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

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

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.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Appendices

Figure Page

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

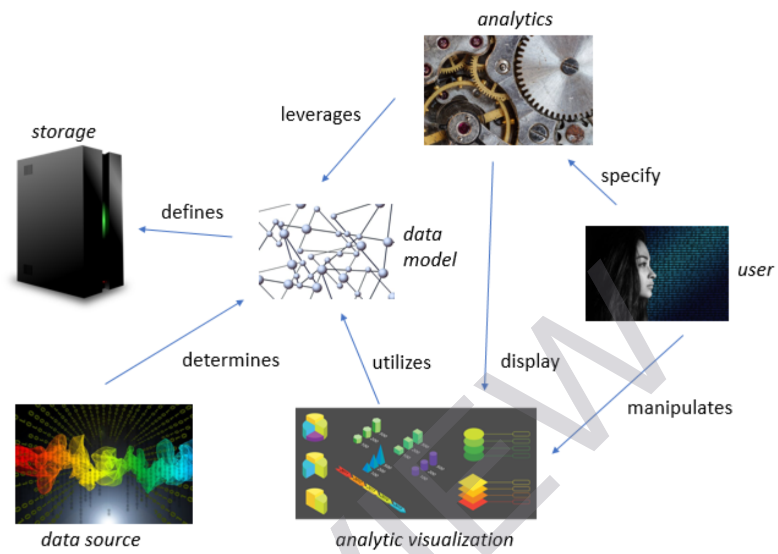

Figure 1. Information System Components.

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.