Data Fusion

Prepared by:
Richard Davies
University of Ottawa Heart Institute
Sonny Lundahl
Amita Corporation

Scientific Authority:
Norman Yanofsky, 613-944-8161
DRDC Centre for Security Science

DRDC-RDDC-2014-C142
May 2014

IMPORTANT INFORMATIVE STATEMENTS

CRTI 08-190RD Towards an Operational Urban Modeling System for CBRN Emergency Response and Preparedness was supported by the Canadian Safety and Security Program (CSSP) which is led by Defence Research and Development Canada's Centre for Security Science, in partnership with Public Safety Canada. The project was led by National Research Council—Institute for Information Technology. Partners in the project include AMITA Corp., Auton Laboratory, Carleton University HOTLab, Carnegie Mellon University School of Computer Science, DRDC Valcartier, Grey Bruce Health Unit, Health Canada, Michigan Department of Community Health, Ontario Agency for Health Protection and Promotion Ottawa Hospital Department of Infectious Diseases and Infectious Control, Ottawa Public Health, PHAC, Queen's University Public Health Informatics, and University of Ottawa Heart Institute.
CSSP is a federally-funded program to strengthen Canada's ability to anticipate, prevent/mitigate, prepare for, respond to, and recover from natural disasters, serious accidents, crime and terrorism through the convergence of science and technology with policy, operations and intelligence.

Government of Canada Gouvernement du Canada

Data Fusion

Richard F. Davies, University of Ottawa Heart Institute, Ottawa, Ontario
Sonny Lundahl, AMITA Corporation, Ottawa, Ontario

Defence R&D Canada – Centre for Security Science

Project Closeout Report

Project Number CRTI 08-0190RD

November 9, 2012

A Partnership Led by Defence R&D Canada – Centre for Security Science

DEFENCE R&D DÉFENSE

Agriculture and Agri-Foods Canada • Atomic Energy of Canada Limited • Canada Border Security Agency • Canadian Food Inspection Agency • Canadian Nuclear Safety Commission • Canadian Security Intelligence Service • Communications Security Establishment • Defence Research and Development Canada • Environment Canada • Fisheries and Oceans Canada • Health Canada • Industry Canada • National Research Council • Natural Resources Canada • Privy Council Office • Public Health Agency of Canada • Public Safety Canada • Public Works and Government Services Canada • Royal Canadian Mounted Police • Transport Canada • Treasury Board Secretariat

Canada

Principal Author

Nov 9/2012

Dr. Richard F. Davies — Date
University of Ottawa Heart Institute

Co-Authors:

Sonny Lundahl, AMITA Corporation
Anu Pinnamaneni, AMITA Corporation
Susan McClinton, University of Ottawa Heart Institute
Berry DeBruijn, National Research Council Canada

Approved by

Nov. 9, 2012

Cheryl Lambert — Date

National Research Council Canada

Approved for release by

Nov. 9, 2012

Norman Yanofsky — Date

Chemistry and Biology Portfolio Manager CRTI Secretariat

Abstract

Data Fusion is a successfully completed Research and Development project funded by The Chemical, Biological, Radiological-Nuclear, and Explosives (CBRNE) Research and Technology Initiative (CRTI). Data Fusion focused on creating a reusable framework for Syndromic Surveillance solutions. It developed and delivered an adaptive process framework and software framework in conjunction with two domain-specific prototypes. These frameworks are extensible and can be configured to other domains and problems that could benefit from surveillance solutions.

The project team comprised of healthcare and technical partners, selected two disparate and important healthcare problems that could benefit from automated surveillance, and developed a solution prototype for each: 1) the detection of serious in-hospital disease outbreaks, and 2) the surveillance of harm related to illicit substance abuse. The goal of each prototype is to enhance early detection and present relevant information to responders to assist in their decision making process. The prototypes were tested on a retrospective dataset from multiple sources within the electronic health record and were shown to be effective and useful.

The prototypes and process framework integrate technological advances developed by the Data Fusion partners: text classification from NRC-ICT Interactive Information Group, data fusion techniques and specialized algorithms from DRDC Valcartier, data integration and management by AMITA Corporation, statistical analysis and display from STATACorp and geospatial mapping from DM Solutions. Non-technical contributions include epidemiological practice principles from multiple stakeholders in health care and public health and domain specific expertise from Infectious Disease and Health Canada, Office of Drugs and Alcohol Research and Surveillance, Controlled Substances and Tobacco Directorate.

The surveillance process and the technology utilized for the prototype development have been documented in a process and software framework. This framework provides a generalizable solution that is ready to be applied to new problems. The goal is to leverage existing expertise and technology and to reduce the effort required to establish automated surveillance.

Data Fusion delivered re-useable surveillance products to facilitate display and communication of data. These are in the form of epidemiological graphs and geospatial maps. Their specifications and re-useable scripts based on commercial-off-the-shelf and open source tools can be applied to new datasets.

The Data Fusion team has completed a proof of concept and developed a surveillance process and software framework that is ready to be applied to an important surveillance domain in real-time.

Résumé

Data Fusion est un projet de recherche-développement qui a été mené à bien grâce à une subvention de l'IRTC – Initiative de recherche et de technologie CBRNE (agents chimiques, biologiques, radiologiques, nucléaires et explosifs). Data Fusion devait permettre de créer un cadre réutilisable à des fins de solutions pour la surveillance syndromique. L'équipe du projet a élaboré et présenté un cadre de processus adaptatif et un cadre logiciel, ainsi que deux prototypes dans des domaines précis. Ces cadres peuvent être élargis et configurés en vue d'une application dans d'autres domaines ou à des problèmes pouvant bénéficier de telles solutions.

L'équipe du projet, composée de spécialistes en santé et de techniciens, a choisi deux importants problèmes de santé susceptibles de bénéficier d'une automatisation de la surveillance, puis a mis au point deux prototypes pour résoudre ces problèmes – le premier, pour le dépistage de graves foyers de maladies nosocomiales et le second, pour la détection des dommages associés à la toxicomanie. Dans chaque cas, le prototype devait rehausser un dépistage précoce et présenter l'information pertinente aux intervenants, afin de les aider à prendre les décisions adéquates. Les prototypes ont été testés au moyen d'un ensemble de données historiques provenant de nombreuses sources du dossier médical électronique. Les résultats ont prouvé leur efficacité et leur utilité.

Les prototypes et le cadre de processus intègrent des progrès technologiques réalisés par les partenaires du projet Data Fusion : le classement des textes par le Groupe de l'information interactive de TIC-CNRC; les techniques de fusion des données et les algorithmes spécialisés de RDDC Valcartier; l'intégration et la gestion des données d'AMITA Corporation; l'analyse et l'affichage des statistiques de STATACorp; ainsi que la cartographie géospatiale de DM Solutions. Parmi les contributions non techniques, mentionnons les principes d'épidémiologie appliquée venant de nombreux intervenants des secteurs de la santé et de l'hygiène publique, ainsi que l'expertise de spécialistes en maladies infectieuses et du Bureau de la recherche et de la surveillance des drogues et de l'alcool de la Direction générale des substances contrôlées et de la lutte au tabagisme à Santé Canada.

La méthode de surveillance et la technologie employées pour créer le prototype ont été documentées dans un cadre qui rassemble processus et logiciel. Ce cadre offre une solution généralisable qu'on peut appliquer à de nouveaux problèmes. L'objectif est d'exploiter l'expertise et la technologie existantes de manière à réduire les efforts nécessaires à l'automatisation de la surveillance.

Le projet Data Fusion a fourni des produits de surveillance réutilisables qui facilitent l'affichage et la transmission des données sous forme de graphiques épidémiologiques et de cartes géospatiales. Les spécifications de ces produits et les scripts réutilisables reposent sur des outils disponibles dans le commerce et de source libre, qu'on pourra appliquer à de nouveaux ensembles de données.

L’équipe de Data Fusion a effectué une validation de principe et a élaboré un cadre logiciel et de processus pour la surveillance, qui est prêt à être appliqué à un important domaine de surveillance en temps réel.

This page intentionally left blank.

Table of contents

List of figures

List of tables

Acknowledgements

This work was supported by Defence Research and Development Canada Centre for Security Science and the Chemical, Biological, Radiological/Nuclear, and Explosives Research and Technology Initiative (CRTI) under project CRTI 08-0190RD and the following participatory and advisory partners:

Development Partners

- University of Ottawa Heart Institute (UOHI)
- National Research Council, Information and Communications Technologies (NRC-ICT)
- Defense Research and Development Canada (DRDC – Valcartier)
- AMITA Corporation
- The Ottawa Hospital Data Warehouse (OHDW)
- The Kingston General Hospital (KGH)
- Ottawa Paramedic Service (OPS)
- Stata Corporation
- DM Solutions

End-user Partners

- Health Canada, Office of Drugs and Alcohol Research and Surveillance, Controlled Substances and Tobacco Directorate (Health Canada)
- Ottawa Public Health (OPH)
- Kingston Frontenac Lennox and Addington Health Unit (KFLA)
- Public Health Ontario (PHO)

Consulting Partners

- Children's Hospital of Easter Ontario Research Institute (CHEO RI)
- Queen's Public Health Informatics (QPHI)
- Toronto Public Health
- York Public Health
- Grey Bruce Health Unit
- Michigan Department of Community Health
- Carnegie Mellon University
- SilvaCorp Inc.

This page intentionally left blank.

1 Introduction

1.1 Surveillance

Surveillance can be broadly defined as monitoring an ongoing data stream to detect an unexpected event, monitoring its progress to track the effects of interventions, and improving ongoing situational awareness in order to inform or improve response capability. This definition is consistent with the PHAC Framework for Evaluating Health Surveillance systems, which acknowledges surveillance as tracking "the life blood of the flow of information in support of crucial decisions that impact on the lives of many citizens."

Technology solutions that enable automated surveillance are capable of improving response strategies at all stages in the time line of an event. Initially, they make it possible to detect an unexpected event at an earlier stage, thereby enabling a more timely and effective response. Subsequently, the same technology makes it possible to track the status of a known event to inform decision-making, measure the results of targeted interventions, and ascertain when an abnormal situation has returned to normal. At all stages of an event, these technologies are also capable of providing data to inform the public and improve their confidence that circumstances are as "under control" as possible and that the best possible decisions are being made. The ongoing situational awareness provided by surveillance systems is also useful in confirming that an unexpected event is not occurring, and in "pre-conditioning" responders to make the most appropriate decisions and responses when a new or unexpected event does occur.

Surveillance is widely used in health care related settings, and it is in this context that this project has been developed. However, the same technology and methods are applicable to any area where counting and analyzing things are useful in predicting events, informing decisions and improving outcomes. The same technologies and methods, when used in other areas, go by different names. These include: "cybemetics", "continuous decision-making", loop of decision-making/intervention", "command and control", "evidence-based decision-making" and "intelligence analysis capability". The framework developed during this project was intended from the outset to include all of these areas.

1.2 The Need for a Generalized Framework

The processes and technology that are needed to establish and conduct surveillance can be collectively considered a "Surveillance Solution". Such solutions are needed in response to a specific need. Often the situation is unexpected and the need is urgent. The resulting time pressure will, understandably, lead the responders to do what is most familiar to them: repeating patterns based on previous training and experiences. Such "silo solutions" will allow responders to get something up and running relatively quickly, but have significant short and long-term disadvantages, including:

- Using the wrong data. Time pressures can unfortunately result in an early commitment to use data that is easily accessible even though it is information-poor.
- A key success factor may be overlooked, resulting in unplanned delays in development and deployment.
- The data management and analytic tools that are chosen may not be the best available.

- The wrong tools may be used for specific tasks.
- The solution developed does not generalize well.
- The solution developed generalizes well enough for systematic mistakes to be repeated each time a new surveillance solution is needed.
- The solution does not take advantage of the opportunity for cross-fertilization across surveillance experts in different subject matter areas.

