

Kamran Sartipi (Adjunct Associate Professor), Norm Archer (Professor Emeritus)  
Bellraj Eapen (PhD Candidate) & Eduardo Lopez (PhD Candidate)

[Information Systems, McMaster University](#)  
sartipi@mcmaster.ca, archer@mcmaster.ca

### *Proposal for Data and Service Intelligence Research Lab*

The proposed “Data and Service Intelligence Research Lab” (short name *DataIntel-Lab*) will provide an advanced research and development datacenter for researchers, graduate students, and post-doctoral fellows to perform research in data science, decision systems, cloud services, and big data platforms. The cloud-based open APIs (application program interface) services in this lab will allow open access by other researchers and industries. As a result, the lab will provide “data-intelligence as-a-service” for its clients to use directly, or use the available infrastructure to develop their own data analytics and decision systems. The techniques and solutions developed in this lab are generic and applicable to different application domains, however, the target domains include medical informatics, cybersecurity, social networking, marketing, and traffic control. Three active research projects in this lab are described below.

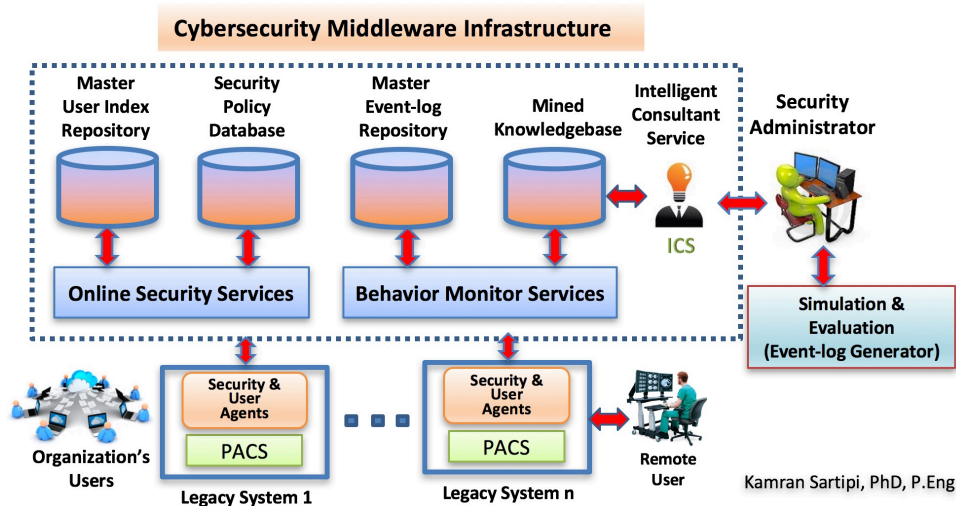
Overall, the research activities in this lab span different aspects of provisioning an intelligent middleware infrastructure for knowledge-driven and customizable decision support systems through cloud and mobile services. The salient characteristics of this infrastructure are as follows:

- *Big data analytics*. It utilizes different machine learning and data mining techniques to process large and heterogeneous datasets using Apache Spark platform to provide fast in-memory and real-time data analytics power.
- *Knowledgebase*. It contains an ontological knowledgebase with the representation format and annotations appropriate to the target application domain.
- *Decision system*. It utilizes virtualized intelligent decision services with the capability of exploring the knowledgebase in order to provide selective and non-overwhelming consultation guides to the users.
- *Customization*. It provides customizable generic agents in mobile devices which invoke cloud-based decision services to effectively assist the users and utilize visually encoded technologies.
- *Simulation*. It provides a domain-specific user-behavior pattern language and an event-log generation engine to synthesize a big data simulation datacenter, which allows the researchers to evaluate their data analytics approaches on an experimental infrastructure before applying on sensitive production datasets.
- *Access control*. It utilizes OpenID Connect access control standards to provide secure cloud-based authentication and authorization mechanisms.
- *System and Semantic Integration*: It provides facilities for integration of legacy and new systems including RESTful messaging and semantic mappings of terminologies.

We have applied the above infrastructure on three major projects, to: i) assist security administrators of large distributed systems to identify malicious user-behaviors to enhance the system’s security policies; ii) assist physicians in underserved and poor regions to access medical knowledge of specific medical

specialties to enhance their care provisioning; and iii) utilizing crowdsourcing to generate clinical knowledge from best-practice evidence in healthcare.

