Designing a Reference Model for Learning Analytics Interoperability

Jae-Hyeong BAE[a], Yong-Sang CHO[b], Jaeho LEE[c*]
[a]*School of Electrical and Computer Engineering, University of Seoul, Republic of Korea*
[b]*Korea Education & Research Information Service*
[c]*School of Electrical and Computer Engineering, University of Seoul, Republic of Korea*
**Corresponding Author: jaeho@uos.ac.kr*

Abstract: The application of the learning analytics should overcome the challenge of effective integration of various information and processes into a unified framework to support the development of an open and extensible learning analytics systems. Based on our preliminary identification of the requirements for the learning analytics systems, we present a refined reference model that fulfills these requirements and provide a guideline for effective integration toward the goal of producing an explicit specification of the learning analytics architecture. We elaborate the comprehensive architecture with detailed descriptions, and then we discuss the experimental implementation of the reference model.

Keywords: Learning Analytics, Reference Model, Software Reference Architecture

1. Introduction

In spite of growing interest for learning analytics, the application of the learning analytics still faces the challenge of effective integration of various information and processes into a unified framework to support the development of an open and extensible learning analytics systems. In our preliminary analysis(Choi, 2014), we presented preliminary identification of the requirements for the learning analytics systems and summarized them as follows:

- **Open and extensible**: It should be open to incorporate new sensors or analytics functionality, desirably without interrupting the task being serviced. It also should ensure incorporation or modification of new workflows at the task level.
- **Distributed**: It should be able to handle multiple sources of data and functionalities distributed over multiple systems. It is also desired to be able to distribute data and to delegate functionality dynamically and transparently.
- **Interoperable:** It should provide compatibility for various learning platforms or VLE by providing interoperable interface to the data and operations.
- **Reusable and configurable**: The functional components and data interfaces should be modular and thus reused and configured for different tasks or more complex tasks as building blocks.
- **Real-time and predictable**: Learning analytics should be performed satisfying the real-time constraints and should be able to estimate the time to completion.
- **Usable:** It should have acceptable user experience (UX) by providing appropriate data visualization and user interfaces for monitoring and tasking throughout the learning analytics process.
- **Secure and traceable:** It should protect personal user information to secure privacy and preserve confidential information. Some analytics functionality should be ensured not to be performed as required. Furthermore, the history of execution of analytics functions and access to data should be recorded, if needed, to ensure traceability.

Based on these requirements, we also proposed an approach to adopt the major processing steps of big data, that is, data collection, data store and processing, analyzing, and visualization of data. In this paper, we present a refined reference model that fulfills these requirements and provide a guideline for effective integration toward the goal of defining an explicit specification of the learning analytics architecture as the international standard so that open and extensible learning analytics systems can be built for worldwide interoperability.

The remainder of this paper is organized as follows. Firstly, we survey some related works in the learning analytics field. Based on this survey, we present the overall architecture along with the basic requirements of the learning analytics systems. The comprehensive architecture with detailed description then follows. Finally we discuss the experimental implementation of the reference model and summarize the main results of this paper.

2. Related Work

In this we survey several related systems in the learning analytics field and compare our approach with them to refine our reference model.

The Society for Learning Analytics Research (SoLAR) is an inter-disciplinary network of leading international researchers who are exploring the role and impact of analytics on teaching, learning, training and development (The Society for Learning Analytics Research, 2015). The Open Learning Analytics (Siemens et al., 2011) project by SoLAR proposes the Integrated Learning Analytics System as an open platform with the following four major components.

- **Analytics Engine**: the analytics engine is a framework for identifying and then processing data based on various analysis modules.
- **Learning Adaptation and Personalization Engine**: the learning adaptation and personalization will include adaptivity of the learning process, instructional design, and learning content.
- **The Intervention Engine**: the intervention engine will track learner progress and provide various automated and educator interventions using prediction models developed in the analytics engine.
- **The Dashboard**: the dashboard presents visualized data to assist individuals in making decisions about teaching and learning. The dashboard consists of four views: learner, educator, researcher, and institutional.

It is worth noting that these components closely matches the major learning analytics steps in our reference model.

ALAS-KA (José A. Ruipérez-Valiente, Pedro J. Muñoz-Merino, Derick Leony, & Kloos, 2014) is a tool that extends the learning analytics features of the Khan Academy platform which includes visualizations for the entire class and individual students with various learning indicators. It helps teachers to make decision supported by the high level information provided and enables students to gain awareness of their learning process for self-reflection. It also can be used by the course instructors to detect class tendencies and learner models. The major components of the system include,

- **Datastore:** The Google App Engine Datastore provides storage for the Khan Academy platform data.
- **Data processing:** This module is in charge of making the proper computation to transform from different low level data from the Khan Academy models into higher level information that is stored as ALAS-KA models.
- **Visualizations:** The Google Charts API 3 was selected for the visualizations because of its simplicity and variety of charts. In our case, the data needed to build the visualizations are requested to the ALAS-KA models in the Datastore. Therefore, this required data could also be received from an external source such as a web service.
- **Recommender:** The function of the recommender is to analyze the results and send warnings to students or professors based on some rules.

