5 Conceptual Models.....	46
2.2.1 Digital Models.....	47
2.2.1.1 Relational Models.....	54
2.2.1.2 Object Oriented Models.....	55
2.2.1.3 Semantic Models.....	57
2.2.1.4 Other Models.....	63
2.3 Patterns.....	67
2.3.1 Model Integration and Mapping.....	69
2.3.2 Heterogeneity.....	75
2.4 Ecosystems.....	76
2.5 Engineering Systems.....	82
2.5.1 System Architecture.....	82
2.5.2 Digital Information Architectures.....	91
2.5.2.1 Dataspaces.....	92
2.5.2.2 Data Vault.....	92
2.5.2.3 Quantum Data Model.....	93
2.5.2.4 Cyc Knowledge Server.....	94
2.5.2.5 Multi-representational Databases.....	95
2.5.2.6 Data Clouds.....	98
2.5.2.7 Peers and Agents.....	99
2.6 Handling Change.....	100
2.6.1 Biological Systems and Biomimicry.....	100
2.6.1.1 Brain and Memory.....	102
2.6.1.2 Genetics and Protein Synthesis.....	104
2.6.1.3 Homologous Sequences.....	111
2.6.1.4 Ecto- and Endosystems.....	114
2.6.1.5 Immunology.....	114
2.7 Types of Change.....	117
2.8 Future Systems.....	121
2.9 Networks and Cyberspace.....	128
2.10 Visualization and Simulation.....	131

2.11 Thought Experiment	133
CHAPTER III APPROACH AND METHODOLOGIES	135
3.1 Innovation for Systems Engineering.....	135
3.2 Demonstration of Evolution in System Data	137
3.3 Ecosystems.....	139
3.4 Metaphors and Patterns in Systems Engineering.....	141
CHAPTER IV EVOLUTION IN SYSTEM DATA.....	143
4.1 Data Evolution Characterization in the US Census Data.....	145
4.1.1 Evolution at the 50 Year Mark.....	146
4.1.1.1 Schema Generalization.	146
4.1.1.2 Structural Impedance Mismatch.	146
4.1.1.3 Value Aggregation and Impedance Mismatch.....	146
4.1.1.4 Additive Schema.	146
4.1.2 Evolution at the 100 Year Mark.....	147
4.1.2.1 Schema Disaggregation.	147
4.1.2.2 Schema Generalization and Representation.....	147
4.1.2.3 Restructuring Schema.	147
4.1.2.4 Schema Value Restructuring.....	147
4.1.2.5 Impedance Mismatch.	147
4.1.2.6 Additive Schema.	148
4.1.2.7 Subtractive/Replacement Schema.....	148
4.1.2.8 Semantic Labeling.	148
4.1.3 Evolution at the 150 Year Mark.....	148
4.1.3.1 Schema Disaggregation.	148
4.1.3.2 Impedance Mismatch.	148
4.1.3.3 Additive Schema.	148
4.1.3.4 Subtractive Schema.....	148
4.1.3.5 Confounding Semantics.....	149
4.1.4 Evolution at the 200 Year Mark.....	149
4.1.4.1 Value Representation.	149
4.1.4.2 Confounding Semantics.	149
4.1.4.3 Restructuring Schema.	150
4.1.4.4 Additive Schema.	150

4.1.5 Evolution Beyond the 200 Year Mark	150
4.1.5.1 Schema Replacement.....	150
4.1.5.2 Restructuring Schema.....	150
4.1.5.3 Additive Schema.....	150
4.2 Discussion.....	151
CHAPTER V THE DIGITAL ECOSYSTEM.....	156
5.1 A Model of Cyberspace	157
5.2 Coevolution Through Dynamics.....	162
5.3 Cyberspace Version of SIR, the UDO	170
5.4 Data of the Ecosystem	176
5.5 Metaphors for Information.....	183
5.5.1 A Pattern Language for a Garden.....	189
5.5.2 A Garden as a Metaphor for a Network.....	192
5.6 Discussion.....	198
CHAPTER VI VERIFICATION AND VALIDATION.....	199
6.1 Evolution of Data.....	199
6.2 Innovation Validation	200
6.3 Digital Ecosystem	201
6.4 On the Use of Metaphor in Visualization	202
6.4.1 Metaphor in a Net Flow Simulation.....	204
6.5 Coping with Evolution Through Metaphor.....	207
6.6 Discussion.....	208
CHAPTER VII CONCLUSIONS AND FUTURE WORK	209
7.1 A Summary of the Journey	209
7.2 Conclusions.....	210
7.3 Future Work.....	210
REFERENCES	213