The goal of the Data Fusion Project was to develop a generalized framework for surveillance that can be used broadly. There are several advantages to this approach, including the following:

- Explicit processes can be developed for identifying the right data, accessing it, and putting it into a format that is amenable to statistical analysis.
- Technological tools can be developed that are generalizable and broadly applicable. It will also be possible to apply improvements and "lessons learned" from each solution that is developed to subsequent projects. In the long term, this will result in the accumulation of a valuable body of knowledge.
- A generalized framework will identify key expertise and allow it to be brought to bear on the project at hand. In this way the most suitable state of the art tools can be identified and used for specific tasks.
- An explicit framework for establishing and conducting surveillance will ensure that all key success factors are identified and dealt with in a timely manner.
- A framework that provides a facility to improve the data quality by reporting deficiencies in data back to the owner of the data source.
- A facility to improve the processing accuracy without having to improve the quality of the data input.

1.3 Meeting CRTI Priorities and Leveraging Previous Projects

Developing generic surveillance technology and capabilities meets the priorities of a broad range of stakeholders including the CRTI. These include: healthcare, public safety and security at the municipal, provincial and national level, Defence Research and Development Canada (DRDC) and the Center for Security Sciences (CSS).

The Data Fusion Project accomplishes this by developing frameworks that encompass 1) the software used to conduct surveillance and 2) the processes employed to put a surveillance solution in place.

1.3.1 Software framework

A comprehensive software framework for surveillance must include state-of-the-art tools for data management, data classification, usual statistical analysis, geospatial analysis and conducting specialized analyses unique to surveillance situations.

1.3.2 Process framework

Specific issues that need to be dealt with in a process framework include methods for identifying the best data with which to conduct surveillance, engagement of stakeholders and subject matter experts, making sure the priorities of stakeholders are aligned and that the surveillance solution put in place meets all of their needs, developing a clear definition of the problem to be addressed, identification of appropriate data sources, analyzing data sources and identifying the best methods for classifying and categorizing the data so that it is amenable to statistical analysis, and dealing with issues of data access, in particular privacy and possible competition among stakeholders that may make data access difficult.

A process framework for surveillance is of critical importance, because most projects that fail do so for non-technological reasons. This usually results from a failure to recognize the importance of specific issues, and the difficulties that may arise in dealing with them. These problems occur repeatedly, and a generic framework to deal with such difficulties will ensure that they are not overlooked, and provide the tools needed to deal with them explicitly.

A generalized framework for surveillance is feasible because the data inputs and desired outputs for surveillance are very similar across solutions, even though the specific data is different. A generalized solution can take advantage of the fact that the graphs, statistical analyses and geospatial displays needed to conduct surveillance are very similar regardless of the problem. A significant additional advantage of a generalized solution shared across different areas is that a specialized analysis used in one area may prove directly applicable to another, and provide the latter with a solution that is unique and innovative to their field.

1.4 Uniqueness of this project

The Data Fusion Project was conducted by a large collaborative group assembled over the course of several previous CRTI projects, and leverages the experience and knowledge gained in conducting them. These include ECADS, CNPHI, CEWS, ASSET, MedPost and the ASSET ILI Watch.

The collaborative group includes technological experts in the areas of Data Management (Amita), data classification (NRC-IIT), statistical analysis (Stata Corp.), geospatial visualization (DM solutions) and specialized statistical analyses (DRDC Val Cartier, Carnegie Mellon University).

It also includes subject matter expertise in healthcare and healthcare delivery (University of Ottawa heart Institute), public health at the municipal (QPHI, Ottawa Public Health), provincial (Public Health Ontario) and national (Health Canada) levels. It also includes subject matter experts in infectious disease (departments of Infectious Disease in The Ottawa Hospital and Kingston General Hospital), and ambulance care delivery (City of Ottawa).

The breadth of expertise contained in this group has made it possible:

- To develop unique “state-of-the-art” solutions for technical and non-technical problems in establishing surveillance.
- For cross-fertilization among different groups of stakeholders involved in the project, allowing for the best tools to be applied to each component of developing a generic surveillance solution.

- Make the process of converting raw data to surveillance output very explicit, and has resulted in a unique capability to apply this knowledge to new situations.

2 Purpose

The current section presents the Data Fusion Project hypothesis, and set out the scope boundaries within which the project was executed.

The purpose of the Data Fusion project was to substantiate the following two hypotheses. First, a generalized framework for surveillance can be developed and proven useful. It should also be reusable across multiple problem areas. Second the capability provided by the process and software frameworks developed within this project are broadly applicable. This will encourage stakeholders to pursue surveillance solutions in domains where they would be beneficial, but in which surveillance is not traditionally done.

To address these hypotheses, the Data Fusion Project was executed as a Research and Development project. Its goal was to bring surveillance technology from a level 3 Technology Readiness to a level 5. Our objective was to ready the technology sufficiently to prove its concept, and pave the path for further development toward deployment. Reaching full deployment readiness was considered beyond the scope of the current project.

Development of the process and software frameworks was centered on two distinct and complementary scenarios. These are referred to in this report as prototype applications. This approach serves to validate the applicability, coverage and benefits of the frameworks as well as their generalizability.

The first of the prototype applications concentrated on hospital-acquired infections. This included a range of clearly defined disease syndromes, as well as micro-scale geospatial analysis. The latter included room-to-room discrimination within hospital wards.

The second prototype application centered on detecting harm related to illicit drug use. There were greater levels of uncertainty built into syndrome definitions, and the geospatial analysis done for this application was on a macro-scale (city-a wide grid).

Given these requirements, data collection and analysis was restricted to retrospective data, as opposed to using live data streams. This allowed for collection and analysis of data over a sufficiently long period of time without the requirement that the project itself run for an equivalent period of time.

An underlying premise of this project is that effective surveillance requires a total solution that includes both a computing infrastructure and its supporting organizational environment. It follows from this that in overall of surveillance framework would include both a software framework and a process framework inter-operating in synergy.

Within the software framework, solutions were developed as much as possible using Commercial -Off-The-Shelf (COTS) components. This was done to leverage the development efforts of software vendors and at the same time avoid reinventing the wheel within the project.

The process framework developed for the Data Fusion project was considered a very important component. We recognize that many projects underestimate the importance of the organizational environment and the impact of non-technical factors, and underachieved as a result. In the development of this framework, subject matter experts were heavily involved in clarifying their usual processes while conducting surveillance and elucidating their requirements and priorities. In this way the process framework supports the development of the software framework by ensuring that the design of the latter remains a user-centered, and reducing the chances that unnecessary "nice to have" features creep into the software framework design.

A defining factor for successful surveillance is data access. A critical factor in achieving this is the adoption of measures to assure compliance with relevant legislation and guidelines to assure the privacy and confidentiality of individuals whose data is being used. Maintaining confidentiality and privacy of personal health information is a core value within healthcare. Similar measures are also increasingly being adopted outside of healthcare. For this reason, the Data Fusion Project has maintained itself on the leading edge of technologies aimed at safeguarding confidential information. We have accomplished this by imposing the strictest standards to all components of the project and its underlying frameworks. The standards applied within this project therefore go well beyond minimum legal obligations. In addition to removing direct identifiers, state-of-the-art techniques have been applied to excess and minimize the risk of re-identification by indirect identifiers and by combining variables within and between data sets.

The goals of this project therefore encompassed a wide scope and numerous high-level considerations. This was made possible by the involvement of a multidisciplinary team who is able to provide expertise in a broad range of technology and process domains. This team included the following:

- Projet Management (NRC-IIT Interactive Information Group),
- Health Care and Scientific Direction (University of Ottawa Heart Institute),
- Data Management and Data Security (AMITA Corporation),
- Data Classification (NRC-IIT Interactive Information Group),
- Data fusion techniques, specialized algorithms, time Series Analysis (DRDC Val Cartier),
- Statistics and Data Visualization (STATACorp),
- Geospatial analysis (DM Solutions),
- Human Computer Interaction (Carleton University and NRC-IIT Interactive Information Group),
- Health Care Informatics (Silvacorp, Queens University, hospital IT/IS staff).
- Infectious Disease (TOH),
- Drug Surveillance (Health Canada),
- Public Health (Ottawa Public Health, Public Health Ontario, Toronto Public Health),
- Data Custody for Ambulance data (Ottawa Paramedic Services),
- Hospital data (TOH data warehouse, KGH Information Technology)
- Privacy and confidentiality (Research Ethics Boards and Privacy experts at TOH and KGH, CHEO Research Institute).

The process and software frameworks developed for this project were intended from the outset to be broadly applicable, configurable and extensible. Our goal was to provide stakeholders with a toolkit that would allow them to pursue surveillance solutions in new domains, and to enhance current surveillance capabilities via cross-fertilization with surveillance experts in other subject matter areas.

3 Methodology

3.1 Project scope

3.1.1 Proof of concept using retrospective data

The goal of the Data Fusion project was to make it possible for responders to adapt existing surveillance technology to new situations. The project's intention was to develop software and adaptive process frameworks that will give responders and decision-makers easy access to state-of-the-art data fusion (DF) technology, and make it possible for them to design and deploy domain-specific DF-surveillance solutions. This was achieved by a developing a proof-of-concept software framework to implement DF-surveillance applications and by testing it with retrospective data.

3.1.2 Two problem areas

Two prototype Data Fusion surveillance applications aimed at important problems were selected for development. Application 1 was built to detect serious in-hospital disease outbreaks. Application 2 was built to conduct surveillance of events related to substance abuse.

3.1.3 Planned output

The output of the project was expected to be specific solutions worthy of future development and deployment. This includes generalized frameworks (software and process) to address new problems that would benefit from ongoing surveillance.

3.2 Selection for problems within area

3.2.1 Problem areas

As per the project charter, the following two problem areas[1] were selected to validate the Data Fusion surveillance applications:

- Serious in-hospital disease outbreaks
- Events related to substance abuse

These areas were selected for the following reasons:

- They are important problem areas in health care and public health.

[1]) A third problem area (monitoring high risk public events) was proposed but was omitted after the selection committee reduced the budget by one third.

- They are areas that are relevant to the goals of the project and priorities of the Centre for Security Science and project partners.
- Relevant data exists for these areas and is accessible.
- The two areas are sufficiently distinct to improve the generalizability of the results.

3.2.1.1 Serious in-hospital disease outbreaks

Infection Control (IC) is concerned with tracking and prevention of infections resulting from treatment in a hospital or other healthcare setting. Current IC practice requires manual review of data from several sources, collating of the data and assimilation of the information in order to provide situational awareness and produce reports. This involves repetitive, manual work that is highly resource dependant, and produces fragmented results.

3.2.1.2 Events related to substance abuse

Current national surveillance on drug abuse is fragmented and does not capture large amounts of information that would benefit the end-users. The Office of Drugs and Alcohol Research and Surveillance, Controlled Substances and Tobacco Directorate at Health Canada holds the mandate to collect information on illicit drug use on a national level. The information is sought to assist:

- Health care personnel
- National anti-drug strategies
- Legislation such as Controlled Drugs and Substances Act
- Public health interventions and health care planning
- Reporting to:
  - International Narcotic Control Board
  - United Nations on Drugs and Crime
  - World Health Organization

Current surveillance involves review of data from several sources:

- general and targeted population surveys
- epidemiological monitoring of high risk groups such as
  - street youth
  - injection drug users
- information from drug seizures
- treatment data
- pilot emergency room data

The work involved in collating this data and assimilating information is intensive and leaves many information gaps. There remains a need for a comprehensive national picture, information on emerging problems and an early warning system to detect new harmful trends.

3.2.2 Selection of specific case definitions

Potential case definitions were evaluated according to the following criteria:

- Generalizability of
  - Results, and
  - Methods;
- Results promise to be important
- The likelihood that the necessary data can actually be obtained
- Relevance to area
- Case mix of syndromes defined by a single observation data, and syndromes requiring a composite of observations.