The integration of Khan Academy platform and the ALAS-KA model by utilizing the Datastore and data transformation is the exemplary approach that is worth of adopting in our reference model.

The exploratory Learning Analytics Tool (eLAT) serves teachers to explore and correlate content usage, user properties, user behavior, as well as assessment results. Based on individually selected graphical indicators it supports reflection on and improvement of online teaching methods based on personal interests and observations (Chatti, Dyckhoff, Schroeder, & Thüs, 2012). The typical Learning Analytics process includes data-gathering, mining of the preprocessed data, and the visualization step, which confirms the cogency of the configuration of our reference model. Additionally, this tool provides an insight into what interactions need to be modeled and how interfaces are designed.

3. Reference model for learning analytics interoperability

A reference model, according to the Wikipedia (https://en.wikipedia.org/wiki/Reference_model), is an abstract framework or domain-specific ontology consisting of an interlinked set of clearly defined concepts produced by an expert or body of experts in order to encourage clear communication. In this section, we present a preliminary reference model for learning analytics by illustrating the set of entities and relationships between them as results of the examination of the systems surveyed in the previous section.

3.1 Abstract workflow of learning analytics

The goal of learning analytics to understanding and improve learning and its environment entails the tasks of the measurement, collection, analysis and reporting of data about learners and their contexts, while preserving confidential user information and protecting the identities of the users at the required level as needed. These abstract steps of learning analytics under the protection of privacy policy can be depicted as an abstract workflow as shown in Figure 1.

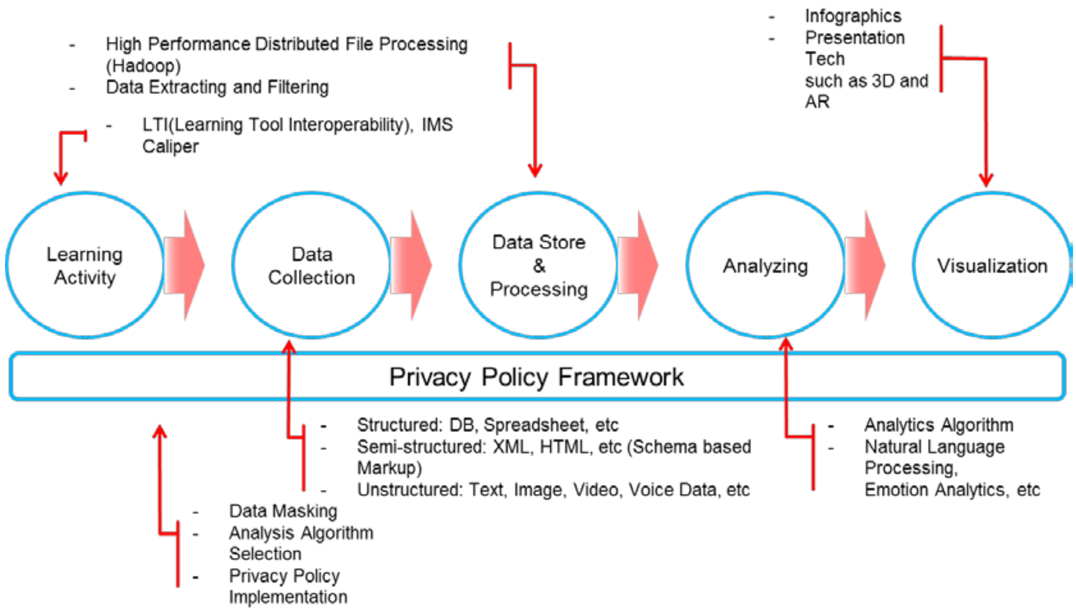

Figure 1. Abstract Workflow of Learning Analytics Service

3.2 Reference architecture derived from workflow and use cases

The abstract workflow in the previous section can be categorized into the four major steps with the associated input/output and processed data, as an overall architecture as shown in Figure 2.

- **Data Collection**: the process of gathering and measuring information on variables of interest in the learning and teaching activities.
- **Data Storing and Processing**: the process of preparing and storing data from diverse and heterogeneous data sources for interoperable data analysis by utilizing the standardized data model and representation.
- **Analyzing**: the process of systematic investigation of learning data by inspecting, and modeling the learning data with the goal of producing descriptive and possibly predictive knowledge.
- **Visualization**: the process of creating visual representation of abstract data including text and geographic information to allow users to see, explore, interact, and understand large amounts of information in analyzing and reasoning about data and evidence.

