

## UALR Master of Science in Information Quality Capstone Project: Requirements and Guidelines

Students pursuing the Master of Science in Information Quality may choose to complete the 33-hour program by either a 6-hour master's thesis or 6-hour capstone project. This document outlines the requirements and guidelines for students selecting the **capstone project** option.

### Table of Contents

UALR Master of Science in Information Quality .....	1
Capstone Project:.....	1
Requirements and Guidelines.....	1
The Capstone Project Process.....	3
Minimum Qualifications .....	3
Courses Completed .....	3
In Good Standing.....	3
Faculty Advisor .....	3
Sponsoring Organization and Supervisor Guidelines.....	3
Qualifying Organizations.....	3
Employers can be Sponsors .....	3
Sponsor Supervisor .....	3
Qualifying Projects .....	4
Project Characteristics .....	4
Must Involve Data and Information.....	4
Must Address Information Quality Dimensions.....	4
Projects are Unpaid.....	4
Project Work Hours.....	4
Two-Semester Projects versus One-Semester Projects.....	4
Project Proposal Guidelines .....	5
Formatting Guidelines .....	5
Writing Style.....	5
Proposal Sections.....	5
Cover Page .....	5
Project Description.....	6
Impact on Information Quality .....	6
Approach.....	6
Deliverables.....	6
Technology.....	6

Dependencies and Limitations.....	6
Ethical/Privacy Concerns.....	7
Project Timeline .....	7
Enrolling in Project Hours .....	7
Project Execution .....	7
Final Project Report Guidelines .....	7
Formatting Guidelines .....	7
Writing Style.....	8
Report Sections.....	8
Cover Page .....	8
Executive Summary.....	9
Background .....	10
Quality Assessment at Start and End of Project .....	10
Description of Project Activities.....	10
Conclusion.....	10
Acknowledgement .....	10
Appendices.....	10
Oral Defense .....	11
Scheduling.....	11
Defense Attendees.....	11
Remote Presentation .....	11
Presentation Content and Format .....	11
Equipment.....	13
APPENDIX A - EXAMPLE PROJECT PROPOSAL.....	14
Project Description.....	15
Problem:.....	15
Proposed Solution:.....	15
Objective: .....	15
Impact on Information Quality .....	15
Approach.....	16
Deliverables.....	16
Technology Description .....	16
Problems and Limitations .....	16
Ethical and Privacy Issues: .....	16
Project Timeline .....	17

## The Capstone Project Process

The MSIQ Capstone Project process comprises the following steps:

1. Meet the minimum qualifications for starting a project
2. Identify a faculty advisor
3. Identify a sponsoring organization and supervisor
4. Identify an appropriate information quality project within the sponsoring organization
5. Submit a project proposal describing the project
6. Upon approval of the proposal, enroll in project hours
7. Complete the project
8. Upon completion of the project, submit a final project report to the MSIQ committee
9. Upon approval of the final report, make an oral defense of the project to the MSIQ committee and sponsor supervisor

## Minimum Qualifications

### *Courses Completed*

To be eligible to begin the capstone project you must have completed at least 9 hours of the MSIQ program. These 9 hours must include:

INFQ 7303	Principles of Information Quality
INFQ 7342	Information Quality Tools
Either INFQ 7367 or INFQ 7322	Information Quality Policy and Strategy  Information Quality Theory

### *In Good Standing*

You must also have regular admission status and otherwise be in good standing in the MSIQ program before starting a project.

## Faculty Advisor

Every capstone project must have a faculty advisor. A project faculty advisor must be a member of the MSIQ faculty and must agree to serve in this role.

## Sponsoring Organization and Supervisor Guidelines

### *Qualifying Organizations*

The next thing that you must do is to find an organization willing to sponsor your project. The sponsoring organization must be external to the UALR Department of Information Science. The organization must agree to sponsor the project. The sponsoring organization may be another UALR academic or administrative department, a business, a government agency, or a non-profit organization.