APPENDICES	241
Appendix A The US Census of Population	241
Appendix B TRIZ for Innovation	309
Appendix C Agent-Based Simulation of Cyberspace.....	399
BIOGRAPHICAL SKETCH	426

PREVIEW

LIST OF TABLES

Table	Page
1. Causality Relationships.....	56
2. Temporality Relationships.....	57
3. Participant Relationships.....	58
4. Spatial Relationships.....	59
5. Conjunctive Relationships.....	59
6. Quality Relationships.....	60
7. Network and Garden Metaphors.....	196
8. Pattern Language for Garden Metaphor.....	205
Appendices	
Table	
A-1. Age ranges 1830 Census.....	250
B-1. Characteristics of Systems.....	314
B-2. Characteristics of Digital Systems.....	321
B-3. Principles of Transformation.....	323
B-4. Extended Contradiction Matrix.....	345
B-5. Contradiction Matrix Application.....	395
C-1. NetLogo Model Default Rate Settings.....	409

LIST OF FIGURES

Figure	Page
1. Information System Components.	4
2. Scope of Work.	15
3. Model, Metaphor, or Imagination?	22
4. MOM is WOW - A Matter of Perspective.	39
5. Chocolate Chip Cookie Recipe.	47
6. Another Chocolate Chip Cookie Recipe.	48
7. Recipe Contextual Model.	49
8. A Different Recipe Conceptual Model.	50
9. Recipe Physical Models.	51
10. A Chemist's Chocolate Chip Cookies.	52
11. ADM Entities and Associations.	64
12. An Architectural Drawing.	83
13. Structure of DNA.	103
14. Building Blocks.	105
15. Encoding Showing Wobble Position.	106
16. tRNA.	108
17. tRNA Binding an Amino Acid.	109

18. Ribosome.	110
19. Homologous Sequence in Histone H1 (Shafee, 2014).....	112
20. An Antibody.....	115
21. TRIZ Methodology.	136
22. 1790 Person Concept Map.	138
23. 1800 Person Concept Map.	139
24. Cyber Warfare Systemigram part I.	158
25. Cyber Warfare Systemigram part II.....	159
26. Cyber Warfare Systemigram part III.	160
27. Cyber Warfare Systemigram part IV.	161
28. Cyber Warfare Systemigram Complete.	162
29. UDO Model of Cyber Warriors.	172
30. Dual Outcome Model.....	173
31. Network Model.	177
32. Portion of the Network Model.	178
33. Events and Health Models.	179
34. Satellite and Traffic Overlays.	183
35. Symbols.....	184
36. Gardens.	191
37. Types of Networks.....	193
38. A Network Garden.	197
39. An Anomaly Becomes Obvious.....	203
40. Net flow in Garden Viewer – PnP Check-ins.	204

41. Net flow in Garden Viewer – Email Exchanges.....	205
42. Net flow in Garden Viewer – Web Browsing.....	206
43. A Summary.	209

Appendices

Figure	Page
A-1. Administrative Data 1790.....	244
A-2. Education Level Encoding.....	270
A-3. 1950 Census Infant Concept Model.	272
A-4. 1790 Census.....	286
A-5. 1800 Census.....	287
A-6. 1810 Census.....	288
A-7. 1820 Census.....	289
A-8. 1830 Census.....	290
A-9. 1840 Census.....	291
A-10. 1850 Census.....	292
A-11. 1860 Census.....	293
A-12. 1870 Census.....	294
A-13. 1880 Census.....	295
A-14. 1890 Census.....	296
A-15. 1900 Census.....	297
A-16. 1910 Census.....	298
A-17. 1920 Census.....	299
A-18. 1930 Census.....	300