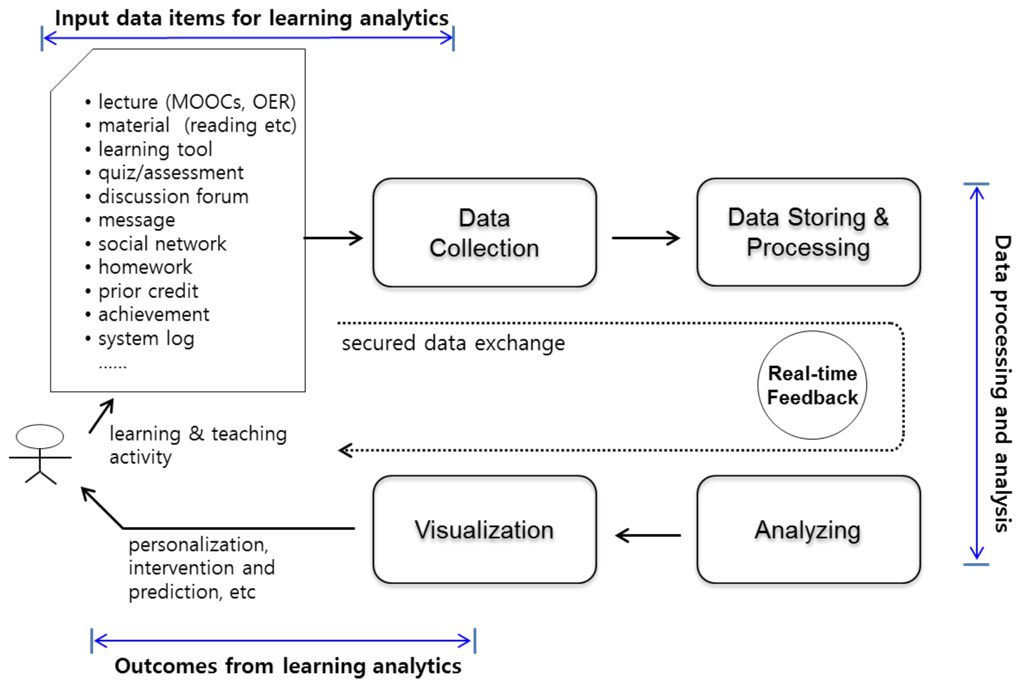

Figure 2: Reference Model for Learning Analytics Service

3.2.1 Zoom-in diagram for Data Collection

Data collection is the process of gathering and measuring information on variables of interest in the learning and teaching activities as shown in Figure 3.

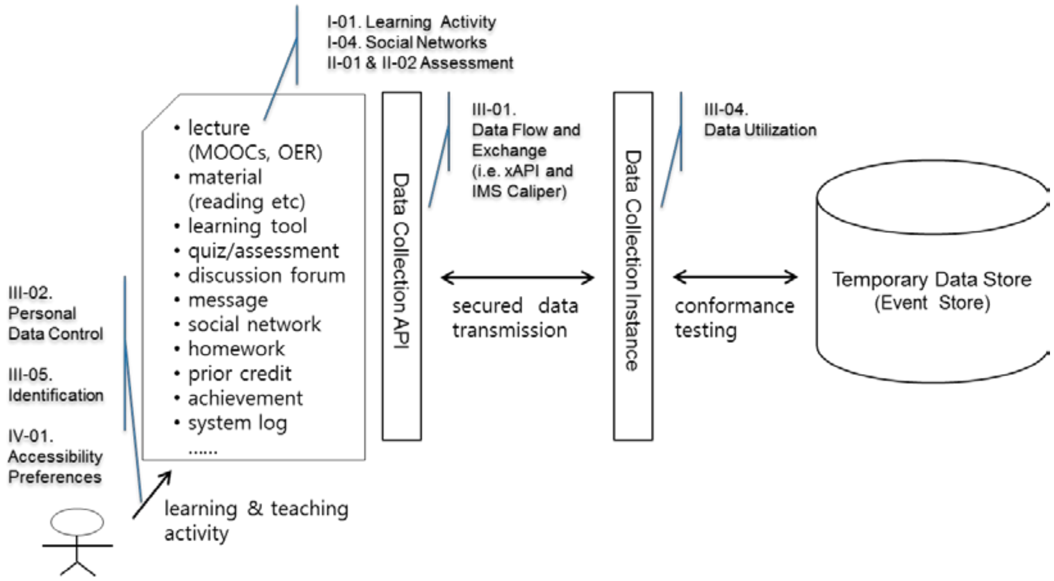

Figure 3: Reference Model of Data Collection for Learning Analytics

- Learning and teaching activities and related data sources such as learning devices and social networks produce various data. The sources include lectures, learning materials, learning tools, quiz or assessment, discussion forum, messages, social network, homework, prior credit, achievement, system log, and so on.
- Diverse learning data need to be collected standardized data collection APIs such as xAPI and IMS Caliper.
- Data collection APIs yields data collection instances, possibly via secured data transmission.

- Data collection instances may go through conformance testing before it is collected in an event store that is a temporary data store, for later processing.

3.2.2 Zoom-in diagram for Data Storing and Processing

Data storing and processing is the process of preparing and storing data from diverse and heterogeneous data sources for interoperable data analysis by utilizing the standardized data model and representation as shown in Figure 4.