### *Employers can be Sponsors*

If you are employed either full-time or as an intern, your employer can qualify as a sponsor as long as the project proposal meets the required guidelines.

### *Sponsor Supervisor*

In addition to identifying a sponsor, you must also identify a sponsor supervisor within the organization. A sponsor supervisor is an employee of the sponsoring organization who will agree to supervise your project and provide input to your advisor on your performance.

## Qualifying Projects

### ***Project Characteristics***

Information Quality is a service science. Accordingly, by fulfilling your MSIQ requirement using the capstone project option you must address and improve some aspect of information quality. This includes producing a final report along with other deliverables appropriate to the project. Project sponsors are expected to help the student identify an information quality problem within the sponsor's organization and work with the student's faculty advisor to develop a plan for addressing the problem.

### ***Must Involve Data and Information***

The primary goal of any project should be for you to have an impact on information quality using the techniques that you learned in your coursework. You should be able to identify an information quality problem, then evaluate, measure, and improve the quality. While technology plays a primary role in information processing, the core focus should be on data and information. Technology improvement may be an important aspect. For example, a project to migrate information stored in a spreadsheet to a relational database in order to improve the dimensions of manipulability and accessibility of the underlying information. However, every project must at its core be related to data and information, and the data and information quality being improved must be clearly identified in the proposal.

### ***Must Address Information Quality Dimensions***

Your project must address and improve at least 3 dimensions of information quality. At least two of these dimensions must have quantitative measures or metrics defined as mathematical formulas. In addition, you should measure the level of quality in all of the dimensions you have identified at the start of the project and end of the project.

### ***Projects are Unpaid***

Unless you are an employee or intern of the sponsoring organization, the sponsor is not expected to pay for work related to the project.

### ***Project Work Hours***

A viable project should comprise 200-300 working hours. If you plan to complete the project in one 15-week semester should expect to spend approximately 15-20 hours per week actively working on the project. Accordingly, a two semester project would necessitate 7-8 weekly working hours to complete.

### ***Two-Semester Projects versus One-Semester Projects***

You are strongly encouraged to complete your projects over two consecutive semesters. It is very difficult to complete an entire project process of appropriate scale within the beginning and end of one semester. You are encouraged to begin your project in one semester and complete the project including the final report and oral presentation in the following semester. For administrative purposes, the final report must be submitted at least 2 to 3 weeks in advance of the oral presentation which must take place before the end of the semester. This leave only 10 to 11 weeks to identify and complete the actual project work and any unanticipated delays. If the project process is not complete by the last day of class in a semester, then you will be required to re-enroll in INFQ 7386 in each following semester until the project is complete. These additional hours are required for incomplete projects even if you have already completed 6 hours of qualifying project hours in previous semesters.

## Project Proposal Guidelines

Your project proposal must be approved by the MSIQ Faculty Committee before you can enroll in either the INFQ 7386 or INFQ 7686 project course.

### *Formatting Guidelines*

Please follow these formatting guidelines

- Letter page size in portrait orientation
- 1" top, bottom, and side margins
- Single column, with single or 1.25 line space
- 11 or 12 pt. font
- Use only black or dark blue font color
- Left justify paragraphs except for cover page
- Signal a new paragraph by either indenting the first line of a new paragraph or spacing between paragraphs
- Number all pages in the body of the proposal
- Figures and tables are encouraged, but all figures and tables should be numbered and captioned.
- Captions for figures and tables should fit on one line.
- All numbered and captioned tables and figures must be referenced by their number at least one time in the body of the proposal.
- All tables and figures must fit on one page. If an important table or figure is longer than one page should be included as an appendix to the proposal.
- Follow consistent indentation of headings and following text throughout the proposal
- Do not split headings and following text between pages

### *Writing Style*

- Use short sentences in direct, first-person language, e.g. "In this project I plan to design and ..."
- Consider rewriting sentences longer than 2 lines.
- Avoid paragraphs of only one sentence
- Do not try to control page breaks by spacing with paragraphs. You should use a page break command
- Spell out "Figure", not "Fig"
- When citing a specific figure or table you should always capitalize e.g. instead of "figure 1 shows that ..." write "Figure 1 shows that ..."
- Avoid the use of the passive voice. Try to use active first person, active voice.
- Always give the complete spelling of an acronym the first time you use it, e.g. "Master data management (MDM) is an important component of ..."
- Have someone else proofread your proposal for spelling, grammar, and punctuation. If you are not a native English speaker, you should have a native English speaker do the proofreading.

