# Knowledge Building and Sharing:
# A Metamodel for Guided Research, Learning, and Application

with

Proof of Concept

MOSAIC:

Model of Securing Application Information Confidentiality

Kimberly Zeitz & Chris Frisina

Client: Randy Marchany
Director of the VT IT Security Lab

Prior Research Contact: Noha El Sherbiny

Blacksburg
May 8, 2013
CS6604

Virginia Tech

This report is in addition to our conference paper aimed for attempted publication in CIKM 2014.

Full details on the project including all of our references, figures, and tables can be found in the above mentioned paper.

## Table of Contents

## Executive Summary

Specific field methodology and models cannot be an afterthought when designing, developing, or administering any kind of technology or system. However, the mass amount of techniques and options can be both overwhelming and confusing leading to the selection of incorrect or insufficient techniques. For an example in the security field, choosing an inadequate methodology can have harmful repercussions including everything from cyber-attacks to illegal data access and retrieval of private information. The solution is a metamodel that combines the most recent techniques and options categorized by common fields and concerns and presented to allow for a user to weigh the benefits, negatives, and particular circumstances needed to meet the unique needs of the user's system or environment. This metamodel would be of particular use for teaching and the sharing of knowledge. Contrary to some models which only present a high level overview, MOSAIC, is our example section of such a metamodel that will guide the user through the learning of and selection of analysis techniques and new security mechanisms. We provide the background and format for such a metamodel, our process for the selection of the security areas we focused on, and the example proof of concept, MOSAIC, Model of Securing Application Information Confidentiality.

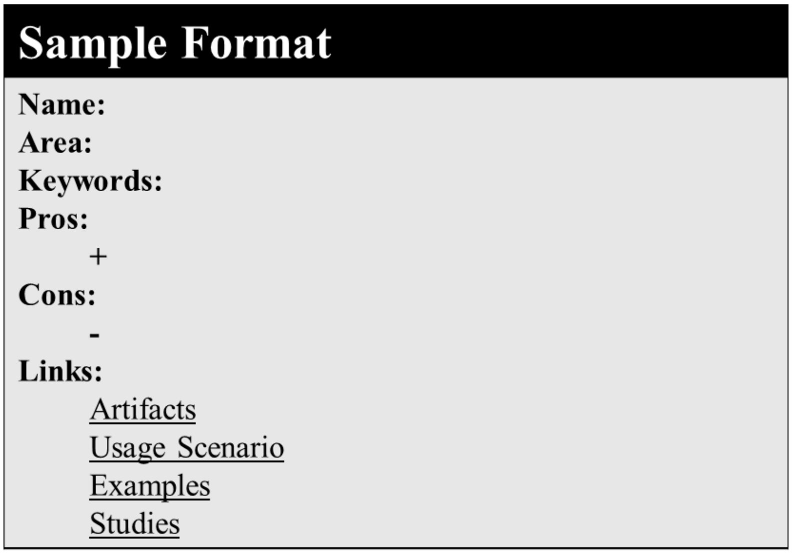
This research consisted of several phases including background research, metamodel development, a proof of concept, and the creation of a future evaluation study. To begin with the background research and metamodel development, we discussed the contributions of a metamodel for knowledge sharing and learning. This included its wide application to different domains and the development of a uniform format for easy comparison of entries and to aid in usability. Next, we addressed the features of a digital library system to support our metamodel and described this based on the 5S model for depicting digital libraries. This included discussion

and consideration for preservation and expansion which are highly linked to the envisioned digital library system and the adoption and use of our metamodel.  Another pivotal section for our progress consisted of a look and model creation of the relevant stakeholders for our metamodel.  This included stakeholder categorizations and their prospective contributions and utilization.

There were also many stages to the development of our proof of concept, MOSAIC, Model of Securing Application Information Confidentiality.  Our initial stage included research of key security domain concepts.  We compiled a list of over fifty areas of common concerns related to security and information confidentiality.  This list was utilized to target our four main areas of private information retrieval, access control, data classification, and threat modeling.  This decision was based on graph analysis of two co-occurrence graphs based on the security concepts scaled and weighted by the strength of their relation to each other.  The graphs included co-occurrences by frequency and by cluster.  Further area research was done to gather basic research of some important concepts and knowledge for each of these four main areas for use to place within our metamodel for the proof of concept and a walk through scenario.  Finally, the evaluation possibilities were outlined and ready for the next phase of execution to include use of the model by a domain expert to organize coursework for one unit of a course to be utilized by students who have given consent.   This will be compared to another unit of similar difficulty taught in the conventional methodology.