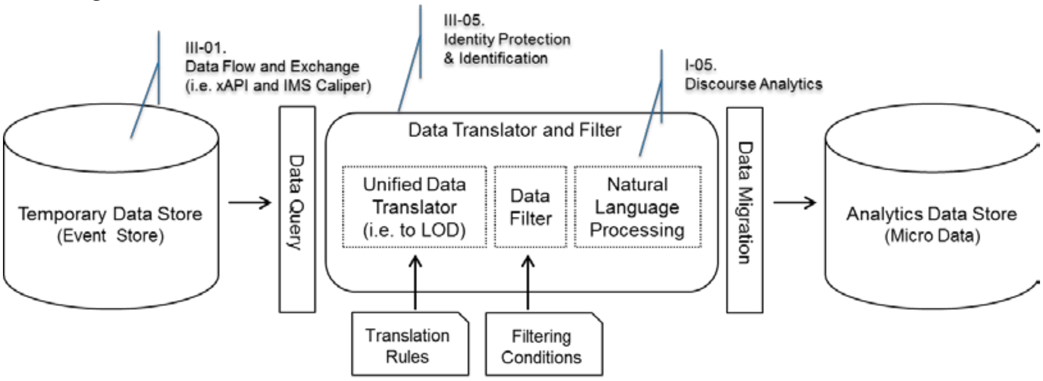

Figure 4: Reference Model of Data Storing and Processing for Learning Analytics

- The learning data stored in Temporary Data Store (event store) are processed by Data Translator and Filter and the processed results are stored into Analytics Data Store (Micro Data)
- The data translator and filter process may have Unified Data Translator which translates various data in heterogeneous representation into a uniform representation, such as LOD, by applying explicit translation rules, for an efficient and interoperable analysis process.
- A general-purpose Data Filter may be applied to the translation process driven by the Filtering Conditions to clean and transform the data.
- One of the main source of data includes discourse, writing, conversation, and communicative events. Such data may need to be processed by Natural Language Processing before the results are in turn translated into a uniform representation.
- The data stored in Temporary Data Store may be accessed via a standardized Data Query interface, and the processed data may be stored to Analytics Data Store via a standardized Data Manipulation.

3.2.3 Zoom-in diagram for Analyzing

Analyzing is the process of systematic investigation of learning data by inspecting, and modeling the learning data with the goal of producing descriptive and possibly predictive knowledge as illustrated in Figure 5.

- As well as the Micro Data stored in the Analytics Data Store, general domain data such curricula, learning resources, and preferences may be stored in Constant Information to be utilized by the Data Analysis.
- Privacy concerns exist wherever personally identifiable information or other sensitive information is collected and stored. Learning data analysis is not an exception. Privacy masking is a way of masking out identifiable information without diminishing the analysis function.
- Various external analysis algorithms such predictive analytics, adaptive analytics, discourse analytics, and other assessment using ICT are applied via Analysis Interface.
- Analysis Processing may consist of statistical analysis, topic analysis, network analysis, and social analysis as the low-level front-end analysis with the data secured with the privacy masking. The results of low-level analysis then may feed into pattern learning, dynamic modeling, and association analysis before they are used by dashboard integration, content recommendation, and learning path recommendation.
- The analysis results may be refined by Data Manipulation interface and then stored into Analytics Data Store for further analysis cycle or later processing steps such as the visualization process.

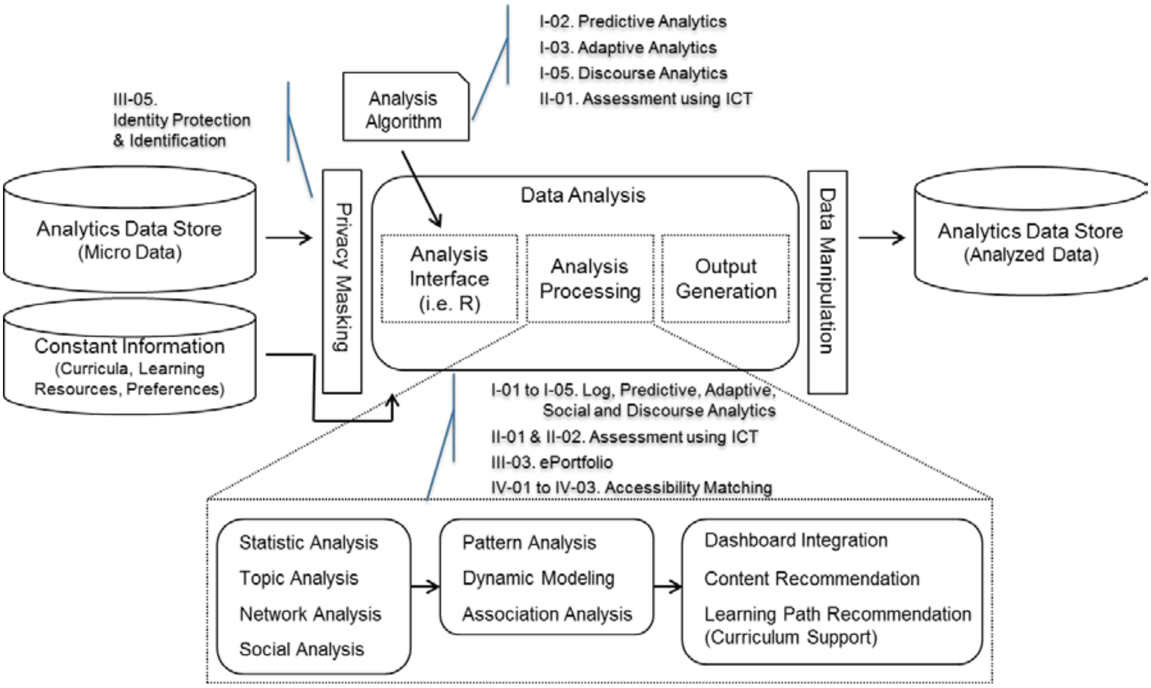

Figure 5: Reference Model of Analyzing for Learning Analytics

3.2.4 Zoom-in diagram for Visualization

Visualization is the process of creating visual representation of abstract data including text and geographic information to allow users to see, explore, interact, and understand large amounts of information in analyzing and reasoning about data and evidence. A primary goal of visualization is to communicate information clearly and efficiently to users via the statistical graphics, plots, information graphics, tables, and charts selected, and thus makes complex data more accessible, understandable and usable as outlined in Figure 6.