### *Proposal Sections*

#### **Cover Page**

Your proposal must start with a cover page. The following items should appear on the cover page in centered paragraphs

- Start with the words "Project Proposal for the Master of Science in Information Quality Program"

- Project Title. The title should be descriptive of the project and must include the word “Quality” and the word “Information” or “Data”
- Your Name
- Proposal Submission Date
- Expected Project Completion Date
- Your Faculty Advisor’s Name
- Name of your Sponsoring Organization
- Name, title and contact information for your Sponsor Supervisor

### **Project Description**

Include a high-level (1- to 2-paragraph) description of the sponsoring organization followed by a

- **Problem Statement** that describes the information quality issues and the impact that they are having on the organization. Include any previous or current efforts by others in the organization to address these issues.
- **Objective Statement** that describes your goals for the project.
- **Proposed Solution** to the problem.

### **Impact on Information Quality**

- Identify at least 3 dimensions of information quality to be improved in the project.
- Indicate how information quality will be improved, along with a metric for how the level of quality will be measured in each dimension (both at the start and end of the project). While some assessments can be anecdotal or subjective in nature, at least two of the measurements be a quantitative calculation based on a mathematic formula, .e.g. “For each file of N records, if M is the number of missing values for item X, then  
Completeness of Item X =  $100 * (1 - M/N)$ ”.
- **WARNING:** In your metrics: do not claim to measure “accuracy” unless you are verifying the correctness of the data from the original source or another independent source know to be accurate. Don’t confuse validation with verification. If you are cleaning the data to make all of the dates in the same format (xx/xx/xxxx), that is validating the dates. If these are dates-of-birth, then you are not measuring the accuracy unless you or someone else is going back to the original source (the person) to verify that is the correct date-of-birth or using some other independent source to verify that the date (such as birth certificate). Be careful with the use of the word “accuracy” in the proposal and in the final report.

### **Approach and Methodology**

Describe the major activities you plan to carry out in solving the problem. If possible, you should show how these steps follow one of the IQ methodologies such as the MIT Total Data Quality Management (TDQM) methodology, Dr. Sebastian-Coleman’s Data Quality Assessment Framework (DQAF), or McGilvray’s Ten Step Method, or the Six-Sigma DMAIC.

### **Deliverables**

Describe the major deliverables of the project including the artifacts you will leave with the sponsor, e.g. software, new procedure or process. Don’t forget to include “Final Project Report” and “Slides for Project Defense”.

### **Technology**

To the extent you know at the time of the proposal, list the software, tools, systems, or other technology you expect to employ in the project.

### **Dependencies and Limitations**

Describe any known issues that may delay or prevent a successful completion of the project. These might include delays in getting software, approval for system access, or dependencies other people in the sponsoring organization completing certain tasks critical to your project.

### **Ethical/Privacy Concerns**

Describe any ethical/privacy concerns, challenges or limitations you may encounter in the project. Will you be working with sensitive information like personally identifiable information (PII) or health information? If so, how will it be secured and protected? Will you be required to sign any non-disclosure or security agreements before you can begin work?

### **Project Timeline**

- Provide a table corresponding to the appropriate number of semester weeks. There are approximately 15 working weeks per semester. A one-semester project should fit into 15 weeks, a two-semester project into 30 weeks.
- You may find it helpful to start at the end of the intended graduating semester and work backwards. Include at least 3 weeks for compiling your findings and producing the Final Report, 2 week for approvals, and 2-3 weeks for revisions before scheduling your oral presentation.