## User’s Manual

Our metamodel is meant for the addition as well as extraction of information for sharing knowledge.  Below is a guide for understanding the metamodel format and structure for the creation and extraction of information into the envisioned digital library system.



**Sample Format**

**Name:**
**Area:**
**Keywords:**
**Pros:**
+
**Cons:**
-
**Links:**
Artifacts
Usage Scenario
Examples
Studies

**Header bar:** The header is in the format of Subarea: Unique title

Example: *Private Information Retrieval: PIR-Tor*

**Name:** Unique title of the concept or work

Example: *PIR-Tor*

**Area:** Where this work resides in the metamodel system structure including Domain/Model Categorization/Subarea

Example: *Security/MOSAIC/Private Information Retrieval*

**Keywords:** User provided and categorization relevant key terms

Example: *PIR, security, anonymous communication*

**Pros:** These are provided benefits and anecdotes based on the research or expert findings and meant to be further supported through the provided linked materials. Indicated with a "+" symbol.

Example: + *A scalable client-server approach to anonymous communication*

**Cons:** These are provided negatives or limitations based on the research or expert findings and meant to be further supported through the provided linked materials. Indicated with a "-" symbol.

Example: - *Clients do not have global system view*

**Links:** The links section is meant to be an open area to provide further needed support, evidence, and materials to allow for the support of the pros and cons, provide the user with further in-depth information, and allow for the addition of concepts which cannot be easily portrayed in the pros and cons format.

Further information important for the creation of content into our metamodel is the selection of questions which can lead the user and aid their decision making by either providing the user with the information or eliminating it as a viable option if it is absolutely not applicable. Ultimately it is up to the user to decide the best fit. However, if there are determining criteria for applicability these should be provided in the form of questions. For example, if a researcher's work is only applicable if the user is dealing with security threat analysis of an existing and not an envisioned system, then an included criteria question for display of this material to the user could be "Do you have an existing system for analysis?." The "yes" or "no" response would then determine if the relevant card in the proper format is returned as a viable option to the user to explore or not.

# Developer's Manual

## History and Phases

### Phase 1: Background Research

Motivations<br>
Domain Applications<br>
Contributions

### Phase 2: Metamodel Development

Uniform Format<br>
Digital Library Envisioned System<br>
Preservation & Expansion<br>
Stakeholders Model

### Phase 3: Proof of Concept

Security Domain Concept Overview Research<br>
Preliminary Modeling with Co-occurrence Graph Analysis<br>
Security Domain Four Major Subset Areas Research<br>
MOSAIC Proof of Concept Scenario and Walkthrough

### Phase 4: Evaluation Plan

- Domain expert to utilize metamodel to organize unit of coursework
- Consenting students learn two units of equal difficulty
    - Metamodel Unit
    - Traditional Unit
- Evaluation based on student feedback and performance

## Inventory of Files

### Preliminary Modeling Co-Occurrence Chart Coding

Our preliminary modeling involved the selection of many facets of the security field and selecting four main areas for our proof of concept. The application code, completed with the scaled and inputted security concepts for modeling is included as files to the project VTechWorks repository.

**Co-Occurrence.zip** - The zip file containing the visualization files
- **occurrence.html** - The HTML file for the visualization
- **js/occurrence.js** - The JS file for the visualization
- **css/occurrence.css** - the CSS file for the visualization
- **data/occurrence.json** - the data file for the visualization

### Stakeholder Model XML

The stakeholder model outlined the four major stakeholders both adding and receiving information from our metamodel envisioned system. This model was created utilizing the xml modeling program draw.io. An xml file included in the project VTechWorks repository can be opened in the online accessible draw.io application and edited.

**StakeholderModel.png** - The static image
**StakeholderModel.xml** - The xml file for draw.io editing

### Midterm Slides

The midterm presentation slides utilized for the Virginia Tech CS6604 course are included in the VTechWorks repository. These slides cover the work related to Phase 1 and part of Phase 2 and Phase 3.

