STATE OF CONNECTICUT
State Innovation Model
Health Information Technology (HIT) Council
2nd Submission: Zato Follow up Questions

1) Is Zato able to provide a healthcare demonstration of their de-identified EHR-indexed solution? Could we visit the BayState Innovation Center to see Zato working in conjunction with a Cerner system?

   Zato would be happy to provide a demonstration of one of more applications of the Zato Health Interoperability Platform software on de-identified patient records (discharge summaries). We would be happy to do so at a meeting place in Connecticut, which is convenient for Committee members. The Baystate Health Innovation Center sponsors open house meetings and visits and demos of activities. The planning for one of the upcoming events includes a demonstration of the Zato Health DRG Dashboard application for a bundling application with Diagnostic Related Groups. The application also reports the relative risk of patients for a particular DRG, in order of those patients with the highest risk based on automated analysis of patient records. Zato has also recently created on IBM hardware at multiple IBM facilities in New York and Massachusetts a demonstration of cross-clinical and genomic data analysis over networks across multiple organizations and data centers using de-identified clinical and genomic data. Baystate Health Innovation Center welcomes visitors to these events. The headquarters offices of Zato Health are located within the new Baystate Health Innovation Center.

   It will not be possible to demonstrate the use of Zato Health software with Baystate Health Cerner data until the Cerner data are de-identified. The data are actual patient records and contain PHI, so our engagement is restricted to internal use. The de-identified Baystate Health medical records data will need to be approved for release by Baystate Health before they could be shown in a public demonstration. Zato Health is working actively on a task under the same Baystate Health contract Statement of Work to implement a de-identification process that could enable that demonstration of de-identified Cerner data.

2) Please provide us with more details on how Zato will work with multiple data sources in terms of indexing/pulling of the data, and the associated timeframes to complete.

   The indexes are normalized representations of the data sources (the application repositories - 'separating the data from the application.' Once the indexes are built, an authorized user can query and analyze across one or many or all indexes in parallel, subject to security constraints on the data and access privileges of the user.

3) Please explain what is stored in the indices. For example, diagrams, examples and a physical demonstration would be useful to understand the data stored in the indices.

   Alpha-numeric data, pointers, and statistics are stored in the indexes. Zato software is read-only with EMR application data, so it is not invasive to existing production data systems.

4) What is the impact on the provider resources in the short term and long term to support the proposed Zato solution?

The indexing process and the query/analysis process run on the edge servers, which carry the load. In terms of putting load on the production EMR servers, the initial pre-indexing process that runs SQL read queries on the productions systems can be automatically scheduled to run on a metered basis during lower-load hours. Alternatively, to be completely unobtrusive, the pre-indexing can be done on another server using a copy of the EMR database. That was the approach preferred by Baystate Health with their Cerner data.

5) What is the tool to store data and perform reporting of that data? How is the data mart or warehouse configured with the system and what tools are required to report from the data mart?

Zato Health creates normalized indexes representing the data stored within diverse applications. The indexes are used for interactive views, analysis, and reporting and for batch analysis and reporting. The load is carried in the index servers ('edge servers'), so that the supplemental capabilities do not overload existing production systems. Zato can index data stored in file systems or tabular databases maintained by applications. Zato can also index the data stored within multiple centralized or de-centralized physical data warehouses. Zato does not create a physical data warehouse. Index servers compute in parallel across the data stored within diverse application repositories and share results cooperatively over networks to create of virtual data warehouses without copying and moving all the data to a central location. The Zato software also can send results to other software applications through a published JAVA API.

In addition to retrospective analysis and reporting from indexes, the Zato Interoperability Platform software also includes a high volume, high speed real-time filtering and routing engine, which searches, analyzes, routes, and sends alerts while processing real time data streams that arrive moment to moment and may not be indexed yet.

6) What analytics capabilities are available for immediate use with Zato (ie. Starter Set)? What tools are available to build on the analytics starter set? What language are they written in?