***A Proposal example is included as Appendix A of this document.***

## **Enrolling in Project Hours**

Once a student has an approved project proposal on file, he or she can enroll in project hours. An MSIQ capstone project requires 6 credit hours to complete. Students can successfully complete the 6-hour project requirements by choosing one of three different paths:

- A. Two (2) 3-hour courses — **INFQ 7386**, Graduate Project, in two successive semesters.
- B. One (1) 6-hour course — **INFQ 7686**, Graduate Project, in a single semester.
- C. One (1) 3-hour course — **INFQ 7386**, Graduate Project, and  
One (1) 3-hour course — **INFQ 7391**, Cooperative Education in Information Quality.

Note: In Option C ideally the work done during the Cooperative Education course is for your project sponsor and directly related to your project. However, this is not required. INFQ 7391 will satisfy 3 hours of the project requirement regardless of the employer and the duties performed or work completed during the internship.

It is very important for you to understand that the MSIQ Capstone Project is governed by a consecutive enrollment policy. This means that once you begin taking project hours, you must continue to enroll in project hours each semester (including the summer semester) until the project process is complete.

## **Project Execution**

You should give your faculty advisor regular updates on the progress of the project. It is important for you to alert your advisor when problems or issues arise that could delay or prevent the completion of the project. Not all projects are expected to follow the project plan exactly as outlined in the project proposal. A change in direction or delay is not necessarily bad as long as there is a justifiable reason. On rare occasions projects are terminated before completion. In these cases, your faculty advisor and graduate coordinator will work with you to restart with a new project.

## **Final Project Report Guidelines**

The final report is one of the most important components of the capstone project process. It should accurately and completely describe what the student accomplished in the project.

### ***Formatting Guidelines***

These are the same as for the project proposal

- The total number of pages in the final report including appendices must not exceed 50 pages. Longer is not necessarily better. Your goal should be a concise, yet complete report that tells the story of your project.
- Letter page size in portrait orientation
- 1" top, bottom, and side margins
- Single column, with single line space
- 11 or 12 pt. font
- Use only black or dark blue font color
- Left justify paragraphs except for cover page
- Signal a new paragraph by either indenting the first line of a new paragraph or spacing between paragraphs.
- Number all pages in the body of the report.
- All figures and tables are encouraged, but all figures and tables should be numbered and captioned.
- Captions for figures and tables should fit on one line.
- All numbered and captioned tables and figures must be referenced by their number at least one time in the body of the report.
- All tables and figures must fit and be placed on one page. If an important table or figure is longer than one page should be included as an appendix to the report.
- Follow consistent indentation of headings and following text throughout the report.
- Do not split headings and following text between pages.

### ***Writing Style***

- Use first-person, past tense language. e.g. "In this project I designed and implemented ..."  
Most of the final report should be written in the past tense describing activities you have already completed. Be careful about simply copying text from your proposal into the final report.
- Write in short, direct sentences. Consider rewriting sentences longer than 2 lines. Avoid paragraphs of only one sentence.
- Do not try to control page breaks by spacing with paragraphs. You should use a page break command.
- Spell out "Figure", not "Fig"
- When citing a specific figure or table you should always capitalize e.g. instead of "figure 1 shows that ..." write "Figure 1 shows that ..."
- Avoid the use of the passive voice. Try to use active first person voice.
- Always give the complete spelling of an acronym the first time you use it, e.g. "Master data management (MDM) is an important component of ..."
- Have someone else proofread your report for spelling, grammar, and punctuation. If you are not a native English speaker, you should have a native English speaker do the proofreading.

### ***Report Sections***

#### **Cover Page**

The final report must start with a cover page. The following items should appear on the cover page in centered paragraphs:

- Project Title: The title should be descriptive of the project and must include the word "Quality" and the word "Information" or "Data". As things change during the course of a project, the title on the final report maybe different than the title on the original proposal to reflect a change in direction, focus, or understanding.