3.2.2.1 Case definitions for serious in-hospital disease outbreaks

Within the Infection Control domain, we prioritized and selected specific health events that are important, considerable, reportable to, and publicly posted by the Ministry of Health and Long-Term Care in Ontario. These health events along with the associated syndrome definitions used for prototype application #1 are summarized in the following table:

Table 1: Prototype Application #1 Syndrome Definitions

Health Event	**Abbreviation**	**Data Fusion Syndrome Definition**		
		Syndrome type	KGH data	TOH DW data
Clostridium difficile	C-diff	Singular	Based on infection control precautions	Laboratory confirmed c-diff toxin text
Methicillan-resistant Staphylococcus Aureus colonization	MRSA C	Singular	Based on infection control precautions	Laboratory confirmed screening swabs
Methicillan-resistant Staphylococcus Aureus infection	MRSA I	Singular	Laboratory confirmed blood, wound or lower respiratory culture	Laboratory confirmed blood, wound or lower respiratory culture
Ventilator Associated Pneumonia	VAP	composite	Not available	Based on components: chest x-ray findings, elevated WBC, antibiotic use,

Health Event	Abbreviation	Data Fusion Syndrome Definition		
				laboratory confirmed lower respiratory cultures
Central Line Infections	CLI	Composite	Based on components: elevated WBC, laboratory confirmed blood culture, antibiotic use	Based on components: elevated WBC, laboratory confirmed blood culture, chest x-ray findings compatible with central line, antibiotic use

3.2.2.2 Case definitions for events related to substance abuse

In consultation with Health Canada, we prioritized and selected specific health events that they consider as being important and that could benefit from enhanced surveillance and an early warning system.

The first event is an example of unexpected harm associated with drug use, namely neutropenia in users of cocaine tainted with levamisole. Levamisole is a veterinarian de-worming agent that causes neutropenia (low neutrophil count, part of the total white blood cell count). It is used as a cutting agent, presumably because it is cheap and enhances or extends cocaine's euphoric effects.

There have been instances in Canada and in the U.S. of patients with life-threatening cases of neutropenia attributed to levamisole-adulterated cocaine. Neutropenia diminishes the immune system's ability to prevent or control infections.

The second health event prioritized by Health Canada is drug overdose due to increased potency, new drugs/combinations/ingredients or usage patterns.
These health events along with the associated syndrome definitions used for prototype application #2 are summarized in the following table:

Table 2: Prototype Application #2 Syndrome Definitions

Health Event	**Data Source**		
	Emergency Room		Ambulance Calls
	The Ottawa Hospital Data Warehouse	Kingston General Hospital	Ottawa Paramedic Service
	Syndrome Definitions		

Health Event	**Data Source**		
	Emergency Room		Ambulance Calls
	The Ottawa Hospital Data Warehouse	Kingston General Hospital	Ottawa Paramedic Service
	Syndrome Definitions		
Neutropenia	Laboratory confirmed low neutrophil count, chief complaint and final diagnosis	Laboratory confirmed low neutrophil count, chief complaint and final diagnosis	
Cocaine use	Based on positive urine toxicology screens for cocaine or cocaine metabolites, chief complaint and final diagnosis	No toxicology available	
Positive microbiology sample culture	Based on microbiology results	Based on microbiology results	
Abnormal Laboratory Result (for unexpected associations)	Based on laboratory results in hematology, chemistry, microbiology	Based on laboratory results in hematology, chemistry, microbiology	
Alcohol use Drug use Substance abuse Poisoning Intoxication Overdose Suicide Attempt Seizure Vomiting Hallucinations	Based on chief complaint and final diagnosis, and toxicology results	Based on chief complaint and final diagnosis	Based on key word search in text fields: Dispatch reason, Paramedic Impression, Treatment and History of Present Illness
Treatment with Narcan (narcotic antidote)	Based on Pharmacy records	Based on Pharmacy records	Based on Treatment field

Health Event	Data Source		
	Emergency Room		Ambulance Calls
	The Ottawa Hospital Data Warehouse	Kingston General Hospital	Ottawa Paramedic Service
	Syndrome Definitions		
Acuity of health event	Based on Canadian Triage Acuity Scale	Based on Canadian Triage Acuity Scale	Based on vital signs

Simple or singular syndromes contain one component and require classification of one data element such as toxicology screens (cocaine metabolite syndrome). Complex or composite syndromes contain several components and require classification of several elements such as positive cocaine toxicology screen *and* low neutrophil count (cocaine plus neutropenia syndrome).

3.3 Development plan

3.3.1 Project Meetings/Workshops

Four project meetings/workshops were held during the project. These meetings brought together subject matter experts, technical experts and project stakeholders. Output from the meetings included the following:

- Development of frameworks
- Discussion of results
- Plotting of new projects
- Expanding partnerships
- new stakeholders and expertise
- information dissemination
- Decision making to resolve project issues
- Identification of risks
- Discussion and development of risk management strategies
- Presentations of interim results
- Planning of publications beyond close date

3.3.2 Iterative Process Solution Development for each problem area

The Data Fusion framework was developed in several stages:

- Stage 0
  - The concept solution and initial framework were developed in collaboration with subject matter experts (SMEs).
  - The concept solution and framework were applied to develop a solution for Prototype Application #1 (*Serious in-hospital disease outbreaks*).
- Stage 1
  - The concept solution and framework were revised during the development of Prototype Application #1 based on lessons learned.
  - The updated concept solution and modified framework were applied to develop a solution for Prototype Application #2 (*Events related to substance abuse*).
- Stage 2
  - The concept solution and framework were revised during the development of Prototype Application #2.
- Stage 3
  - The results of Prototype Application #1 and Prototype Application #2 were validated with SMEs and technical experts.
  - Lessons learned were applied.
  - The framework was enhanced based on the lessons learned.
  - The framework was generalized to be applicable to newer problem areas and applications

3.3.3 Generalized process

- Consult with SME[2]s, develop concept solution
- Use concept to develop solution for Prototype Application 1.
- Apply lessons learned to update concept solution
- Apply updated concept solution to Prototype Application 2.
- Review lessons learned during Prototype Application 2
- Update concept solution
  - Apply lessons learned to finalize frameworks.
  - This final version would be applied to the next project
  - this is a major output of project
- Evaluate framework
  - final output: framework to apply to new problems

[2] Subject Matter Experts

4 Results

The Data Fusion project achieved all of its goals, including the development of a generalized surveillance capability in the form of process and software frameworks. This was accomplished in the context of conducting retrospective analyses in two domains. These analyses provided useful research into their specific subject matter, and also served as proofs of concept for the utility of a generalized surveillance framework. The technology developed during this Research & Development project is now ready to advance to Technology Demonstrations that will show its operational capability.

The conduct of this study involved accessing and analyzing several large data sets. The data for prototype 1 was derived from multiple sources within the inpatient hospital record, including ADT (Admissions, Discharges, Transfers); and radiology, pharmacy, laboratory (microbiology, chemistry and hematology) results. The data set for project two consisted of data derived from three years of ambulance call reports and Emergency Room visits.

The data consisted of numeric, categorical, free text, time and geography variables in multiple formats. In total, over 14 million records were accessed and acquired that retrospectively encompassed over 1000 days.

The software and process frameworks developed by this project deal with the process of identifying, acquiring, managing and analyzing data of this scope, to produce meaningful analyses and useful outputs. Together, these frameworks provide a usable surveillance capability.

4.1 Software framework

Our software framework for surveillance uses state-of-the-art tools to deal with data collection, data organization, analyses and output generation. It consists of an HL-7 compliant Enterprise Database Management System (EDMS), an open source enterprise service bus, data classification tools including natural language processing provided by NRC-IIT, sophisticated tools for traditional data analysis provided by Stata Corp., tools for geospatial analysis provided by DM solutions, and specialized analytic tools provided by DRDC Val Cartier.

4.2 Retrospective analyses

The Data Fusion Process and Software Frameworks were developed while conducting retrospective analyses that addressed important problems using real data. The results provide a convincing proof of the project concept and insight into how better surveillance can be useful in addressing real-life situations.

4.3 Process framework

We developed an explicit framework for the process of implementing a surveillance solution. This includes engaging stakeholders, defining the problem of interest and its solution, dealing with issues of data acquisition and data sharing including privacy, defining specific syndromes to be monitored, and specifying how data is to be analyzed and interpreted. The process framework also deals with issues of information communication and dissemination in the form of useful surveillance products that can be utilized directly by responders.

The process framework developed for the Data Fusion project follows and in some cases expands on established guidelines and methods for project management.

4.3.1 Management and coordination

An essential component for success that was identified by this project is the need for a core team dedicated to developing and improving the process of surveillance. This team need not be focused in any one subject matter area, but needs access to a network with expertise in many such areas.

The key personnel and expertise required by this core team are as follows:

1. A project champion
2. Non-technical expertise skilled in organizational and process factors
3. Technical expertise

If this core team can be established and maintained, it will be able to assemble a wider team of surveillance experts as long as there is a problem to focus on and a commitment and resources to solve it. For experts in the area of surveillance, the opportunity to address an important problem using their skills and expertise is in itself a powerful motivator

Perhaps the most important lesson of this project is that if you build a good team and give them an opportunity and resources to focus on an important problem, they will predictably provide a result that is better than you initially expected.

4.3.2 Key steps of process framework

The key steps of the process framework developed for this project are as follows:

1. Stakeholder identification
2. Initial problem definition
3. Problem definition-stakeholder engagement loop
4. Data acquisition
5. Development of a data management plan

6. Development of a data analysis plan.

4.3.3 Stakeholder Identification

Stakeholder identification begins with a core group that is dedicated to solving the problem at hand. For this project the core group consisted of the project partners, and consisted of the following.

- The NRC Institute for Information Technology,
- The University of Ottawa Heart Institute,
- AMITA Corporation,
- Defence Research and Development Canada-Val Cartier
- Health Canada Drugs Directorate Surveillance Division.

The initial task of this core group was to identify and engage other stakeholders relevant to the problem being addressed. For the Data Fusion project, these consisted of the following.

- Local and Provincial Public Health in Ottawa, Kingston, Toronto and Grey Bruce (prototypes 1 and 2)
- Infectious Disease specialists in Kingston and Ottawa (prototype 1)
- The Ottawa Paramedic Service (prototype 2),
- Emergency Room health care providers (prototype 2)

The next task was to identify and engage technical development experts. For this project these included:

- Stata Corp.
- DM Solutions
- TOH Information Technology
- KGH Information Technology
- TOH Data Warehouse
- City of Ottawa IT
- Silvacorp.
- Carnegie Mellon University
- CHEO Research Institute
- Queens Public Health Informatics

A major goal for this project was to develop it into a win-win proposition for each stakeholder by providing new information and opportunities for collaboration.

4.3.4 Problem definition

Once stakeholders were identified and engaged, problem definition was accomplished by an iterative process. This involved the following steps:

- Defining the surveillance needs,
- Developing an initial plan to meet these needs,
- Reviewing the plan with stakeholders to see how well it met the defined needs of the project,
- Re defining the plan,
- This process was repeated until there was convergence between the plan and the projected needs. At this point the plan was finalized.

Throughout this process the team consulted with all stakeholders to make sure that their goals remained aligned with those of the project, and that the scope of the project had been appropriately managed.

4.3.5 Problem Definition/Stakeholder Engagement Loop

This proved to be an important component in planning each project. Stakeholders had to be engaged while being respectful of their time, and providing them with the expectation of reasonable tangible benefits of participating in the project. Their expectations had to be managed, and their expertise and other commitments acknowledged and respected. Much of this was accomplished at project meetings. These had to be well prepared, include relevant educational content, and have skilled facilitation.

4.3.6 Data Acquisition Plan

A data acquisition plan was developed for each project that proceeded in the following steps.

- Potential data sources were identified
- Data elements in each source were analyzed with regard to information content, data quality and the need for processing and or categorization.
- Data owners were identified and engaged as stakeholders.
- Applicable legislation and regulations, including research ethics when applicable were explicitly addressed.

- Develop Data Sharing Agreements between data owners, researchers, and other stakeholders were developed and finalized.
- Data de-identification requirements and data security were explicitly dealt with.
- Technical and non-technical barriers to data access including cost were identified and dealt with.
- A technical data acquisition plan was developed and validated. Success in all the previous steps culminated in the implementation of this plan.
- At all steps, critical success elements were more often non-technical rather than technical issues.

To manage the risk of not obtaining relevant data with which to work, we purposely sought data that would substantially exceed the scope of the originally proposed project. Not all of it was obtained, but we still exceeded the original project goals and scope with regard to data access by a wide margin. The relevant data sources identified for this project were as follows.