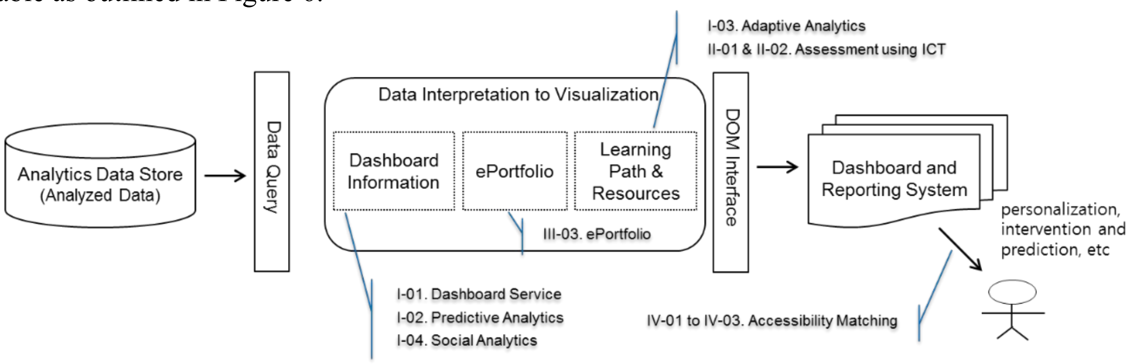

Figure 6: Reference Model of Visualization for Learning Analytics

- The data in Analytics Data Store may be accessed by the visualization process via Data Query interface.
- Visual representation for learning analytics may include dashboard information, ePortfolio, Learning Path & Resources.
- Various external analysis algorithms such predictive analytics, adaptive analytics, discourse analytics, and other assessment using ICT are applied via Analysis Interface.
- The dashboard information may show comparisons or progresses, recommendations, and real-time assessments, topic-based assessment, social-network graph, and so on.
- DOM Interface may provide an open interface to the external Dashboard and Reporting System to realize personalization, intervention and prediction for specific users with additional accessibility requirements as needed.

4. Deployment of the Reference Model

The conceptual reference model described in the previous section serves the purpose of designing the system clearly reflecting identified requirements. In this section, we present an experiment test system to validate the reference model and provide developers with guidelines in implementing the reference model.

4.1 Deployment for Data Collection

In our implementation of the data collection process, learning activity data are generated by the Readium (Readium.org, 2015), a reference system for rendering EPUB 3 publications, as depicted in Figure 7.