- The subtitle should be “Submitted in Partial Fulfillment of Requirements for the Master of Science in Information Quality”
- Your Name
- Report Submission Date
- Your Faculty Advisor’s Name
- Name of the Sponsoring Organization
- Name and title of the Sponsor Supervisor

### ***Executive Summary***

The Executive Summary must be the first section of the final report. It is the most important part of the report. Don’t make it all background, it should clearly state the results of the project. The executive summary should cover three topics in the following order:

**Background:** First, start with a brief background. For example, “XYZ is a (*company/agency/etc. - describe the sponsor and what the sponsor does, but be brief*). The information quality problem experienced by XYZ is that (*describe the problem or problems that were addressed by this project*). The goal of this project was (*use the past tense in the summary because you should be describing things you have already done*) to (*design/build/implement/modify/etc. use a verb here*) a (*system/process/database/... – describe the main project deliverable*) for XYZ that (*describe briefly how the deliverable addresses the sponsor’s IQ problem*).”

**What you did:** Next, describe what you did in the project. For example, “In this project, I (*designed/ programmed/ analyzed/ implemented/ tested/etc. – a verb that describes specifically what you did*) a (*dataset/GUI/SQL procedure/Run Book/... some deliverable*) so that (*describing the purpose of what you did*).”

This sentence should be repeated for each major task. For example, if you profiled a data source, then developed and tested programs or scripts to cleanse and standardize the data.

**The Benefits to the Sponsor:** Finally, describe the benefits of your project. For example, “As a result of my project (*describe a direct benefit to XYZ, e.g. the time required to receive a file and load it into the database has been reduced from an average of 2 days to 4 hours. In addition, the number of non-conforming data items has been reduced from 60% to 5%.*)”

This sentence should be repeated for each major benefit. The metrics or statistics stated here should also appear in the body and in the Conclusion Section of the report.

#### **Other guidelines for the Executive Summary section include:**

- The Executive Summary must not be longer than one page.
- It should be broken into different paragraphs.
- It should recap the most important quantitative and qualitative results of the project.
- It should describe what you did in the project including measurements and demonstrating improvements.
- It should be written retrospectively. It should be about what you have already done (past tense) and not what you propose to do.
- Don’t save any surprises for the conclusion, disclose everything important in the executive summary.
- The last paragraph of the Executive Summary should be strong. It should be more quantitative and cite the most important improvement statistics from your report to back-up your claims of improvement.

## ***Background***

- Describe the information quality issues and their impact on the organization.
- Describe any previous and/or current efforts by others in the organization to address these issues.

## ***Quality Assessment at Start and End of Project***

It is critical that the report clearly shows the improvement of information quality in some quantitative way. Quantitative measures of information quality taken at the both the start and the end of the project should demonstrate this improvement. Although some assessment measures can be anecdotal or subjective in nature, at least two of the measurements should be quantitative.

For the quantitative measures:

- Identify the dimensions or characteristics of the information quality being measured.
- Describe the methodology for taking the measurements.
- For smaller sets of measurement data, include in the body of the report, otherwise include in an appendix.

## ***Description of Project Activities***

In this section, clearly explain which parts of the project were completed by you, and which parts were completed by other participants.

It is always helpful to include figures to illustrate your results and other work. Things like a schema for the new database, a new process flow, screen shots of a new data entry GUI, or an example business requirement document.

Be quantitative: The review committee expects to see tables and graphs of data quality measurements. The metrics that you defined in your proposal should be calculated and shown in the report with measurements at the beginning of the project, during the course of the project, and at the end of the project.

**WARNING:** If the data you are working with contains personally identifiable information (names, addresses, phone numbers), do not show these in your report. Be careful when including examples, tables, and screen shots. If you want to show a screenshot that has personal data you should edit out the identifying information or use dummy (not-real) information in your report.

## ***Conclusion***

The conclusion should:

- Clearly state the benefits of the project to sponsor.
- Recap and summarize your key improvement statistics for each dimension and focus on the benefits for the company that you achieved. You can add a table or a scorecard for this purpose.
- Focus on the results of the project and not the background or activities of the project.