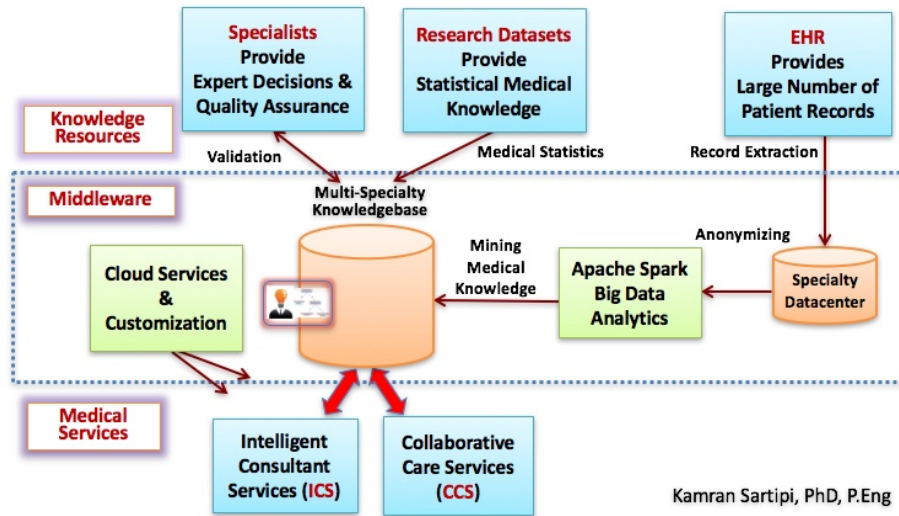
### *Intelligent Consultant-based Cybersecurity Middleware Based-on User-Behavior Pattern Discovery*



This research project (shown in figure above) provides a security middleware infrastructure for a large distributed medical system including several legacy and new PACS (picture archiving and communication systems) that desire to securely integrate their services and share medical images and reports. In such an environment, the authenticated and authorized users (trusted users) in one PCAS system can access to the resources of other systems with no proper security control that cause serious damage to such a sensitive information system. The security administrators cannot effectively monitor and control such situations due to huge volume of dynamically changing user behaviors. The objective of this research is to assist security administrators to enhance the security policies of such a large distributed system. To achieve this goal, a harmonized set of scientific techniques from data mining, knowledge engineering, constraint pattern matching, customizable agents, and decision support systems provide an effective and non-overwhelming consultation service for monitoring user behaviors and enhancing security policy rules. This service provides step by step consultation to the security administrator on the suspicious behaviors to identify affected users, resources, and locations precisely. A data analyst utilizes our behavior query language (BQL) to specify high-level clusters of events and their constrained relationships which allows for extracting user behavior patterns using data mining techniques to populate a knowledgeable. Our behavior pattern language (BPL) is used to compose suspicious complex user-behavior patterns to be searched for approximate matches in the event-log repository to identify anomaly behaviors and the involved resources and locations. The proposed infrastructure also utilizes cloud-based open access authentication and authorization mechanisms for secure sharing of documents and resources among heterogeneous legacy PACS systems.

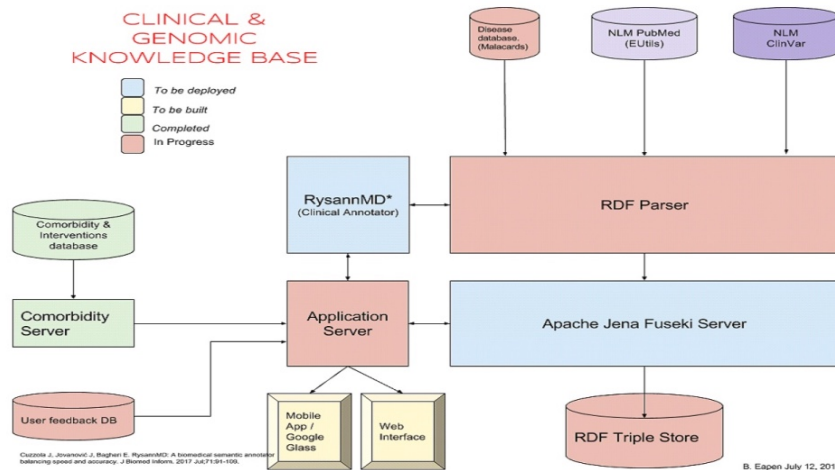
### *A Cloud Infrastructure for Intelligent and Customizable Medical Consultation Provisioning*