- 3-year retrospective data sets obtained from Kingston General Hospital,
- 3-year retrospective data set from The University of Ottawa Heart provided by the Ottawa Hospital Data Warehouse.
- 3-year retrospective data set of consecutive Ambulance Call Reports provided by Ottawa Paramedic Service
- 3-year retrospective data sets consisting of ER visit clinical data and associated hematology, biochemistry and toxicology laboratory results from Kingston General Hospital and The Ottawa Hospital.
- Poison control data was sought for prototype 2, but was not acquired in the required time frame.
- Coroner's data from Kingston was sought for prototype 2, but was not acquired in the required time frame.

Once was clear what data was to be acquired and used, an explicit data management plan was developed. This included plans for its collection, storage, security, organization and distribution to project team members. We also assessed the quality of the data and developed appropriate tools to identify and deal with duplicate or incomplete data.

One of the lessons learned in the current project is that the wrong technological tools are frequently used to manage epidemiological data. An explicit process for accomplishing this, using the right tools, can realize substantial gains in efficiency and timeliness.

4.3.7 Data Analysis Plan

A data analysis plan was developed for each prototype that identified and dealt with issues that would impact the results of surveillance and their interpretation. These included syndrome definition and the development of simple and complex syndromes. For the Data Fusion project we developed an operational definition of "syndrome" as follows:

A syndrome is something derived from an ongoing data stream that will be counted and used for surveillance.

A simple syndrome was defined as a syndrome that can be derived from a single element in the data stream. A complex syndrome was defined as a syndrome whose definition required the combination of two or more data elements.

The Data Analysis Plan specified exactly what would be counted for each syndrome in each project (e.g. exact syndrome definition, incidence versus prevalence). These decisions were usually context-specific. For all comparisons it was also essential to explicitly define the appropriate denominator for each syndrome. The Data Analysis Plan specified which initial analyses and data visualizations would be used, along with expected results and expectations of what an abnormal result would look like.

4.3.8 Data Dissemination and Utilization Plan

An important part of the process framework is to explicitly identify the data users and their needs. Intended users should be included as stakeholders. As new potential data users are identified they will need to be engaged, and it may be necessary to broaden the stakeholder group to include users that were not planned at the beginning of the project.

4.3.9 Cross Fertilization

A major advantage of the approach taken by this project is that it brings together surveillance experts in different subject matter areas and allows them to share their experiences and knowledge. Under these circumstances, it will be possible to identify known solutions in one area that are directly applicable but unknown in another area. This will lead to lateral transfers of expertise, which are an important source of innovation.

4.4 Software Framework

The premise of the data fusion project is that a framework, that identifies the most suitable available technology for each task, will have significant advantages over silo solutions developed in isolation. The software framework developed by this project was designed to be: 1) generalizable, 2) scalable, 3) easily deployable, 4) leveraging existing technology, 5) modular, 6) well integrated, 7) extensible. It is critically important that the framework not be absolutely dependent on any one component, such that other available technologies could be used if necessary.

The following components were identified as necessary to achieve these characteristics: 1) Enterprise Service Bus (data bus), 2) data classification, 3) general statistical analysis, 4) specialized statistical analysis, 5) mapping and geospatial visualization.

4.4.1 Enterprise Service Bus

Incorporating an enterprise service bus (data bus) addresses the reality of disparate data streams presenting data in different formats, needing to be collated together so they can be analyzed. Most currently available statistical programs require flat files and are slow and cumbersome when dealing with complex multidimensional data. Ultimately, to be usable in an "off-the-shelf" statistical package such data must be converted into a series of well-designed flat files.

This problem was addressed by adopting an HL-7 compatible open source data bus (MIRTH). This solution was chosen because of its cost effectiveness, freedom from vendor lock-in, and low number of deployment issues given its open source nature. Development on the MIRTH data bus was undertaken by AMITA Corporation. Combining the power of a data bus with the data management tools in the off-the-shelf statistical package chosen for this project (Stata) resulted in a robust and scalable data management solution superior to any previously available option.

4.4.2 Data Classification

One of the goals of a generalized framework is to make it possible to use the best available data for surveillance. In general, data is ignored if it is not presented in a readily analyzable format, even if it contains highly relevant information. A particular example of this is data that exists as free text in its native format. The alternative, which is often adopted, is to accept lower quality data that requires no pre-processing.

We overcame this obstacle by building an explicit step in the process framework to identify the best data, and by including the tools in the software framework that allowed us to categorize it into useful information. As part of the process framework, this information was specifically developed into syndromes, broadly defined as information derived from the available data stream that could be counted and used for surveillance and tracking. A simple syndrome is information that can be derived from one data stream (e.g. a chest x-ray finding is present or absent). A complex syndrome is information that is derived from more than 1 data stream (e.g. a positive chest x-ray plus a high white blood cell count).

This approach gives us the ability to combine multiple data streams into syndromes that are useful for surveillance. Within the healthcare domain, we now have the unique capability of being able to conduct surveillance directly from EHR-derived data in its native format. This capability will prove very useful in implementing online, real-time surveillance solutions.

4.4.3 Usual Statistical Analysis

Many surveillance systems include basic tools for graphing and statistical analysis as part of a "dashboard" solution. These tools are often rudimentary, and there is little flexibility for the end-user to modify or use different analytical methods. Rather than devote time and resources to developing such tools, we strategically partnered with Stata Corp. (a major developer of Commercial-Off-The-Shelf (COTS) statistical software) enabling end-users access to a broad range of sophisticated statistical and graphing algorithms. These algorithms have been extensively validated, and have a user base that is familiar with them and available for peer-support. Stata also contains a powerful scripting language perfect for developing complex re-useable routines for and by end-users. End-users are provided the opportunity to use the best tool for a specific task from a very large toolset, to leverage established methods that are proven and validated, and to communicate their results easily and accurately.

4.4.4 Specialized analytical and data fusion techniques

When faced with an unexpected situation that requires surveillance to be established, end-users will often develop “tunnel vision”, and stick with those techniques they have used in the past. Not infrequently, experts conducting surveillance in other areas will have developed techniques that are directly applicable to the current problem but are unknown to those who are attempting to manage or mitigate it. The process developed for this project prioritized the inclusion of surveillance experts from a broad range of subject areas; utilized Data Fusion technology developed at DRDC-Val Cartier; and made it available for analysis of health-care related data. By allowing for cross-fertilization from different areas, this approach gives a much greater breadth to the analytic capability that can be applied to a problem.

4.4.5 Mapping and geospatial visualization

Mapping is often very effective for visualization of surveillance-type data, but is often unavailable to end-users. This visualization is frequently complementary to statistical analysis and epidemiological plots in the understanding of data. Effective mapping is often hampered by data that must be processed into a format that can be mapped and usually requires unique software to accomplish this. Much of the time and resources devoted to mapping are spent on data manipulation, using sub-optimal tools contained in a mapping program.

We engaged the expertise of an industry partner with mapping expertise. We also undertook major data manipulation processes using proficient tools, in this case the enterprise data bus. This approach let us work effectively with mapping experts to generate reusable maps.

This approach made it possible for us to map disease syndromes derived directly from EHR data down to the bed level in both Kingston General Hospital and the University of Ottawa Heart Institute. This level of granularity for data display for these institutions is currently not available from any other source. The technology developed for this project now positions us to accomplish this in real-time in a production system.

4.5 Prototype Application 1

The objectives of the Data Fusion Project were to develop process and software frameworks for surveillance while implementing two prototype applications. Prototype application 1 deals with hospital-associated infections, more specifically C. Difficile infection, MRSA colonization and infection, ventilator acquired pneumonia, and central line infections.

4.5.1 Data Sources

Application 1 was constructed using retrospective data from the University of Ottawa Heart Institute and Kingston General Hospital. The latter provided a three year set that includes native EHR data streams of the following:

- Admissions, discharges, transfers and bed changes
- Infectious disease precautions instituted on admitted patients
- Laboratory
- Microbiology
- Emergency room visit
- X-rays ordered (but not results, these were scanned and therefore unavailable)
- Pharmacy

Data were provided on all patients admitted to Kingston General Hospital over a 3-year period of time encompassing 2009-2011.

The Ottawa Hospital Data Warehouse provided data for the University of Ottawa Heart Institute. It consisted of a complete three-year set of data in native EHR format. Separate data feeds were provided for the following:

- Admissions, discharges, transfers and bed changes
- Microbiology
- Radiology chest x-ray results
- Hematology laboratory results on white blood cell counts
- Pharmacy results regarding the administration of intravenous antibiotics, and antibiotics used to treat C. Difficile.

4.5.2 Data Management

Data were de-identified at source for primary identifiers (e.g. name, hospital number, etc.). They were then transferred to AMITA Corporation in an encrypted format and transferred to the project Enterprise Database Management System developed by AMITA. Data that required classification (e.g. free text test results) were transferred to NRC-IIT in encrypted format. Data classification was accomplished at NRC according to rules developed in conjunction with healthcare experts. Free-text classification was designed to account for negations and other modifiers in the context of the findings. The classified data was then transferred back to AMITA in an encrypted format. The EDMS was used to organize and combine data into a usable format for transfer to STATA Corp. for statistical analysis, to DM solutions for geospatial visualization, and to DRDC Val Cartier for specialized analyses.

Prior to transfer from AMITA, data elements were specifically analyzed to avoid the risk of indirect identification of individual patients (e.g. by combining two variables which were not individually identifying but which might be identifying in combination). All data transfers had to be approved by the project data custodian (Dr. Richard Davies).

Using the data derived directly from EHR records, after categorization by NRC, it was possible to define meaningful simple and complex disease syndromes. The following example shows a data stream of chest x-ray, laboratory and pharmacy data on an individual patient, and illustrates how these data are interpreted in order to derive cyclic syndromes related to the diagnosis of Ventilator Associated Pneumonia (VAP). VAP is one of the quality indicators used by the Ontario Ministry of Health to assess quality of care in hospitals.

Figure 1: Chest X-Ray (Filtered by test results within encounters)

tdatetime	nursingstation	bed	intubated	centralline	pneumonia	pleuraleffusion	Edema
-03 20:30:00	HCSA	CSIA03	INTUBATED	CL	NULL	NULL	EDEMA
-03 22:50:00	HCSA	CSIA03	NULL	NULL	NULL	NULL	NULL
-04 01:32:00	HCSA	CSIA03	INTUBATED	CL	INFILTRATE	PLEURAL_EFFUSION	NULL
-04 07:00:00	HCSA	CSIA03	NULL	NULL	INFILTRATE	PLEURAL_EFFUSION	NO_EDEMA
-04 10:00:00	HCSA	CSIA03	NULL	NULL	INFILTRATE	NULL	NULL
-06 10:45:00	HCSA	CSIA03	INTUBATED	NULL	INFILTRATE	NULL	NULL
-08 09:45:00	HCSB	CSIB10	NULL	CL	NULL	NULL	NULL
-10 16:00:00	HCSB	CSIB10	NULL	CL	INFILTRATE	NULL	NULL
-12 11:10:00	HCSC	1332-1	NULL	NULL	NULL	PLEURAL_EFFUSION	NULL
-16 20:30:00	HCSC	1332-1	NULL	CL	INFILTRATE	PLEURAL_EFFUSION	EDEMA
-17 08:55:00	HCSC	1332-1	NULL	CL	INFILTRATE	PLEURAL_EFFUSION	EDEMA
-29 10:15:00	H3	3330-1	NULL	CL	INFILTRATE	PLEURAL_EFFUSION	NULL