## ***Acknowledgement***

It is always a good idea to acknowledge the help you received. If you worked with other people, and part of their work is included in the report, then be sure to acknowledge the specific work that they did.

## ***Appendices***

Appendices allow you to include detailed information that would be distracting in the main body of the report. Examples of items you might have in an appendix include things like reference tables, questionnaires, SQL scripts, and programming code.

## Oral Defense

The final step in the capstone project process is to make an oral defense of your project to the MSIQ committee.

### ***Scheduling***

Scheduling is coordinated by the Information Quality Administrative Assistant well in advance of the end of the semester. All defenses must take place on or before the last day of class. However, your presentation cannot take place before your final report has been approved. Failure to get approval for your final report will result in the cancellation of your defense.

### ***Time Limitation***

Each defense presentation is limited to a total of 50 minutes including introduction, presentation, demonstrations (if any), and questions and answers by the committee.

You must leave the last 10 minutes open for questions and answers by the committee, therefore the total time for your part of the presentation must be no more than 40 minutes. Manage your time carefully.

### ***Defense Attendees***

**Sponsor Supervisor:** In addition to yourself and the members of the MSIQ Committee, you should invite your Sponsor Supervisor. Although his or her attendance is not mandatory, it is always beneficial to have the supervisor present to speak about the benefits of the project to the organization. If the supervisor is not present, then the faculty advisor will solicit feedback on your performance by email.

**Guests:** The oral project defense is a public event and you may invite guests to attend your defense.

### ***Remote Presentation***

Local students living within 50 miles of campus are expected to make their oral defense in person. Remote students should make arrangements at the time of scheduling for a live webcast presentation using Blackboard Collaborate or another mutually agreeable webcasting technology.

### ***Presentation Content and Format***

The oral presentation should be supported by PowerPoint slides that illustrate the major points of the project. You are expected to explain the content of the slides, not to read the slides. The slides should focus on the key elements of your project. The following is a suggested layout for your PowerPoint deck.

#### **Suggested Layout**

- Title Slide
  - Title of your project as it appears on your final project report
  - Your name
  - Name of the sponsoring organization
  - Name of the person in the organization supervising your project
  - Name of your faculty advisor
- Background Slide
  - One slide describing your sponsoring organization, Where is it? What does it do?
- Problem Statement
  - One or two slides that explain the problem you addressed in your project
  - Why was solving this problem was important the organization?

- Project Objectives
  - List all of the things that you originally planned to accomplish and deliver during the course of your project. List everything in the original plan even if some of them were not accomplished or changed during the course of the project.
- Project Plan or Approach
  - Explain your step-by-step strategy for solving the problem.
  - Use graphics where appropriate
- Activity Slides
  - One or more slides that explain your activities (what you did) in the project.
  - This is where you tell the story of your project.
  - Be sure to credit others who may have helped you
  - Use graphics where appropriate
  - Point out any problems you encountered and how you overcame these problems, or how these problems caused the project deliverables to change from the original design
- Project Accomplishments
  - List and explain what you accomplished.
  - Try to be quantitative (use numbers and statistics) as well as qualitative in your descriptions
  - If the project was to improve data quality, this is where you show the before and after measurements of the dimensions of DQ that you improved.
  - Use charts, graphs, and other graphics where appropriate
  - Explain any differences between your original plan of accomplishments and what you actually delivered, and what caused these differences.
- Conclusion
  - Explain how the project benefited the sponsor.
  - Explain what you learned from doing this project
- Acknowledgements
  - (Optional) Special thanks to anyone who was especially helpful to you
- Questions? Slide
  - A slide to leave on the projection screen while you answer questions from the committee