Ad-hoc analysis can be run against any or all indexes across structured and unstructured data at the same time based on security access constraints, using a combination of structured and unstructured operators. Operators include probabilistic inferencing, concept/entity recognition/extraction, proximity, part-of-speech recognizers, and normalized regular expressions for text rich data; Boolean, numeric, date, range operators for fielded data; and SQL for indexing from relational databases. End users typically interact with screen forms and menus. Developers have access to a published JAVA API. Minimal training is required by end users creating queries. Licensees have the tools necessary to create reporting applications or integrate with third party applications. Applications that are created by Zato Health for the Interoperability Platform are distributed to licensees at no charge.

7) There is concern that normalization could add inherent distortion into the data. How is the data normalized? Who is involved in performing the normalization? How much time is required to perform this process?

Normalization can be done at indexing time or on the fly in ways that do not distort the source data. For data tags (field names, etc), an administrative user interface manages the field mapping process. For the data values, the data are normalized to be consistent with regard to integer or floating point values and formats, such as multiple date formats for the same date. The tools are included. The only manual process is the initial schema analysis and field mapping. The amount of time is needed is relatively low and depends on the complexity of the table space.

8) Is the data always kept behind the site's fire wall? There are concerns of privacy and patient consent. For example, is the data encrypted at rest and in transport?

Sensitive data do not move from the location where they are stored, processed, and protected. The indexes are stored in the same location behind the same firewall. The data are always encrypted end-to-end during network transmission. The data can also be encrypted at rest if desired. PHI data that may be contained in Zato indexes are protected the same as production data.

9) Many of the states are using the state HIE or a centralized database. What are the advantages of edge servers relative to these solutions beyond keeping the data behind the firewall?

Data modeling and ETL, consolidation, and aggregation in a separate centralized facility represents the most expensive, time consuming, and rigid method for attempting the capabilities required of an HIE. Making copies of PHI data and trusting other parties to manage the data increase risk. In most cases, well over 90% of data do not need to be centrally processed. The requirements and specifications of an HIE are better served in most respects by a secure network based de-centralized Healthcare Information Sharing Environment (HISE) accessed by cooperating organizations across diverse applications than by traditional data aggregation techniques.

10) Do you follow a "proof of concept" methodology? If so, can you describe the major steps of this methodology and where it has been used? Is there an example work plan you might share?

All systems built with any components of Zato's software suite have been created after a pilot engagement using a proof-of-concept methodology and subsequent evaluation before a production roll-out. The methodology for the proof of concept varies to some extent from organization to organization. Common methodology steps include creation of a user group and use cases for initial workflow analysis and subsequent testing and assessment; initial IT system and network loading analysis with the IT team and subsequent testing; initial access security analysis with the security team and subsequent testing; initial privacy protection analysis and subsequent testing.

11) How does your work in the intelligence industry translate into healthcare? In particular it would seem that provider and vendor expertise would be needed to work on the schema and mapping for the indexing. If not, how does the software understand syntactic and other differences in the data?

We agree that the input, feedback, validation, and evaluation of the customer is an important factor to the success of an implementation. There are many similarities between secure information sharing, knowledge discovery, and knowledge extraction for intelligence analysis and healthcare analysis. Some of the similarities are listed below.

- Mission critical applications with potential for life or death implications
- Role-based security access control, data protection, and auditing in both domains
- Privacy protection and auditing are paramount under HIPAA and the US Privacy Act
- Global views needed across data silos - many diverse sources in multiple data centers and organizations
- Expensive, time consuming, and risk increasing to model, copy, move, ETL, and centralize data
- Content ownership is better controlled by leaving data where they are rather than proliferating copies
- Big Data volumes - billions of structured database records and text rich documents research documents
- Ease of use - typical physicians, nurses, intelligence agents are not data scientists
- Timeliness - response latency for query and analysis within seconds at massive data scale
- Productivity - health care providers needs to achieve the same dramatic labor productivity gains realized by the Intelligence Community to be cost effective - unified interface, single pane of glass
- Accountability - consistency and ease of analysis, result verification, and auditing

12) What individual and state-wide challenges should we anticipate based on your experience implementing your system for other customers?