Figure 2: WBC (Filtered with only either high for tohguideline or mohltc guideline)

ldatetime	nursingstation	bed	labresult	tohguideline	mohltcguideline
l-03 20:30:00	HCSA	CSIA03	11.8	high	normal
l-05 03:45:00	HCSA	CSIA03	15.5	high	high
l-05 22:00:00	HCSA	CSIA03	14.9	high	high
l-06 03:50:00	HCSA	CSIA03	16.2	high	high
l-06 22:20:00	HCSA	CSIA03	13.0	high	high
l-09 21:30:00	HCSB	CSIB10	12.3	high	high
l-10 05:35:00	HCSB	CSIB10	13.6	high	high
l-11 04:00:00	HCSB	CSIB10	12.4	high	high
l-11 18:30:00	HCSC	1332-1	12.9	high	high
l-12 02:30:00	HCSC	1332-1	12.5	high	high
l-12 06:15:00	HCSC	1332-1	11.8	high	normal
l-13 04:05:00	HCSC	1332-1	13.9	high	high
l-14 05:29:00	HCSC	1332-1	13.9	high	high
l-15 03:20:00	HCSC	1332-1	14.6	high	high
l-16 04:50:00	HCSC	1332-1	14.1	high	high
l-17 04:50:00	HCSC	1332-1	13.2	high	high
l-22 08:25:00	H3	3330-1	11.9	high	normal
l-23 09:38:00	H3	3330-1	13.5	high	high
l-24 08:16:00	H3	3330-1	12.3	high	high
l-25 08:36:00	H3	3330-1	13.8	high	high
l-26 08:14:00	H3	3330-1	14.5	high	high
l-27 08:01:00	H3	3330-1	12.5	high	high
l-28 08:40:00	H3	3330-1	13.1	high	high

Figure 3: Microbiology by encounter (Some patients do have multiple tests classified with different values per encounter, this one is one of the simple ones)

datetime	nursingstation	bed	mrsac	mrsai	cdiff	lrculture	bloodc	bloodi	bcorganism
-03 20:30:00	HCSA	CSIA03	NI	0	0	NULL	NULL	NULL	NULL

Below is what a more complex one would look like (from another unrelated encounter)

-03 00:30:00	H1CA	1325-1	NI	0	0	NULL	NULL	NULL	NULL
-04 21:20:00	H1CA	1325-1	0	0	0	P_LR	NULL	NULL	NULL
-05 14:15:00	H1CA	1325-1	0	NC	0	NULL	N	N	NULL
-05 15:00:00	H1CA	1325-1	0	NC	0	NULL	N	N	NULL
-07 10:35:00	H1CA	1325-1	NI	0	0	NULL	NULL	NULL	NULL

Figure 4: Pharmacy (Filtered on positives for either IV or CDiff Antibiotics)

rtdatetime	orderenddatetime	nursingstation	bed	ivantibiotics	cdiffantibiotics
-18 08:00:00	[redacted]22 08:45:00	HCSC	1332-1	1	0
-18 08:00:00	[redacted]20 08:00:00	HCSC	1332-1	1	0
-18 08:00:00	[redacted]-18 17:00:00	HCSC	1332-1	1	0
-18 08:00:00	[redacted]18 17:00:00	HCSC	1332-1	1	0
-22 08:00:00	[redacted]25 08:00:00	H3	3330-1	1	0
-25 17:24:00	[redacted]-30 11:45:00	H3	3330-1	1	0
-30 08:00:00	[redacted]-28 23:59:00	H3	3330-1	1	0

4.5.3 Preliminary Results of Prototype 1

Using the data management algorithms developed by the team, it was possible to track and chart individual disease syndromes by ward, and to map the incidents of specific syndromes down to the individual bed. This includes syndromes derived from free-text data (e.g. chest x-ray reports). To our knowledge this is a unique capability not previously available.

Some examples of case frequency reports created by the Data Fusion project in the form of maps are provided in the following figures.

Figure 5: Mapped Cases of Clostridium difficile by ward over 1 year

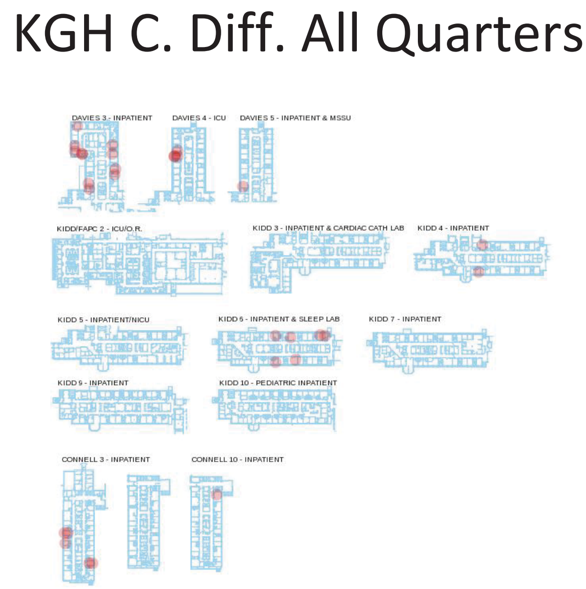


Figure 6: Mapped Cases of Clostridium difficile by ward over 3 months

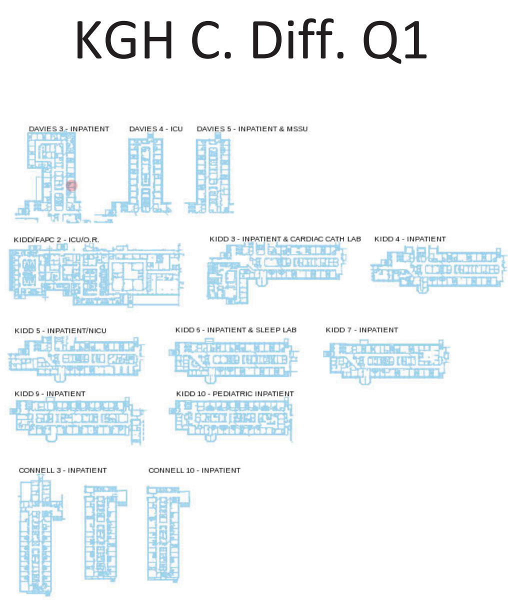


Figure 7: Mapped Cases of Methicillin Resistant Staph Aureus by ward over 1 year

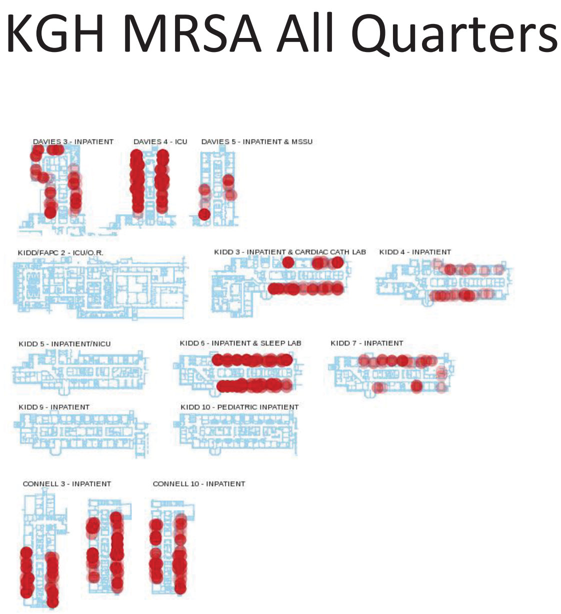


Figure 8: Mapped Cases of Methicillin Resistant Staph Aureus in ICU over 3 months