A-19. 1940 Census.....	301
A-20. 1950 Census.....	302
A-20. 1960 Census.....	303
A-21. 1970 Census.....	304
A-22. 1980 Census.....	305
A-23. 1990 Census.....	306
A-24. 2000 Census.....	307
A-25. 2010 Census.....	308
B-1. Erector Set TM	312
B-2. Contradiction Matrix Use.	343
B-3. Strategic Themes of an Information Architecture.	372
B-4. Accommodating New Items in a Coat Check System.	376
B-5. eProfile (First Pass).....	380
B-6. eProfile with address Module.	382
B-7. avatar Model.	386
B-8. Potential Overlap of eProfile and avatar Models.....	388
B-9. Elephantpigapottamus.....	390
B-10. Information System Operational Zones.	392
C-1. QURDO Model.....	401
C-2. Population Study (100, 250, 500 for 500 ticks).....	402
C-3. 616 Ticks.....	404
C-4. 1046 Ticks.....	405
C-5. 3555 Ticks.....	406

C-6. 52128 Ticks.....	407
C-7. Defense Recruitment Twice Offensive for 847 Ticks.	411
C-8. Defense Recruitment Twice Offensive for 1516 Ticks.	411
C-9. Defense Recruitment Twice Offensive for 4680 Ticks.	412
C-10. Defense Recruitment Twice Offensive for 21728 Ticks.	412
C-11. Decreased Conversion of Defense to Offense at 5749 Ticks.	413
C-12. Arrest Rate of 0.200 for 14670 Ticks.	413
C-13. A Different Stability at 22386 Ticks.....	414
C-14. Stability in Jeopardy at 5446 Ticks.....	414
C-15. Arrest Rate of 0.250 for 310033 Ticks.	415
C-16. Lower Arrest Rate with Higher Recruitment at 1395 Ticks.	415
C-17. Rate Sliders and Counter Displays.	416
C-18. User Interface of Model.....	418

ABSTRACT

Lippert, Kari, J., D. Sc., University of South Alabama, May 2018. Towards the Evolution of Information in Digital Ecosystems. Chair of Committee: Robert Cloutier, Ph.D.

Digital ecosystems are the next generation of Internet and network applications, promising a whole new world of distributed and open systems that can interact, self-organize, evolve, and adapt. These ecosystems transcend traditional collaborative environments, such as client-server, peer-to-peer, or hybrid models (e.g., web services) to become a self-organized, interactive environment. The complexity of these digital ecosystems will encourage evolution through adaptive processes and selective pressures of one member on another to satisfy interaction, adaptive organization, and, incidentally, human curiosity.

Building an ecosystem, whether natural or digital, requires the correct components. At present, digital ecosystems cannot be built because critical components which support adaptive evolution are missing. An adaptive, evolving data model is one such component. Like the DNA of living systems, this model is a type of encoding that allows and supports the evolution necessary to cope with a changing environment. However, as in a living system, the evolving data model does not stand alone. It must be an integral part of agile information architecture. Such architecture will require more than standard systems engineering; it will require evolution engineering and systems thinking.

This new architecture abandons the traditional systems engineering strategies of well-planned and fully understood systems, and it replaces them with the creation of a planned environment which fosters learning by doing and which enables unanticipated change. Systems thinking will help in this creation by asking not only what systems are involved, but also what kinds of systems might be involved that has not been thought about before. This type of innovation is critical as the choices and paradoxes of the ecosystem are considered and reconciled. The foundations of a system of systems are coexistence, cooperation and coeducation. These three are also foundations of an ecosystem.

This work addresses one of the essential parts of the digital ecosystem – the information architecture. The research, inspired by systems thinking influenced by both biological models and science fiction, applies the TRIZ (теория решения изобретательских задач, teoriya resheniya izobretatelskikh zadatch) method to the contradictions raised by evolving data. This inspired the application of patterns and metaphor as a means for coping with the evolution of the ecosystem. The metaphor is explored as a model of representation of rapidly changing information through a demonstration of an adaptive digital ecosystem. The combination of this type of data representation with dynamic programming and adaptive interfaces will enable the development of the various components required by a true digital ecosystem.

Nothing endures but change.