*Infrastructure for Intelligent and Customizable Digital Health*



This research project (shown in figure below) provides a datacenter and corresponding techniques and tools to extract information from heterogeneous sources, process information to generate mined-knowledge, and integrate and store them into a specialty knowledgebase within the middleware infrastructure. This knowledgebase is further utilized by a cloud-based intelligent decision support system which provides effective, personalized and customizable assistance to medical professionals in diagnosis and treatment of their patients. The proposed datacenter utilizes three sources of medical knowledge which will be validated, integrated, properly annotated, and stored in the specialty knowledgebase using a new knowledge representation format we developed. The cloud services provide customization of the mobile devices for a new type of knowledge-driven and context-based decision support services (called consultant services), and for different collaborative tasks among physicians.

*Clinical Knowledgebase (ClinBase): Bringing Evidence to the Bedside by Crowdsourced Benchmarking of Fuzzy Clinical Rules.*



Clinical research generates evidence for best practices in healthcare, but the domain experts find it difficult to keep updated with the latest evidence from the published biomedical literature. ClinBase is a framework for semantic annotation and extraction of clinical rules from published research and storing them as RDF. The framework provides web and mobile interface for domain experts to benchmark the extracted rules to reflect clinical experience. The system includes a diagnostic clinical decision support system that utilizes the benchmarked rules to interpret patient history. The system is being prototyped for Dermatology.

Below, some relevant projects within our research team are briefly described, with links to the PDF publications.

- **Intelligent middleware security provisioning.** A secure, central and service-based “intelligent middleware” consisting of: multi-agent technology (smart local agents and administrative middleware agents) for two-level decision-making process; central policy repository and management; central metadata repository for images; and a centralized authentication and decentralized authorization model [[j12](#), [j13](#), [js4](#), [c48](#), [c46](#), [c45](#), [c44](#), [c42](#), [c22](#)].
- **Smart decision support systems.** This project aims at providing a new generation of decision support systems where mined-knowledge at decision points (as reminders, alerts, recommendations) will assist the physicians (for patient diagnosis) or administrators (for resource allocation) to make effective decisions. In this context, mined-knowledge refers to the extracted patterns and trends from clinical data using data mining techniques. This research covers both rule-based and flow-based decision support techniques [[j10](#), [j4](#), [ch3](#), [ch2](#), [c43](#), [c41](#), [c31](#), [c13](#), [c11](#)].
- **Knowledge-driven user behavior-pattern discovery.** This research provides an intelligent decision support system that effectively assists the system administrators to obtain deep insight into the system user's dynamic behavior patterns in order to refine the existing security policies using a new behavior pattern query language (BPQL) [[j9](#), [j14](#), [j8](#), [c49](#)].
- **User behavior simulation environment.** In this research, we developed an event-log generator engine, which receives administrator-defined user-behavior patterns using our behavior pattern language (BPL) and produces corresponding events in the context of noise events which allows us to effectively test and fine-tune the above techniques before applying them on the production event-logs. [[j11](#), [j14](#)].
- **Service customization in SOA.** This research enhances data privacy and security, reduces network traffic, and provides new enterprise level features. It introduces two new concepts “task service” and “service representative” in the SOA environment. Task service is a multi-component (model, knowledge, data) web service that can process the client data locally at the client side. Service representative is a generic agent at the client side that will be customized by the knowledge component and will execute the model component on both client and task service data. [[j6](#), [js2](#), [js1](#), [c39](#), [c38](#), [c36](#), [c35](#), [c34](#), [c32](#)].
- **Mobile eHealth.** This project provides efficient techniques to integrate different devices such as cell phones, tablets, and specialized devices such as pacemakers to be used seamlessly with other software services of the electronic health record (EHR) systems. In this approach data will be

collected, maintained, analyzed and communicated using HL7 information models and messaging. [[c35](#), [c34](#)].

- **Knowledge translation to HL7 FHIR messages.** This project provides a well-defined process to guide translation of healthcare transactions (obtained from healthcare scenarios) to HL7 Interactions. The approach re-categorizes Interactions based on their behavioral traits in a messaging context and uses semantic analysis to automate the process of identifying such interactions [ [c40](#), [j7](#), [j2](#), [c30](#), [c25](#), [c21](#), [c19](#)].