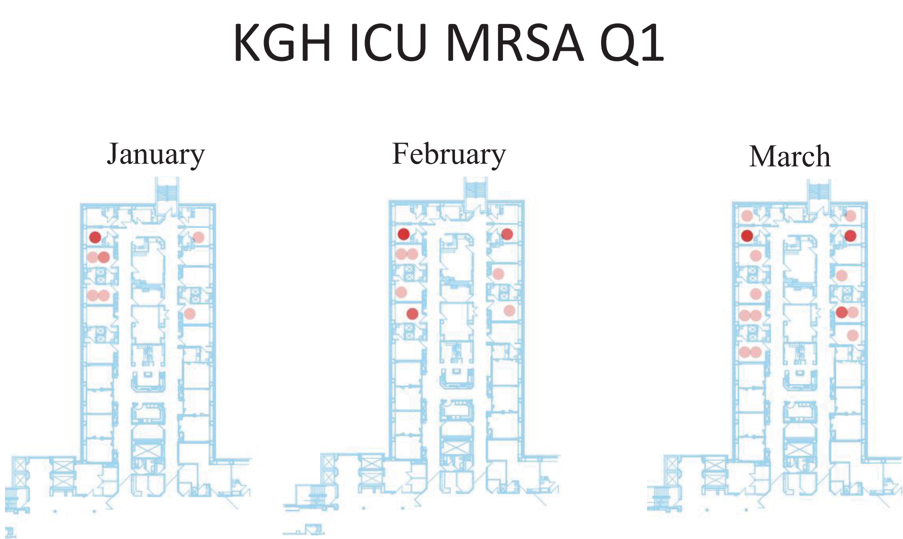


Figure 9: Cases of Chest x-ray Infiltrates over 1 year, all wards

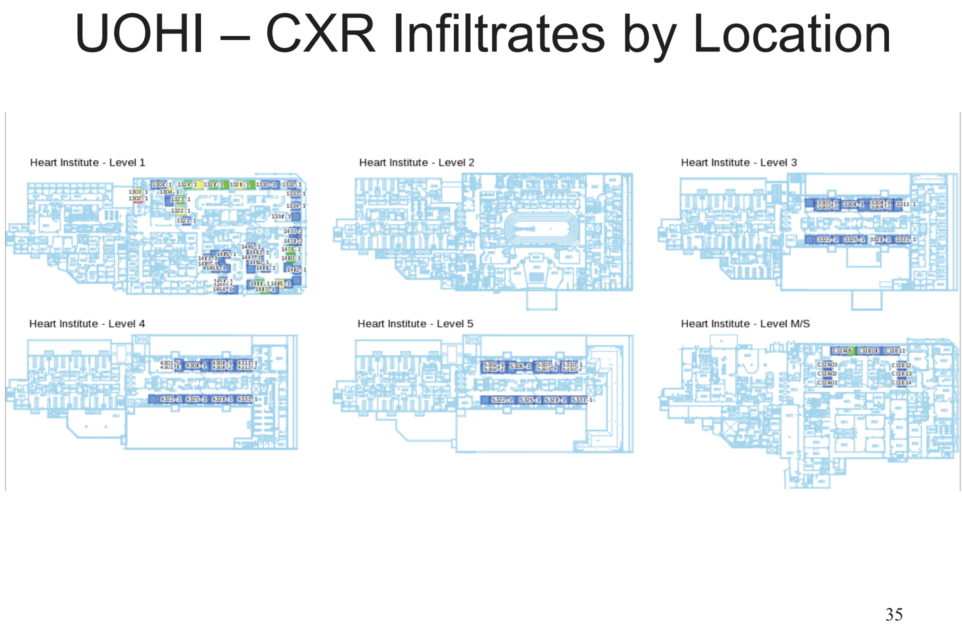


Figure 10: Cases of Chest x-ray Infiltrates over 1 year, by ward

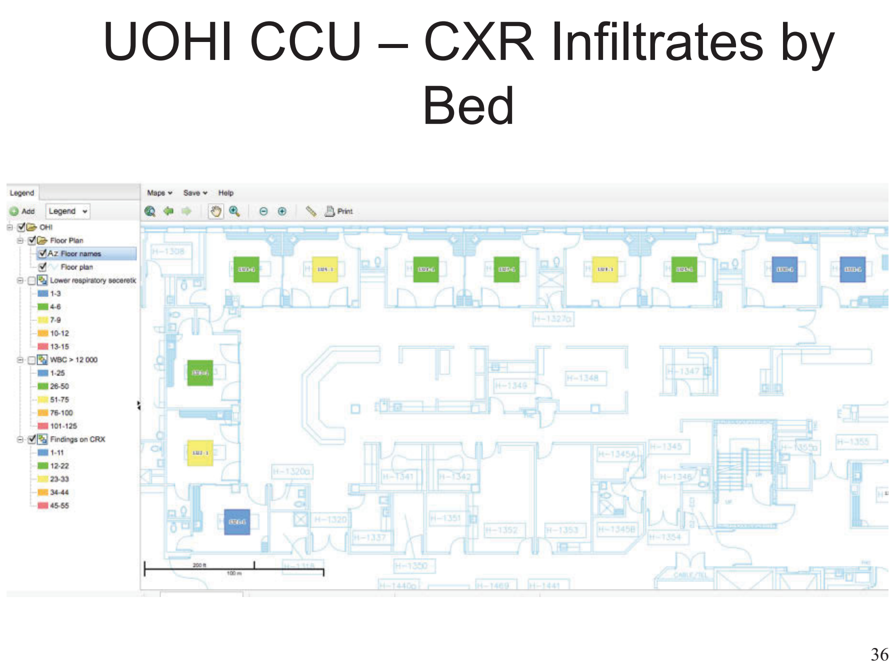
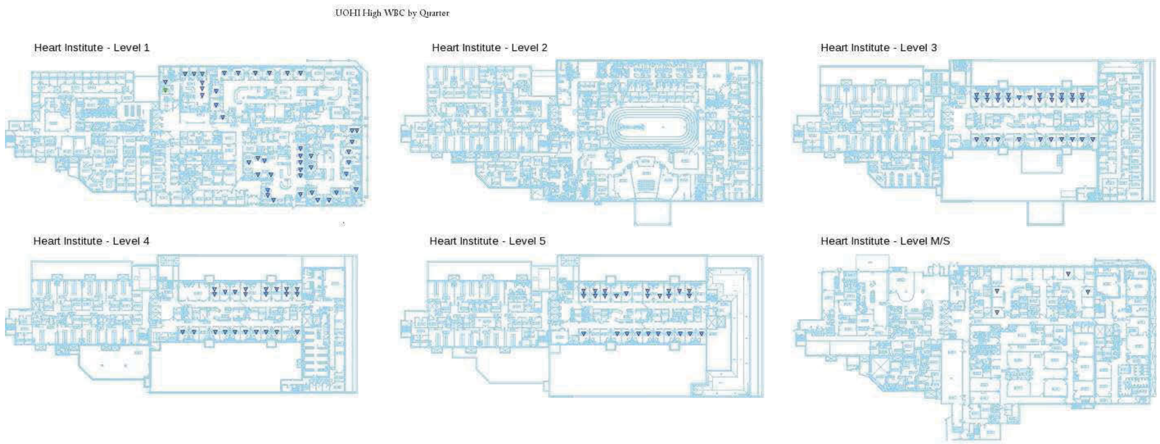
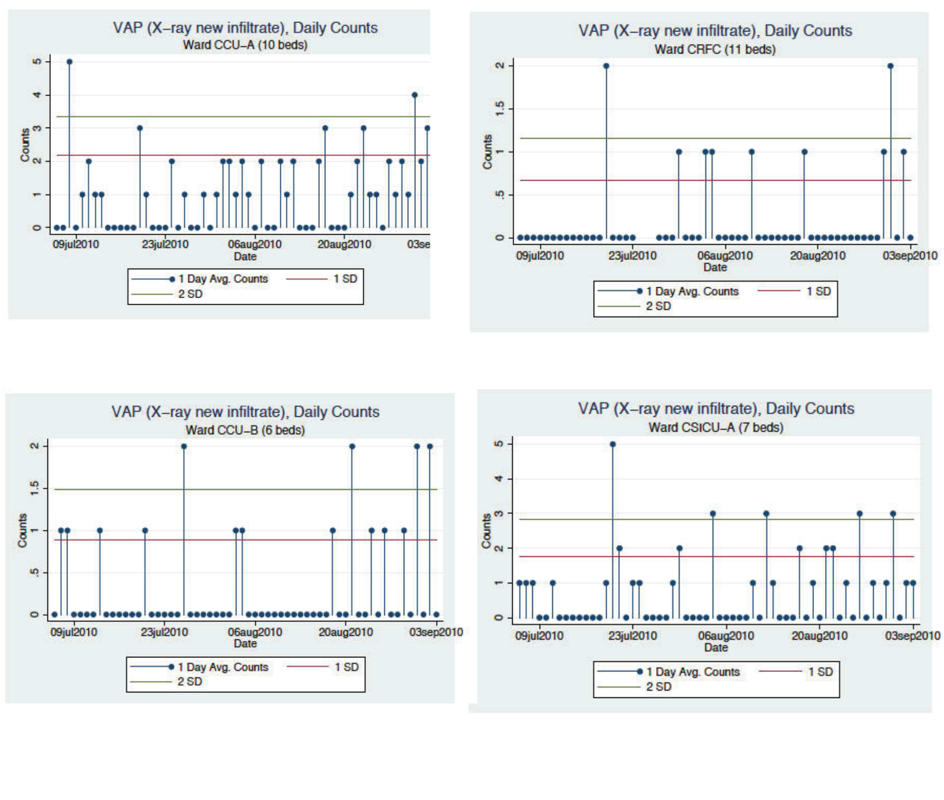


Figure 11: Cases of elevated White Blood Cell Count over 3 months by ward



Some examples of case frequency reports created by the Data Fusion project in the form of epiplots are provided in the following figure.

Figure 12: Cases of Chest x-ray Infiltrates over 1 year



Within prototype 1 we also demonstrated the capability of tracking hospital-acquired infections of interest such as Vancomycin Resistant Enterococcus (VRE), Methicillin Resistant Staphylococcus Aureus (MRSA) and Clostridium Difficile (C. Diff). The following shows example results.

4.5.4 Data Quality

The data provided by the hospitals were generally of very high quality. However, one of the lessons learned is that it is important to be cognizant of the limitations inherent in any data stream used for surveillance. The following examples illustrate the need for high quality data.

The first example relates to the precautions data obtained from Kingston General Hospital. This data identified all instances where precautions were instituted because of infection or colonization with C diff, VRE or MRSA. However, we discovered that the date attached to each precaution was the date of hospital admission, rather than the date of a positive test. In the cases of VRE and MRSA this is not important. A positive test result simply identifies a patient who is colonized. However, in the case of C diff this is crucial, because the test identifies patients who have become acutely ill. Often C diff infection only occurs after several weeks in hospital, so this would result in cases being identified, in some instances, weeks before they occurred. This would not be a problem for tracking long-term accumulations of cases, nor would it be a problem for an on line system that acquired new data on a daily basis (in which the date a disease occurred could be inferred from the date the report was received). It did however pose a problem for retrospectively tracking disease incidence, which was the goal of this proof of concept project.

Similarly, the categorization of free-text data can sometimes result in missing syndromes. This occurred in the example above for chest x-ray reports. It is apparent that in some chest x-rays the radiographer did not explicitly report on findings which were likely there. This would result in individual chest x-rays not identifying the presence of an endotracheal tube or a central line. The presence of these findings would have to be inferred from the results of previous and subsequent tests, and this would have to be taken into account in designing syndromes used to track disease.

4.6 Prototype Application 2

The objective of Prototype 2 was to further develop the process and software frameworks developed in Prototype 1, and adapt them to detecting and tracking harm related to illicit drug use

4.6.1 Data Streams for Prototype 2

Two sources of data were used for prototype 2. The first is a set consisting of information from Ambulance Call Reports (ACRs) for the Ottawa Paramedic Service encompassing the years 2009-2011. It includes the following elements:

- Dispatch Reason
- Paramedic Impression
- History of Present Illness
- Treatments Administered
- Vital Signs

- Time
- Location of pickup (expressed as Universal Transverse Mercator (UTM), which defines geographical location within one Km).

The second set consists of emergency room data for the same three-year period of time obtained from The Ottawa Hospital. It includes the following elements:

- Age and Gender
- Chief complaint
- Final diagnosis
- CTAS Triage code
- Five digit postal code
- Results of usual laboratory tests (CBC including WBC count, electrolytes BUN and creatinine)
- Results from toxicology screening if sent.

One of the challenges faced in conducting this study was the unexpected length of time necessary to access the data. As a result, only preliminary analyses for prototype 2 are available for this report. These are sufficient for proof of concept and to meet the primary goals for this project. Further analyses will be conducted beyond the fiscal year end with the goal of publishing our results.

4.6.2 Hypothesis 1

Prototype 2 was directed at addressing two hypotheses.

The first hypothesis addressed the feasibility of using ER data to detect unexpected associations, such as those observed between cocaine abuse and neutropenia (low neutrophil count, part of the white blood cell count). The latter is known to occur because of the contamination of cocaine with levamisole, which is an antiparasitic agent known to induce neutropenia in some patients.

- The emergency room data needed to test hypothesis 1 are acquired.
- Chief complaint and final diagnosis data have been sent to NRC for categorization.
- Cases have been classified with regard to the possibility of drug abuse. Associated hematology laboratory data [white blood cell counts] have been classified as normal, low (neutropenia) and high.
- Analysis to determine whether or not it is possible to detect an association between drug abuse and neutropenia is underway.

One specific challenge encountered stemmed from privacy concerns, resulting in a lack of an identifier linking cases across ER encounters, and it is possible that patients classified as potential drug users on one visit could present with neutropenia on another visit and we would not detect the association. A live surveillance system capable of linking patients across encounters would therefore be more capable of picking up abnormal associations. This is particularly true if linkage could also be established to inpatient records.

4.6.3 Hypothesis 2

The second hypothesis addresses the question of whether ambulance call data can be used to track the geographic location of cases possibly related to drug overdose. This relates to the well-documented problem of narcotic overdose related to unexpectedly pure drugs or to drugs laced with the anaesthetic Ketamine. The use of either could result in an unexpected overdose. Geographic clustering of such cases might be useful in detecting the presence of contaminated drugs, and could lead to the identification and elimination of the source.

- The Ambulance data to test hypothesis 2 has been provided by the Ottawa Paramedic Service and loaded into the project EDMS.
- Dispatch Reason and Paramedic Impression data (semi-free text from a restricted vocabulary) were then transferred to NRC-IIT where it was categorized as follows:
  - Alcohol (yes or no)
  - Drugs (yes or no)
  - Specific drugs (yes or no)
    - Amphetamines
    - Cocaine
    - Antidepressants
    - Ecstasy
    - Heroin
    - Marijuana
    - LSD
    - Stimulants

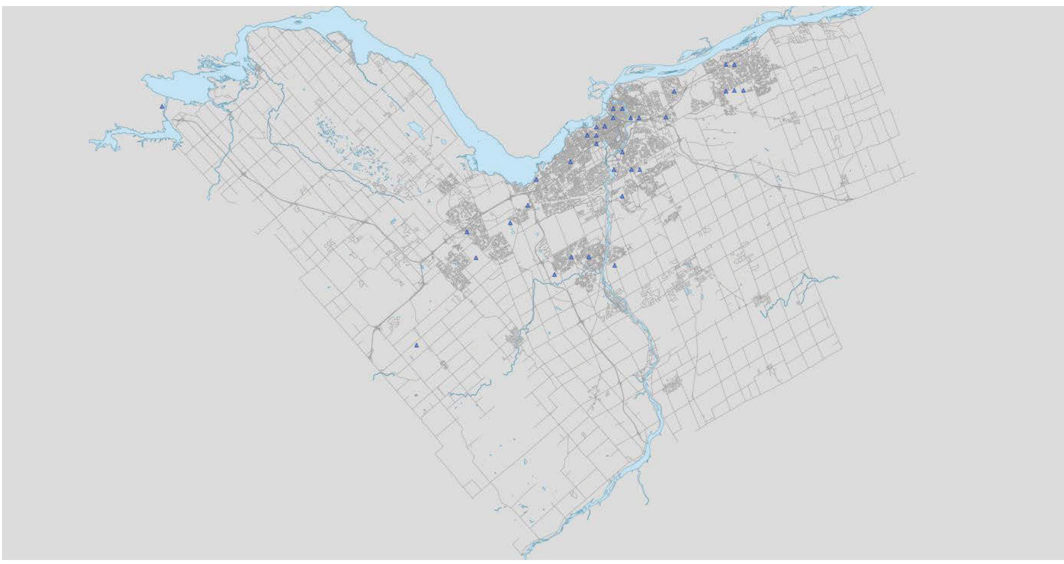
- Cases were categorized on the basis of keyword searches on the following single words:
  - Substance
  - Poison
  - Intoxication
  - Overdose
  - Intent
  - Suicide

- Cases were also categorized according to the occurrence of the following adverse effects:
  - Vomiting
  - Hallucinations

- Cases were also categorized based on the following compound definitions:
  - Drug plus overdose
  - Overdose plus intent
  - Overdose plus adverse event

Using these case definitions, we demonstrated the capability of mapping incidence by time and geography. The latter is accurate to the square kilometer, and would be adequate for the identification of abnormal clusters. Sample maps are shown below:

Figure 13: Map showing incidence of drug overdose based on ambulance call reports in Ottawa in a three month period



This demonstrates the capability of acquiring electronic health record data, deriving disease syndromes of interest, and geographically mapping these syndromes for the purposes of surveillance and tracking.