–Heraclitus

CHAPTER I

OVERVIEW

Improving information analysis is a pressing challenge (Desouza, 2009; Obama, 2009). In many fields such as laboratory science, law, meteorology, medicine, marketing, intelligence analysis, security, and even in news reporting, the ability to collect data is increasing at a faster rate than the ability to analyze it (Thomas & Cook, 2005).

Accommodating both changes and the unknown without making storage decisions based on incomplete information requires that today's systems store everything. This, in turn, significantly increases the space required for storage in the attempt to gather data that are more complete and postpone the development of analytic sense-making tools that will manipulate, combine, and understand the data. Massive collections of data continue to grow within every information domain.

Much effort is being devoted to automating analysis and the application of artificial intelligence will allow "data to find data" (Jonas, 2009). However, current information systems depend on either a static data model or on a model which has a very slow rate of change. As the rate of change of the data increases these systems are, in turn, increasingly unable to keep up with the changing demands on their underlying information architecture. Systems need a degree of foreknowledge about the message environment (Morningstar & Farmer, 2001) and systems that cannot adapt must be

replaced. This tightly couples the cost of a system to the clock speed (rate of change) of its data (Fine, 1998). Even when cost is no object, replacement development cannot keep pace with the dynamic nature of today's information space. As the volume and complexity of information being processed continues to increase, there is a growing desire to know tomorrow's data yesterday and to create the program to process it today. Dealing with tomorrow's unknown data is challenging, but it is equally difficult to know today what data will be of worth tomorrow – what data will an analyst need tomorrow to answer a question not yet envisioned?

As the value of cross-domain analytics emerge, their domain specific data stores collide when they use different semantics for similar labels. With the increased emphasis on information sharing among and between analytic domains, there is an increased need to be able to facilitate understanding for humans as well as machines. Resolution through mapping requires significant human input even in the most sophisticated machine guided systems (Ives *et al.*, 2009). The development of large centralized data models to be used by all is also fraught with problems (Doan & Halevy, 2005; Hurson, Bright, & Pakaad, 1994). In these systems, it is required that the same real-world object in local databases map to a single, global representation and that semantically different objects map to different global representations. Issues arise in the development of the global representation, and these include labeling (synonyms and homonyms), formatting differences, structural variation (relationships and other constraints), as well as the level of abstraction used at the local level (Mark & Roussopoulos, 1987). These issues are often rooted in cognition or perception as the data itself is always present; it is our focus that is evolving. Critical to future analytic success is the development of a framework that

supports spatial, temporal and contextual understanding; that supports uncertain, incomplete, misleading, and contradictory information; that supports multiple, simultaneous perspectives; and that supports manual and automated discovery and synthesis.

1.1 Problem Statement

The systems being built today are larger, more complex, and much less focused than their predecessors, such as the FAA Next Gen system, Smart Grid, Smarter Planet, Next Warrior, and the DARPA F6 modular satellites. These systems, and other information processing and analysis systems, are deliberately and carefully designed by system engineers, and implemented to meet known and specified needs about known information and environments. Current systems engineering practice requires knowledge in advance about the information a system will process to construct the data models which then become the foundation of the system, as shown in Figure 1. Increasingly, data requirements are less well defined, and the data itself is growing in complexity. Future information processing systems must be capable of embracing change since their complexity is driven by both multidisciplinary scope and the increasing complexity of the information which they process. This represents a significant contradiction in systems engineering: *the data evolves but there are no methods, patterns, or techniques currently available to engineer a system that supports or embraces this evolution.*