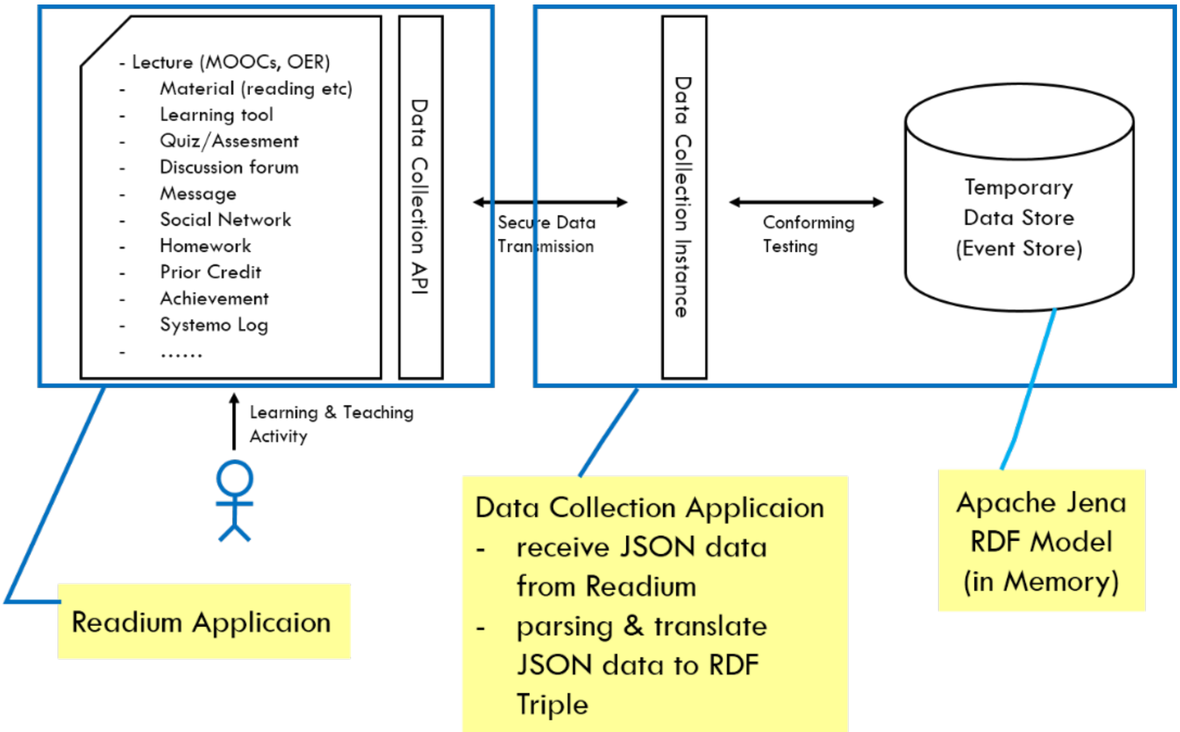

Figure 7: Deployment for Data Collection

Readium in Figure 8 and Figure 9 generates the Caliper (IMS Global Learning Consortium, 2015) event data in JSON format.

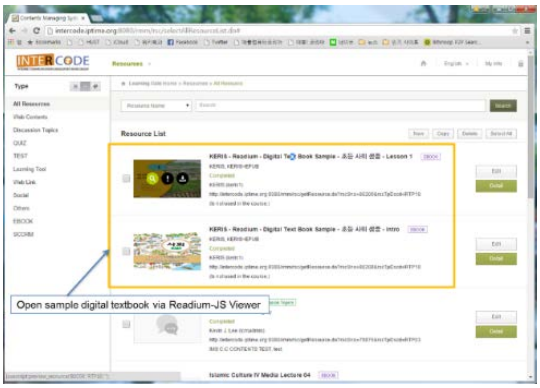

Figure 8: Bookshelf of Digital Textbook

Figure 9: Readium-JS Viewer

4.2 Deployment for Data Storing and Processing

In our implementation, the data stored in the Event Store are accessed via a data query interface using the web protocol. The RDF data are stored into Virtuoso Triple Store (OpenLink Software, 2015) for efficient access by the analysis applications as shown in Figure 10.

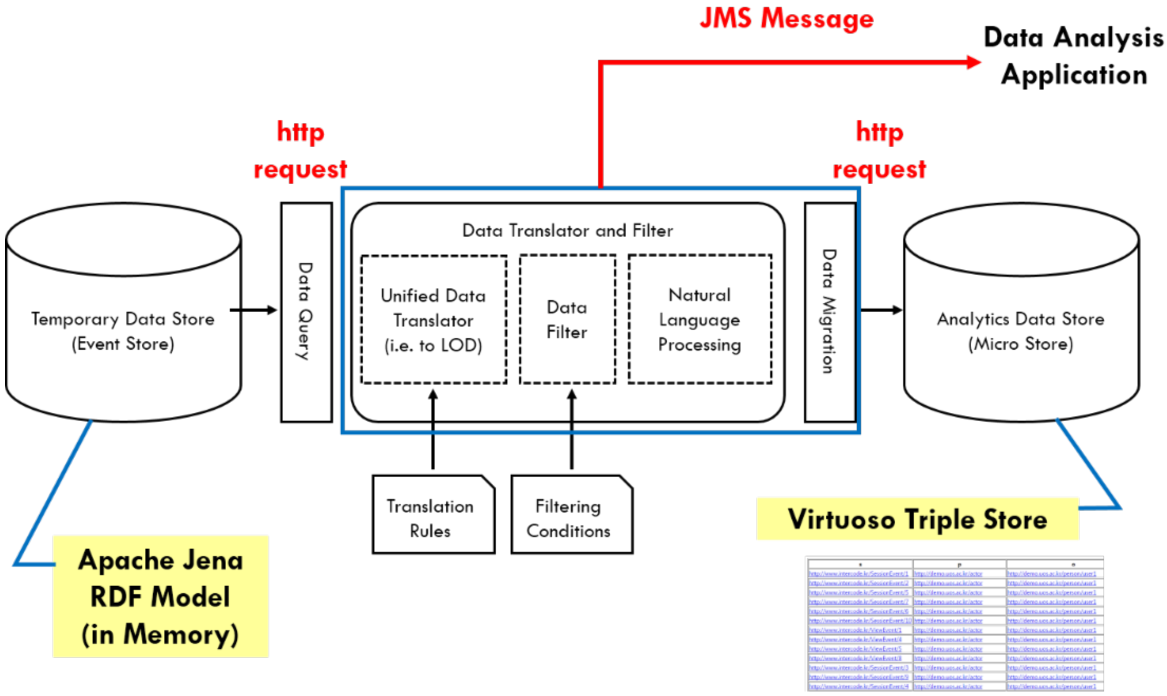

Figure 10: Deployment for Data Storing and Processing

The generated data in JSON format are in turn translated into RDF triples by the Data Collection Interface before they are stored into the Event Store as shown in Figure 11 and Figure 12.