5 Transition and Exploitation

5.1 Transition to End Users

In the Data Fusion project we have successfully developed a comprehensive framework for a multi-disciplinary scientific team doing applied research in a real life setting. The science has moved from the labs into the real world. Twenty partner organizations were directly involved with more organizations indirectly engaged. The professions involved in the project included: medical, information science, information processing, legal and commercialization.

The Data Fusion framework has proven the concept of detecting complex outbreaks with multiple data streams using retrospective data. The next step is to demonstrate the system in real-time, in an operational environment.

The problem domain explored in this project was the medical domain. However, by design, the framework can be generalized to other domains such as: defence, intelligence, search and rescue, environment (e.g. health impact of air quality), the safety of the food and water supply as well as advertising. This makes possible the establishment of a multi-sector partnership to maintain and develop this technology.

Potential end-users of an operational Data Fusion framework are public health organizations at the municipal, provincial and federal levels, including the 120 Public Health Units and Regional/District Health Authorities in Canada. Exceedingly few currently have access to such a system. The Data Fusion framework will push this technology toward a solid foundation that favours uptake and use, by making a common system easily available to all Canadian users at a low cost. This will also facilitate data sharing and ensure that subsequent technology development benefits users across Canada.

5.2 Follow-On Commercial Development or R&D Recommended

5.2.1 Commercial Development

The Data Fusion project team integrated the business of many government departments, academia and health science institutions. Thus, the four businesses were immersed in the scientific and end-user community.

As a result of the Data Fusion project, the foundation has been prepared for a commercial product.

In parallel with the Data Fusion project, AMITA has invested in the commercialization of the framework that was developed. The goal of this investment is to offer a world-class service product to domestic and international customers, thereby creating a situation where high quality jobs in the knowledge industry can be created.
The findings from the project are currently being used to prepare a competitive proposal for a specific operational domain (Search and Rescue, a demonstration that leverage is already being created in a competitive situation.

NRC will be in an excellent position to collaborate with Canadian industry to advance the Data Fusion framework and embed it into commercial grade integrated systems or system components.

5.2.2 R&D Recommended

The following Research and Development (R&D) is recommended:

1. The Data Fusion framework with its data feeds should be funded to keep the project alive another two years.

   The funding should include:

- Management of the servers with data collected
- Serving of data requests for the scientific members
- Administrative resources for government reporting
- Assistance with complex statistical questions
- Two scientific conferences annually
- Outreach and communication activities
- Keeping the data sources feeding data
- Allowing for two new data sources a year

2. Make linkages with the recently started FUTURE INTELLIGENCE ANALYSIS CAPABILITY (FIAC) led by DRDC Val Cartier.
   - This would include collaborative sessions and presentations at the FIAC conferences.
   - The Data Fusion project team end-goal was similar to the FIAC aims although Data Fusion covered a smaller information domain and was restricted to two problem areas.
   - Lessons learned can be leveraged by the FIAC team with the benefit of shortening the FIAC start up time significantly (to months rather than years).
   - Three Data Fusion partners are currently involved in the FIAC initiative (DRDC Val Cartier, AMITA and NRC).

3. Fund the "monitoring high risk public events" problem area that was cut when the budget for the Data Fusion project was reduced by one third.

5.2.3 Outreach

Funding should be secured to give presentations and provide information pamphlets to the scientific domain users and the public.

5.3 Intellectual Property Disposition

This project followed the principles and approach to intellectual property set forth in the CRTI Guidebook:

At the outset of the project, proprietary software, methods or practices used in this project which do not fall under the open source GPL, and in which project partners have Intellectual Property (IP) will be identified. The team members will finalize how Intellectual Property (IP) will be addressed, including the particular ambitions and desires of the team members in terms of use of the resulting software product. The right to use of the foreground IP for the team members will be determined in light of the ambitions of each of the partners. Should there be difficulty resolving any issues, the Treasury Board Policy on Title to Intellectual Property will be used For Disclosure and Use of Information.

Discussions yielded the following:

- All NRC IP will remain the property of NRC.
- All DRDC IP will remain the property of DRDC.
- All AMITA IP will remain the property of AMITA Corporation.

5.4 Public Information Recommendations

In October 2010, the Expert Panel on Research and Development was charged by the Government of Canada to examine how to strengthen the impact of federal investments in support of a more innovative economy. Innovation Canada: A Call to Action [1], the Expert Panel's final report, published in October 2011, lays out a series of recommendations for government's support to innovation.

Recommendation 6 of this report is as follows:

"Establish a clear federal voice for innovation and engage in dialogue with provinces to improve coordination and impact."

Taking that further, the Panel stated their vision as follows:

"The Government of Canada must assume a leadership role by establishing business innovation as a whole-of-government priority and consequently restructuring the governance of its business innovation agenda, while developing a shared and cooperative approach with provincial and business leaders."

The Data Fusion framework developed by this project is a prime example of Canadian innovation, where federal funding and research capacities worked in close collaboration with provincial and local governments, private sector, and academia. The shared and cooperative approach has produced new technology that fills a gap and has great potential for commercialization.

One of the primary tasks of the Centre for Security Science (CSS) is to enable and foster partnerships across departments and agencies in the Government of Canada, and among federal, provincial, and municipal levels of government, private-sector industry, academia, and responder and operational communities. The role of the Public Security Technical Program (PSTP) in this task is to reach out to potential partners involved in the PSTP mission areas and provide a forum through which the partners can develop into Communities of Practice.

Good communication among project partners, PSTP programme participants and industry professionals in the public domain is required to maintain product momentum with Data Fusion. The media can be valuable and effective by informing the public about the benefits to Canada, especially if this is done in a non-technical and straightforward way.

6 Conclusion

For the problem areas under study, it was shown that syndromes could be defined to allow surveillance. These syndromes require input to be combined from disparate data streams. It is demonstrated that obtaining the individual data elements is relatively straightforward and that the final result from combining these exceeds the sum of the parts. Three key components result from the process framework: (a) the knowledge of which data elements are required, (b) the knowledge of how to combine these data elements into the data composite, and (c) the choice of analysis method to interpret these composite data signals as being part of a normal state or an aberration.

Our project team identified two areas that could benefit from automated surveillance. We developed an adaptive process and software framework for each of them: 1) the detection of serious in-hospital disease outbreaks, and 2) the surveillance of harm related to illicit substance abuse.

6.1 Conclusion Prototype Application 1

In Prototype Application 1, in-hospital infections are detected and tracked per patient (or per bed) over time. This required collection and combination of data streams from multiple sources, including Admission and Discharge, Bed transfers, Drug prescriptions, Lab and Imaging diagnostics, and Contact Precautions. The statistical aberration detection algorithms and the mapping functions both successfully revealed elevated infection rates for several types of infection. It was shown that this solution would augment the agility of the hospital's Infection Control responders. The steps involved to construct the solution were documented. This prototype application is expected to generalize well to scenarios where causes and effects are relatively well understood, but where tighter integration of data, technologies, and processes are paramount to success.

6.2 Conclusion Prototype Application 2

In Prototype Application 2, illicit drug use signals are captured to detect the surfacing of new drugs on the market, contaminated batches of drugs appearing, or other circumstances in which substance abuse lead to adverse medical events. In this scenario, both the incidence and the scope of measurable data elements contain greater uncertainty. The Data Fusion project collected and allowed the combination of data obtained from Ambulance Dispatch, Emergency Room presentations, and Toxicology laboratory reports. It was shown that the Data Fusion approach led to signals being picked up by combining data elements, where originally the environment was too noisy for the signals to be detected. This prototype application is expected to generalize well to scenarios of considerable cause/effect uncertainty where flexibility in data choice and analysis methods is required.

6.3 Summary

The Data Fusion project has successfully developed a robust and multi-functional prototype framework for surveillance. The generalized adaptive framework, which arches across the software and process environments, is usable over a wide range of subject areas addressing multiple problem areas. It will lead to improved access to surveillance tools for environments where some surveillance tools already exist, and will lead to new capacities for surveillance in other environments where surveillance is traditionally not or marginally present.

References.....

[1] *Innovation Canada: A Call to Action*, Cat. No. Iu4-149/2011E-PDF, ISBN 978-1-100-19384-7, 60957

This page intentionally left blank.

Annex A Project Team

Position	Name	Title	Phone Number and email address
Project Role			
Project Champion	Ms. Cheryl Lambert	General Manager, Information and Communications Technologies, National Research Council of Canada	613-991-4650 Cheryl.Lambert@nrc-cnrc.gc.ca
Project Manager	Dr. Peter Wojcik	Project Manager, NRC-ICT	613-993-0186 Peter.Wojcik@nrc-cnrc.gc.ca
Portfolio Manager	Mr. Norman Yanofsky	CRTI Portfolio Manager, Defence R & D Canada	613-944-8161 613-995-0002 Norman.Yanofsky@rddc.gc.ca
Deputy Project Manager, Health Care/Public Health Team Leader	Dr. Richard Davies	Professor of Medicine, University of Ottawa	613-761-4729 613-724-6234 RFDavies@ ottawaheart.ca
Deputy Project Manager, Technical Team Leader	Mr. Sonny Lundahl	Vice President, Research and Development, AMITA Corporation	613-742-6482 613-742-8188 sonnyl@amita.com
Financial Officer	Mrs. Roxanne Proulx	Finance Officer, Institute of Information Technology, National Research Council of Canada	613-993-2494 roxanne.proulx@ nrc-cnrc.gc.ca
Procurement Lead	Ms. Debbie DeLeavey	Head, Services and Construction Contracting, National Research Council of Canada	613-990-0874 Deborah.deleavey@ nrc-cnrc.gc.ca
Project Team, Technical Team	Dr. Norman Vinson	Associate Research Officer, Human Computer Interaction, Institute of Information Technology, National Research Council of Canada	613-993-2565 Norm.vinson@ nrc-cnrc.gc.ca
Project Team, Technical Team	Dr. Joel Martin	Group Leader, Interactive Information, Institute of Information Technology, National Research Council of Canada	613-990-0113 Joel.Martin@ nrc-cnrc.gc.ca

Position	Name	Title	Phone Number and email address
Project Team, Technical Team	Dr. Patrick Maupin	Defence Scientist and Group Leader, DRDC Valcartier	418-844-4000 Patrick.maupin@drdc-rddc.gc.ca
Project Team, Health Care/Public Health Team	Dr. Suzanne Desjardins	Director, Office of Drugs and Alcohol Research and Surveillance, Controlled Substances and Tobacco Directorate, Health Canada	(613) 946-4223 Suzanne.desjardins@hc-sc.gc.ca
Partner, Health Care and Public Health Team	Dr. Natasha Crowcroft	Director, Surveillance and Epidemiology, Ontario Agency for Health Protection and Promotion	416-314-5188 natasha.crowcroft@oahpp.ca
Project Team, Health Care/Public Health Team	Dr. Mark Raizenne	Director, Foodborne, Waterborne and Zoonotic Infections Division, Public Health Agency of Canada	613-948-6883 Mark_Raizenne@phac-aspc.gc.ca
Project Team, Health Care/Public Health Team	Dr. Amin Kabani	Director, CNPHI Project, National Microbiology Laboratory, Public Health Agency of Canada	204-792-6183 Amin_Kabani@phac-aspc.gc.ca
Project Team, Health Care/Public Health Team	Dr. Isra Levy	Medical Officer of Health, Ottawa Public Health	613-580-6744 x23681 Isra.Levy@ottawa.ca
Project Team, Health Care/Public Health Team	Dr. Kieran Moore	Director QPHI Project, KFLA Public Health, and Associate Professor of Medicine, Queen's University Department of Emergency Medicine	613-329-1191 Moorek1@ kgh.kari.net
Project Team, Health Care/Public Health Team	Dr. Daniel Neill	Associate Professor of Robotics and Computer Science, Carnegie Mellon University	412-621-2650 neill@cs.cmu.edu
Project Team, Health Care/Public Health Team	Dr. Raphael Saginur	Ottawa Hospital, Division of Infectious Diseases, Chair Ottawa Hospital Research Ethics Board	613-798-5555 ext 14155 rsaginur@ottawahospital.on.ca
Project Team, Health Care/Public Health Team	Dr. Kathy Suh	Division of Infectious Diseases, Ottawa Hospital,	613-737-7600 ext2651 ksuh@cheo.on.ca
Project Team, Health Care/Public Health Team (advisory)	Mr. Stephen D'Silva	SilvaCorp Inc.	905-875-1608 Stephen@silvacorp.com