**ProjectMidtermMOSAIC.pdf** - The static presentation slides
**ProjectMidtermMOSAIC.pptx** - The editable powerpoint

## Final Slides

The final presentation slides utilized for the Virginia Tech CS6604 course are included in the VTechWorks repository. These slides cover the work related to all of the Phases. This brief presentation covers examples and the major contributions of the work, however, are not an exhaustive reference for the work. Full documentation can be found in the CIKM paper.

**ProjectFinalMOSAIC.pdf** - The static final presentation slides
**ProjectFinalMOSAIC.pptx** - The editable final powerpoint

## ACM CIKM 2014 Paper (Not Yet Submitted)

A conference paper has been written in the ACM format for potential submission to CIKM 2014, the International Conference on Information and Knowledge Management. The paper provides details through all phases of the project up to the design of further evaluation of the metamodel through its use in a classroom.

- **CIKM.pdf** - The static paper
- **CIKM-ACM-2014.zip** - The zip file containing all LaTex files
  - **CIKM.tex** - The LaTex file
  - **CIKM.bib** - The BibTex file
  - **CIKM.pdf** - The generated pdf
  - **acm_proc_article-sp.cls** - The ACM class file
  - **AC1.png** - Access Control Card 1
  - **Cluster.png** - Co-occurrence graph by cluster
  - **DC1.png** - Data Classification Card 1
  - **Flow.png** - The flow model graph
  - **Frequency.png** - Co-occurrence graph by frequency
  - **Overview.png** - Model Overview Image
  - **PIR1.png** - Private Information Retrieval Card 1
  - **SampleFormat.png** - Sample Card Format
  - **StakeholderModel.png** - Stakeholder graph
  - **TM1.png** - Threat Modeling Card 1
  - **TM2.png** - Threat Modeling Card 2
  - **TM3.png** - Threat Modeling Card 3
  - **Figures.pptx** - Editable formats for the images

Auto Generated Files:

**CIKM.aux**<br>
**CIKM.bbl**<br>
**CIKM.blg**<br>
**CIKM.log**<br>
**CIKM.synctex.gz**

## Lessons Learned

### Timeline/schedule

The outline of our semester long project is presented below. We have learned that project phases often overlap and require re-iterations as progress is made. In addition, the evaluation of research is no trivial matter and proof of concepts are a start, but further investigation and time will greatly enhance the validity of even theoretical works. Our metamodel requires further time to fully evaluate its utilization with an IRB approved study exploring its use in a classroom setting. The phase breakdown is included below.

#### Phase 1: Background Research

January

#### Phase 2: Metamodel Development

January-March

#### Phase 3: Proof of Concept

February-May

#### Phase 4: Evaluation Plan

May

The evaluation plan was outlined to be done as future work to be able to gain IRB approval and to allow for the schedule and time of a domain expert to convert course materials into the format of our metamodel for use with the instruction of a unit of a course to be taught in addition to an equally difficult unit in the regular format. This will provide a comparison to receive consenting students' feedback and assess their performance in the two units.

## Problems

As mentioned, evaluation is not a step to be taken lightly. This proved to be a section and phase of our project which could not feasibly fit into our timeline for this semester. A future work extension had to be made. Further problems also involved deciding how in-depth of a proof of concept needed to be done to highlight the application without having to organize a mass amount of the available resources and findings from countless research papers and works.

## Solutions

The extension of the timeline past the semester was better suited to finding a domain expert and allowing for plenty of time for the organization of a unit's worth of course materials into the metamodel format and for the planning of release to the students. This also allows for the selection of a unit which is compatible in difficulty for comparison to the unit utilized for the metamodel. To solve the issue of selecting the amount of works and papers to be represented in the proof of concept, we scaled the example to a chosen subset of works in each of the selected four core areas of MOSAIC including Private Information Retrieval, Access Control, Data Classification, and Threat Modeling. The scenario was then chosen and explained through the use of this small subset.

## Acknowledgements

We would like to acknowledge several people who have provided invaluable guidance and support to us throughout this semester. Dr. Edward A. Fox has welcomed our questions and challenged us to strive for excellence by considering key research questions related to the design and evaluation of our metamodel. Randy Marchany, the director of the IT Security Lab at Virginia Tech has graciously committed his time and acted as our client and domain expert. Noha El Sherbiny has acted as our prior research contact and directed our thinking towards the area of security within the realm digital libraries.