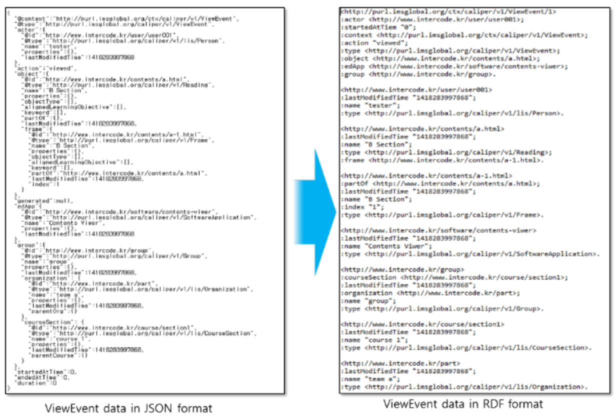

Figure 11: Reading Activity Data Translated to RDF format

Figure 12: Collected Data on Event Store in RDF Triples

4.3 Deployment for Analyzing

In this simple test implementation, we utilized a versatile class-based interface from R program to access the database for analysis as shown in Figure 13. R is a free software environment for statistical computing and graphics (The R Foundation, 2015). In order to achieve a uniform handling of input and output data, the Virtuoso Triple Store is used for the Analytics Data Store along with SPARQL ("SPARQL 1.1 Query Language," 2013)

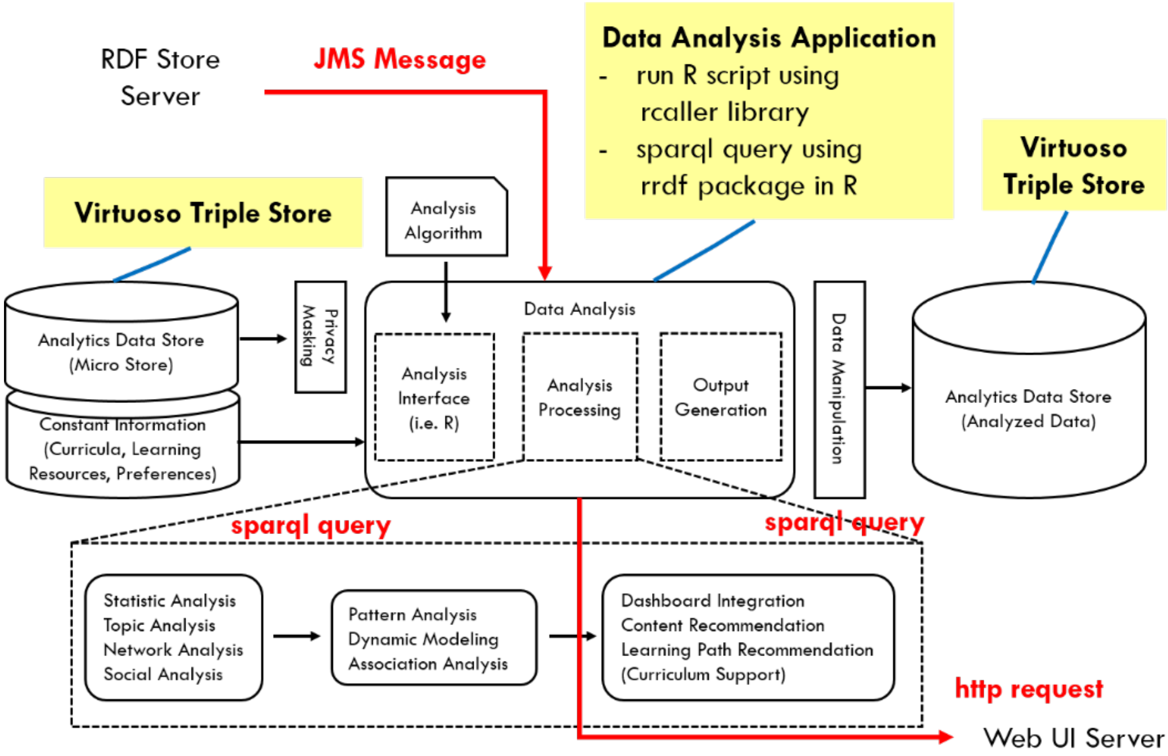

Figure 13: Deployment for Analyzing

4.4 Deployment for Visualization

The process of creating visual representation of the analyzed data stored in the Analytics Data Store is implemented using an asynchronous event driven framework, Node.js (Node.js Foundation, 2015) and JavaScript chart libraries as shown in Figure 14. The Web UI server not only query to pull data from the Analytics Data Store, but also receive updated data as new analytics data are stored to the Analytics Data Store to allow responsive visualization on the dashboard in Figure 15.