#### **Overall Guidelines for PowerPoint Presentation**

- Don't prepare too many slides! You should be able to tell your story with about 15 slides, don't go over 20 slides, otherwise you will exceed your 40 minute time limit for presenting
- Use graphics such as screen shots, charts, flow diagrams, and graphs as much as possible. Try not to make the entire deck one long set of bullet points.
- For slides where you need to use text, don't make the text too small or too dense. Any text should be in a font size large enough to be easily read by the attendees.
- Check your slides to be sure they do not to reveal any confidential information from your sponsor. For example, check screen shots to be sure they don't show any personally identifiable information.
- Only use dark color fonts on white or very light color backgrounds. Be careful about using color in your presentation. The colors rendered by the projector may look different than the way they appear on your computer screen. When this happens, some combinations of colored fonts and backgrounds can become unreadable.

#### **Live Demonstrations:**

Live demonstrations of systems or websites can be an optional part of your presentation. Demonstrations can be very helpful, but can also be risky. Be sure to test your demonstration in advance using the same equipment you will use in your presentation, and if possible, in the same room where your defense will take place. If you plan to give a demonstration, make allowances for

the extra time by reducing the number of presentation slides appropriately. In addition, it is also a good idea to prepare extra slides covering the same content as your demonstration in case your demonstration fails to operate properly.

### **Equipment**

- You are expected to bring your own laptop for the presentation. If you don't have one, please let the Information Quality Administrative Assistant know that at the time you schedule your presentation.
- Your oral defense will be scheduled in a room in the EIT building that has UALR WiFi Internet access and a digital projector.
- The projector will have a VGA input connector. Be advised! Many newer laptops do not have a VGA output connector. If your system does not have a VGA output connector, then you are responsible for bringing an appropriate adapter that will connect your laptop to the projector's VGA input. For example, if your laptop only has HDMI output, you need to bring an HDMI-to-VGA adapter to your presentation.
- Bring a backup copy of your PowerPoint presentation on a USB drive in case your system fails or is incompatible with the projector. This will allow your presentation be made using a different computer.

## **APPENDIX A - EXAMPLE PROJECT PROPOSAL**

Project Proposal for the  
Master of Science in Information Quality Program

### **Improving the Quality of Alaska Law Enforcement Information through Inter-Agency Data Sharing**

Submitted by: **Samantha Jones**

Faculty Advisor: **Dr. Emily Smith**

Sponsoring Organization: **Alaska State Police**

Sponsor Supervisor: **Sgt. Millard Fillmore**  
Chief Technology Officer, Alaska State Police

[mfillmore@aksp.org](mailto:mfillmore@aksp.org)

907-456-3898

Date Submitted: **8/20/2014**

Estimated Completion Date: **4/30/2015**

## Project Description

The purpose of this project is to help the Alaska State Police with the creation of a fusion center. A fusion center is simply a way of bringing together (“fusing”) disparate information about the same case that is spread across different law enforcement agencies. A fusion center is an effective and efficient mechanism to exchange information and intelligence among agencies. A Fusion Center can also help to maximize resources, streamline operations, and improve the ability to fight crime and terrorism by merging data from a variety of sources (Taken from the “Fusion Center Guidelines: Executive Summary”).

**Problem:** The U.S. Department of Justice's Global Justice Information Sharing Initiative sets out the guidelines of what a Fusion Center should be, but they did not offer any standards on how a Fusion Center should operate or how the data should be collected, used, or reconciled. The Alaska State Police has asked for help in evaluating the technologies that exist in this market space. The items of interest in the evaluation this includes Entity Resolution, Pattern Analysis, and Case Management.

**Proposed Solution:** My portion of the overall project is to help show the State Police how data from multiple sources can be consolidated to provide a more complete view. Possible sources of data include Police data from local city and county agencies, property taxes, DMV records, Marriage/Divorce records, civil court records, etc. Other possible sources of data include Department of Corrections (DOC), Alaska Crime and Information Center (ACIC), and the Department of Finance Administration (DFA).

**Objective:** The objective of my project is to demonstrate how the integration of data can be accomplished. In producing the report and other deliverables, I hope to provide the State Police with an understanding of how they can begin to assemble a Fusion Center from an IT perspective.