We can anticipate several challenges to individual and state-wide implementation of any new systems designed to improve data interoperability among healthcare information applications within a provider organization and across multiple providers and payers. The challenges described below are similar to challenges experiences by our team in improving data interoperability among national security organizations. Each of these challenges were overcome in that domain, and the explanations below will include examples of how these challenges can and will be overcome in healthcare. Overcoming these challenges to healthcare data interoperability is the essential enabler to facilitate comparable and verifiable reporting of quality measures by providers in order to achieve measurable quality of care improvements and cost effectiveness improvements for healthcare.

1. **Innovative new technology.** Resistance to new technology innovation is typical in any domain until the implementation and benefits of the new technology innovations are proven by early adopters. Every application in the FBI was a data silo until the office of the CTO became an early adopter and implemented new techniques. A subsequent study measured that certain tasks that had previously taken 32,000 hours to complete were being completed with the new data interoperability software in 30 minutes or less by a single analyst. Similar resistance to new technology in healthcare was heard in a classic comment in the first HIT Council meeting that we attended, in which an individual commented that because the new technology was not yet in wide use, how could it do things that they have not done before.

2. **Staff availability.** Connected to number 1 above is the factor of no staff being available to try new data interoperability innovations, so the problems persist. The explanation is typically that current staff are already 100% utilized and therefore, there are no staff available to implement and

prove anything new that could result in improvements over the existing methods. In the first introduction of new technology for a national security customer, thousands of staff were occupied with existing applications running existing technology in stove pipes. The first allocation of staffing to the new interoperability technology was just .5 FTE – one half of one full time engineering position. That staffing was soon increased and the staffing of existing applications decreased as existing approaches failed to show improvement and the new approach showed both mission improvement and improved user satisfaction.

3. **Available budget for innovative solutions**. Funding a new cost to achieve a new benefit is a classic chicken and egg problem. In the Intelligence Community, there were two examples of pivotal catalyst funding to improve data interoperability. The first example was an injection of a huge amount of counter-terrorism funding after 9/11. In that example, over 90% of the available new funding fir data interoperability was squandered by allocating those funds to an older existing technology. The FBI 'VCF' initiative spent over $200 million to try to solve the data interoperability problem with older technology (http://en.wikipedia.org/wiki/Virtual_Case_File). Meanwhile, the CTO and Deputy CTO of the FBI took a small budget 'out of hide' to fund the most successful data interoperability system in the history of the FBI using new technology implemented by members of the Zato Health team: http://en.wikipedia.org/wiki/Investigative_Data_Warehouse. Success stories included press stories from CNN, Washington Post, CBS Evening News, and powerful Congressional testimony by Directors of the FBI and the Treasury Departments Financial Crimes Enforcement Network. The second example was the funding of a data interoperability pilot by the Department of Defense. With a budget of one million dollars over a two-year period, members of the Zato team implemented a successful data interoperability capability across data sets of multiple DOD agencies. The highly successful and lauded results are summarized in a report recently submitted to the Senate Armed Services Committee.

4. **Resistance to change by existing software and services vendors.** In spite of Congressional testimony by EHR vendors that their EHR applications are open and not obfuscated, the practical reality is that each of the largest EHR applications is a data island or silo to each other application and to other healthcare data applications. While these EHR systems provide excellent data collection systems for the respective population in its database, these systems represent barriers to meeting data interoperability challenges. Recent Congressional testimony several weeks ago highlights this problem http://www.fierceemr.com/story/legislators-grill-karen-desalvo-interoperability/2015-05-06. Baystate Health chose to be an early adopter by initiating a pilot with Zato Health, in which Zato has just been issued the task to index the data stored in a 13 Terabyte Cerner EHR. Baystate Health has targeted 'data liquidity' and the need to be able to generate reports more flexibly and affordably across EHR data stores and other healthcare application repositories. Data interoperability success stories across EHR repositories and other healthcare databases systems will prove and demonstrate not only a direct cost-benefit advantage for implementations, but also a direct path to fund from cost savings and revenue generation the healthcare data verifiability and accountability for State, Federal, and Insurance payers needed for cost and quality comparisons among providers - resulting in both improved quality and improved cost effectiveness of healthcare services to the benefit of patients and payers and likely prompting some EHR vendors to make their data more accessible to see the opportunity for greater market share from data interoperability.