A case study is presented to illustrate data evolution (Eisenhardt, 1989). This

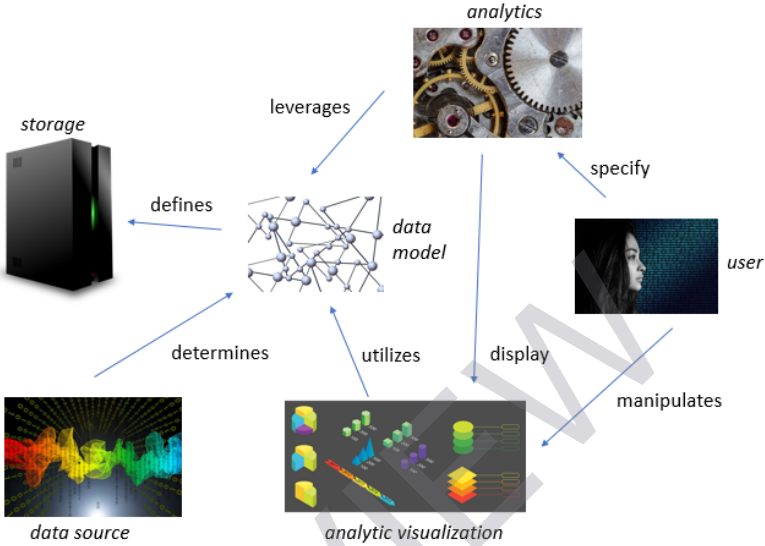


Figure 1. Information System Components.

research addresses the application and extension of an innovation methodology to study this problem of evolving data. The adaptive data model is created, managed, maintained, and utilized by the agents and the users alike. It combines the concepts of patterns (Alexander, Ishikawa, & Silverstein, 1977) and metaphor (Lakoff, 1990; Lakoff & Johnson, 1980; Pylyshyn, 2007) to enable the construction of a representation of a digital ecosystem. The research then discusses the development of a digital ecology, a foundation for the interconnected, interrelated, and interdependent digital species that manage data and information. It is easiest to envision as a multi-agent system (Jeffery,

2002) where the ecological species are agents who mediate with brokers on behalf of users to find any desired information. The agent populations in this envisioned digital ecology would co-evolve with new data appearing in the environment and new user requirements and age off in a manner determined by dictate or by lack of use.

1.2 Background

People depend on information to carry out the activities of daily life, including the assembly of ideas from multiple sources. Therefore, people create mental models which evolve over time as they adapt to new information, adapt to the viewpoints of others, and as their focus or interest changes (J. M. Carroll, 1988; Fodor, 1985; Gentner & Stevens, 1983; Hawkins, 2004; Johnson-Laird, Girotto, & Legrenzi, 1998; Schuck, 2010). Current digital systems are not able to mimic this adaptation, severely limiting their capabilities in information analysis. Useful data representations mimic the way in which humans think about information thus facilitating understanding, synthesis and retrieval (Korfhage, 1997).

The mental representation of information is part of cognitive science, a field focused on understanding human thought, memory and reasoning (Stillings *et al.*, 1995) which is also being tackled by complexity studies (Bar-Yam, 2005) and through neuroscience (Hawkins, 2004). Much of this research is centered on how these mental models evolve over time. Mental models form the foundation of communication; they require the information receiver to understand and either adapt to or adopt the mental model of the information provider.

The ability of an individual to communicate information to another depends largely on their ability to share this mental model. The model represents the sender's point of view about the information, or evokes a particular emotional response from the recipient (Gelernter, 1994; Rosenblatt, 1994, 1996). Mental models are representations of real, hypothetical, or even imaginary situations and they are used to remember, to learn, and to make decisions (Johnson-Laird *et al.*, 1998). Models used in computer systems, on the other hand, are deliberately and carefully designed by system engineers. The system design process begins with the definition of the data flowing into and out of the system – and this process has long held the resulting data model to be immutable.

Without data, computer programs would have nothing to do. Computer programs are built to receive, process, present, and store data and information.¹ The process of designing computer programs (and computer systems) begins with an examination of the data – the inputs and outputs. Historically, the information portion of a system focused on what could be done with the data at hand (e.g., sort, search, compute, display). Systems were built for a purpose, often to process specific data and to produce the specified reports. The requirements and specifications of the system provided a more or less complete understanding of the data and the ways in which it would be used. An engineer

¹ Arguably not identical, the terms *data* and *information* are nonetheless used interchangeably in this paper as the concepts can be applied to both and the distinction is immaterial in this context. *Information* is a set of data that has been matched to an information need. The concept of information has both personal and time-dependent components that are not present in the concept of data. *Data* can be organized independently of individual users, thus the organization of information is a more personal thing, requiring the active intervention of a user. In database systems where the same types of needs and questions arise repeatedly, much of the organization appropriate to these needs can be built into the storage system. In general information retrieval system, however, it may be impossible to anticipate fully the most appropriate organization until the various needs are expressed.