## Impact on Information Quality

The primary information quality issues in this project are related to the dimensions of

- **Completeness**—Officers in the field and case managers often do not have access to all the information about a suspect or person of interest.  
*Metric:* I will provide a form to the case managers that they will use to rate completeness for the information on a monthly basis. The metric will be calculated as the average of the rating over all case managers.
- **Access**—Important case information resides in many different agencies and databases is not readily available in one place.  
*Metric:* The number of agencies participating and number of records contributed on a monthly basis.
- **Timeliness**—Currently it can take several days to request and receive information from another agency.  
*Metric:* All of the completed information request forms that were sent to other agencies will be analyzed each month. For each completed request I will calculate the difference in days between the “request time” and the “delivery time”. The metric will be the average time difference calculated over all of the requests for that month.
- **Value-Added**—No single piece of information may be that important by itself, but when assembled into a complete picture, it could solve or prevent a crime.  
*Metric:* The percentage of cases closed each month as reported in the monthly statistics.

(NOTE: Other dimensions that may be relevant for a project include: Free-of-error (accuracy), believability, reputation, relevancy, amount-of-data, interpretability, representational consistency, conciseness of representation, manipulability, security)

## Approach

Because this project will be about the reconciliation and recognition of disparate data sources and differing data collection methods, I have requested access to a sample dataset(s) from the participating parties. If unable to gain access, schemas for the data will be requested to create simulated datasets. The activities will be carried out over a **30-week** (two-semester) timeframe.

The source data will be stored in a MySQL database. A “query engine” will access the different tables to simulate data acquisition from multiple data sources. The results will be returned, cleansed and standardized, then presented to the user. Counts will be produced by running queries against the database.

## Deliverables

1. **An Information Product (IP) Map** (or series of maps) of the proposed data sources. This will allow the Alaska State Police to see how the data can be integrated and where problems can occur in the integration process.
2. **A written plan** of how the disparate data can be integrated and reconciled. This plan will be based on what is learned from the IP Maps and the interviews with the participants.
3. If time permits, **a simple prototype** demonstrating the entity resolution and data integration process.
4. Final Project Report
5. Slides for Oral Project Defense

## Technology Description

The software program language and specifications that I plan to use in this project are:

- SAS dfPower Studio—I will use this software as the entity resolution tool for this project.
- MySQL 5.0—a free Relational Database. I will use it to store the demo data.
- Global Justice XML (GJXML)—an XML specification for allowing interoperability between data sources containing law enforcement data.
- National Information Exchange Model (NIEM)—the new specification that will supersede GJXML as the exchange mechanism.
- I will use Java to develop any software tools needed for the project.

## Problems and Limitations

The major challenges I see with the project are around the data, both the availability and the quantity of data. If the data are not available, then the schemas for the participating agencies will be requested, and I will have to create simulated data based on these schemas. That may make the project go longer than anticipated. However, by creating my own simulated data, I can control the size of the datasets and avoid overloading the test system, thereby solving the quantity problem.

## Ethical and Privacy Issues:

There are several ethical issues that I will have to deal with in this project. The first is access to information. I will first have to sign a Confidentiality Agreement and undergo a standard State Police Background Check. Also, the data that will be used will be restricted to Freedom-Of-Information (FOI) data and other publicly available data.



## Project Timeline

First Semester	
Week 1	Secure necessary forms; Request data and/or schemas
Weeks 1 – 9	Initial research on Fusion Centers, Entity Resolution and interviews with participants. Start IP Maps
Weeks 10 - 12	Preliminary data analysis
Weeks 13 - 15	Develop data validation rules and data cleansing routines
Second Semester	
Weeks 16 - 18	Complete IP Map; Start Work on Prototype
Weeks 19 - 24	Complete Prototypes
Week 25	Submit First Draft Report for Approval; Schedule oral defense
Weeks 26 - 28	Finish Final Project Report
Week 29	Oral Defense