Position	Name	Title	Phone Number and email address
Project Team, Technical Team (advisory)	Dr. Avi Parush	Associate Professor Department of Psychology Carleton University Human Oriented Technology (HOT) Laboratory	613 520-2600 ext. 6026 avi_parush@carleton.ca
Project Team, Health Care/Public Health Team (advisory)	Dr. H.R. Lynn	Medical Officer of Health, Grey Bruce Health Unit	519-376-9420 hlynn@publichealthgreybruce.on.ca
Project Team, Health Care/Public Health Team (advisory)	Dr. Melinda Wilkins	Director, Division of Communicable Disease, Bureau of Epidemiology, Michigan Department of Community Health	517-335-8165 wilkinsm@michigan.gov

Annex B Project Performance Summary

B-1 Technical Performance Summary

Key Technical Goals

Following were the key technical goals of the Data Fusion project:

- Develop a reusable process framework to implement systems to conduct Surveillance
- Develop a software framework for:
  - Data capture and processing
  - Statistical analysis, visualization and mapping, advanced analytic techniques
  - Applicability to diverse situations for problem solving

The frameworks were to be developed in the context of two surveillance scenarios:

- – Detection of In-Hospital Disease Outbreaks
- – Detection of Harm Related to Illicit Drug Use

Key Technical Accomplishments

The Data Fusion project resulted in the development of the following capabilities:

a. Real-time identification and tracking of user-defined disease syndromes using EHR data streams.
   i. In Hospital
   ii. Outpatients (e.g. microbiology reports for infectious disease surveillance)
b. Conversion of data streams into useful surveillance information.

A re-usable framework and several tools were developed:

- HL7 Data capture and organization
- Privacy protection framework
- Categorization and classification of data elements
- Creation of syndromes relevant to specific problems
- Statistical analysis and reporting
- Mapping and visualization
- Risk factor analysis
- New measures for outbreak detection, new decision fusion techniques, survey of distance measures in the framework of evidence theory
- Feature extractor, selection and data fusion for time series

The project addressed the Risk Assessment and Priority Setting priority area, specifically RA 1, RA 2 and RA 3. The performance is summarized in the following table:

Table: Data Fusion Priority Areas

Specific Priority	Description of Gap	Data Fusion Project Target to address Gap	Attained
RA 1 - Science dimension of risk	• Examine the differences between risk domains (C, B, RN, and E) and attempt to create a common reference frame for assessing the risks across these domains. • Explore the quantification of risks across various domains.	The project combines DRDC's expertise in Situation Analysis, Monitoring (SAM) and Data Fusion (DF) with that of the other team members to develop a service oriented CBRNE threat detection and monitoring framework that will allow responders to implement advanced RA2 solutions that effectively bridge RA1 and RA3.	Success: **Full** Rationale: Two scenarios and prototype systems have been developed (detection of severe infections in hospitalized patients and detection of harm related to illicit drug use) that address specific gaps not covered by existing systems. These are highly relevant to the detection of both bioterrorism and naturally occurring disease outbreaks. Technological development focused on Data Fusion, supporting decision-making processes, allowing efficient human-system interactions and moving electronic threat monitoring into its next stage of development. The Data Fusion Surveillance system developed allows responders to rapidly evaluate potential threats, respond appropriately to incoming alerts and permit analysts to explore relationships between data streams and thus enhance their ability to extract relevant features from the environment. The project leveraged the knowledge of all project partners to develop a statistical threat monitoring capability applicable to multiple domains of risk. The Data Fusion surveillance capability that was developed will promote public confidence and trust by providing new sources of credible information relevant to CBRNE risk.
RA 2 - risk cataloguing, modeling/visualization	• Propose new concepts for the capture and inventory of risk related data, with a view to supporting modeling and visualization of the risks across the domains. • Adapt or develop new techniques for representing multiple risks visually in a geographical reference frame.		
RA 3 - threat proliferation monitoring	• Explore the sciences, techniques and concepts behind foresight and future visioning that would support risk assessment and capability goals development. • Examine guidelines, protocols, tools and techniques for the monitoring of threats through various concepts/approaches, such as knowledge mining.		

Technology Readiness Level of Deliverable (TRL)

The project technology started at a TRL Maturity of 3 and moved to a 5 by the end of project. The estimated time to reach TRL7 Maturity is 36 months.

Advantages Over Existing/Competing Technologies

Following are the advantages of the technology developed under the Data Fusion project compared to other existing technologies:

- The framework and tools developed are adaptive and leverage existing and emerging EHR technology infrastructure
- The framework and tools developed can be applied to diverse problem areas

B-2 Schedule Performance Summary

The Data Fusion Project started on May 1st, 2009 and version 1.0 of the Project Charter was completed by July 15th, 2009. The document planned the completion date of the project for November 30th, 2011.

Unfortunately significant delays in securing access to Personal Health Information (PHI) data and difficulties in negotiating data sharing agreements between partners caused slippage of the original project schedule. As a result, the project team suggested that the project duration be extended until March 31st, 2012. Members of the Project Review Committee, at the meeting that was held on December 9th, 2010, recommended that a formal project extension request be submitted to CRTI, which subsequently approved the change to the project completion date. Version 3.0 of the Project Charter was issued to reflect schedule delays and adjustments in milestone dates.

Although the Project Team encountered further delays in acquiring access to PHI data in 2011, most of the milestones were completed on time (before March 31st, 2012). The project team decided to ask for a second extension of the project until December 31st, 2012 to enable the team to refine the results, finalize a close-out report and organize a final Project Review Committee meeting.

The second extension was granted by CRTI. During this additional time (April – December 2012) the Project Partners contributed in-kind to put final touch-ups on the final report and formally close the project.

The Data Fusion project was officially closed at the Project Review Committee meeting, held at NRC on November 9th, 2012.

In summary the project was completed with an approximately 12% variance in schedule.

B-3 Cost Performance Summary

The Data Fusion project has been completed at a total cost of $3,572,094, which is about 15% over the initially planned $3,108,502 (version 1.0 of the Project Charter). This cost increase, absorbed by in-kind contributions of the Project Partners, was caused by unexpected delays in gaining access to PHI data and difficulties in negotiating multiple data sharing agreements. As mentioned in the previous section these difficulties contributed to the schedule slippage. That in turn was a major cause of some cash flow problems with respect to CRTI funds.

The Project Team requested two rollovers of CRTI funds from a previous fiscal year to the next and received approval from CRTI/CSS Authority. The amounts of the rollovers were:

- $86,396 from 2009/10 FY into 2011/12
- $286,326 from 2010/11 FY into 2011/12

These changes in the cash flow were incorporated into the subsequent versions of the Project Charter.

In summary, 58% of the $3572094 total cost was covered by CRTI/CSS funds and 42% was covered by in-kind funds of the Project Partners.

Annex C Publications, Presentations, Patents

Richard Davies, University of Ottawa Heart Institute, Data Fusion presentation at Public Security S&T Summer Symposium, Ottawa June 14, 2012, CRTI 08-190RD: Data Fusion Solutions for Monitoring CBRNE Threats

Ko, A., Jousselme, A.-L, Maupin, P., A Novel Measure for Data Stream Anomaly Detection in a Bio-surveillance System, Int. Conf. on Information Fusion, Chicago, IL (USA), July 2011.

Jousselme, A-L., Maupin, P., Distances in evidence theory: Comprehensive survey and generalizations, International Journal of Approximate Reasoning, Available online 31 August 2011 (20 pages).

Davies, Richard F., Morin, Jason; Bhatia, Ramanjot; deBruijn, Lambertus A System for Surveillance Directly From the EMR, has been accepted for inclusion in the 2012 International Society for Disease Surveillance (ISDS) Annual Conference program. The ISDS Conference will be held December 4-5 in San Diego, California

Bibliography

DATA FUSION PROJECT CHARTER 08-0190RD TO THE MEMORANDUM OF UNDERSTANDING CONCERNING THE CHEMICAL, BIOLOGICAL, RADIOLOGICAL OR NUCLEAR AND EXPLOSIVE (CBRNE) RESEARCH AND TECHNOLOGY INITIATIVE (CRTI) , DATED 5 March 2012, VERSION: 6.0

CRTI Proposal # CRTI-08-190RD, Data Fusion Solutions for Monitoring CBRNE Threats (DF-Surveillance), 2008-2009 CRBNE Research and Technology Initiative (CRTI), Full Proposal Application Research and Technology Development- RD

List of symbols/abbreviations/acronyms/initialisms

Table 3: Abbreviations, Acronyms, and Initialisms

ACR	Ambulance Call Report
ADT	Admission Discharge Transfer
ASSET	Advanced Syndromic Surveillance and Emergency Triage
BUN	Blood Urea Nitrogen
CBC	Complete Blood Cell count
CBRN	Chemical, Biological, Radiological, Nuclear
CBRNE	Chemical, Biological, Radiological, Nuclear, Explosive
C-DIFF	Clostridium difficile
CEWS	Canadian Early Warning System
CHEO	Children's Hospital of Eastern Ontario
CHEO RI	Children's Hospital of Eastern Ontario Research Institute
CLI	Central Line Infection
CNPHI	Canadian Network for Public Health Intelligence
COTS	Commercial off the Shelf
CRTI	CBRNE Research and Technology Initiative
CSS	Centre for Security Science
CTAS	Canadian Triage Acuity Score
DND	Department of National Defence
DF	Data Fusion
DRDC	Defence Research and Development Canada
ECADS	Early CBRN Attack Detection by Computerized Record Surveillance
EHR	Electronic Health Record
ER	Emergency Room
ESB	Enterprise Service Bus
EDMS	Enterprise Database Management System
GBHU	Grey Bruce Health Unit
GPL	General Public License
FIAC	Future Intelligence Analysis Capability
HL7	Health Level Seven
HCI	Human Computer Interaction
ILI	Influenza Like Illness
IC	Infection Control
IP	Intellectual Property
IT	Information Technology
IS	Information Services
KFLA	Kingston, Frontenac, Lennox & Addington Health Unit
KGH	Kingston General Hospital
MDCH	Michigan Department of Community Health
MIRTH	Open source enterprise service bus
MRSA	Methicillin Resistant Staphylococcus Aureus
MRSA C	Methicillin Resistant Staphylococcus Aureus Colonization

MRSA I	Methicillin Resistant Staphylococcus Aureus Infection
NRC IIT	National Research Council Canada, Institute for Information Technology
OAHPP	Ontario Agency for Health Protection and Promotion
OCRI	Ottawa Centre for Regional Innovation
OHDW	Ottawa Hospital Data Warehouse
OPH	Ottawa Public Health
OPI	Office of Primary Interest
OPS	Ottawa Paramedic Service
PA 1	Prototype Application #1
PA 2	Prototype Application #2
PHAC	Public Health Agency Canada
PHO	Public Health Ontario (formerly OAHPP)
PHIPA	Personal Health Information Protection Act
PSTP	Public Security Technical Program
QPHI	Queen's University Public Health Informatics
R&D	Research & Development
S&T	Science and Technology
SARS	Severe Acute Respiratory Syndrome
SME	Subject Matter Expert
TOH	The Ottawa Hospital
UOHI	University of Ottawa Heart Institute
UTM	Universal Transverse Marcator
VAP	Ventilator Associated Pneumonia
VRE	Vancomycin Resistant Enterococcus
WBC	White Blood Cell count