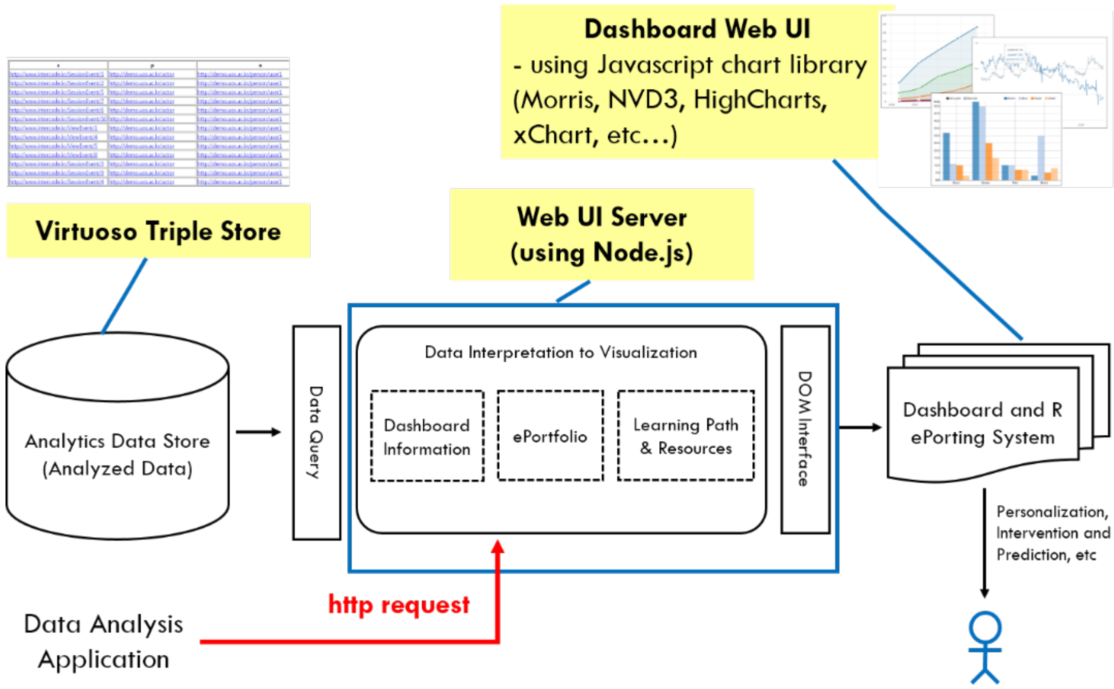

Figure 14: Deployment for Visualization

5. Discussion and Future Work

In this paper, we presented a reference model and an experimental implementation of a learning analytics system based on our preliminary identification of the requirements for the systems to be open, distributed, interoperable, reusable, real-time, usable, and secure. We believe that our approach of designing a reference model and testing with the experimental implementation is laborious but assured tactic to reach the goal of defining an explicit specification of the learning analytics systems as the international standard so that open and extensible learning analytics systems can be built for worldwide interoperability. Accordingly, our next step is to extract explicit data model and interfaces for the possibly distributed components based on our experiments and extensions of the experimental implementation.

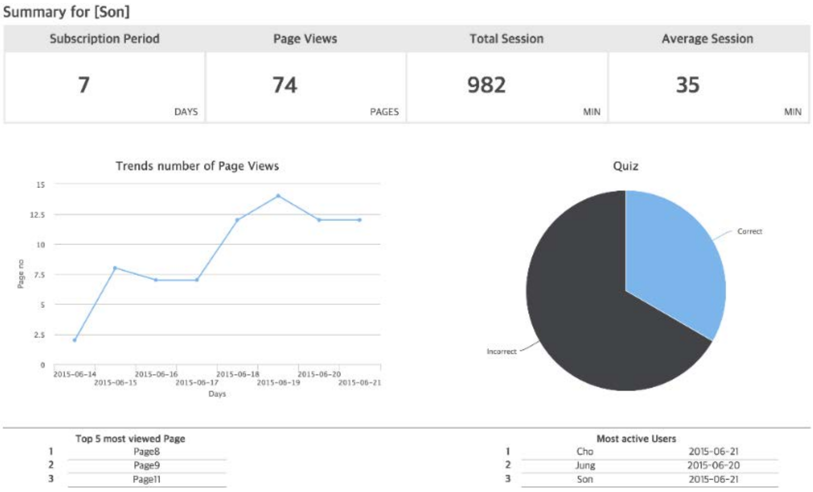

Figure 15: Simple Dashboard for Engagement Profile

Acknowledgements

This research was supported by the ICT Standardization program of MISP (The Ministry of Science, ICT & Future Planning)

References

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning, 4*(5), 318-331.

Choi, B.-g. C., Yong-sang; Lee, Jaeho. (2014). *Preliminary Requirements Analysis towards an Integrated Learning Analytics System.* Paper presented at the The 1st ICCE Workshop on Learning Analytics (LA2014).

IMS Global Learning Consortium. (2015). Caliper Analytics. Retrieved from http://www.imsglobal.org/caliper/

José A. Ruipérez-Valiente, Pedro J. Muñoz-Merino, Derick Leony, & Kloos, C. D. (2014). ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform. *Computers in Human Behavior, 47*, 139-148. doi:10.1016/j.chb.2014.07.002

Node.js Foundation. (2015). Node.js. Retrieved from https://nodejs.org/

OpenLink Software. (2015). Virtuoso Universal Server. Retrieved from http://virtuoso.openlinksw.com/

Readium.org. (2015). Readium Project. Retrieved from http://readium.org/

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., . . . Baker, R. (2011). Open Learning Analytics: an integrated & modularized platform. Retrieved from http://www.solaresearch.org/OpenLearningAnalytics.pdf

SPARQL 1.1 Query Language. (2013). W3C.

The R Foundation. (2015). The R Project for Statistical Computing. Retrieved from https://www.r-project.org/

The Society for Learning Analytics Research. (2015). About SoLAR. Retrieved from http://solaresearch.org/about/