The Concept of Big Data Reference Model

Sungjoon Lim, KoDB*, joon@kodb.or.kr
Dongwon Jeong, KNU**, djeong@kunsan.ac.kr
Jangwon Gim, KISTI***, jangwon@kisti.re.kr
Hanmin Jung, KISTI, jhm@kisti.re.kr

*Korea Database Agency
**Kunsan National University
***Korea Institute of Science and Technology Information
Contents

• Brief History of discussions
• 32N2386 Reference Model for Big Data
• Concept for Big Data Reference model
• Uses of the Big Data reference model
• Possible Structure of parts
• Issues
## Brief History of discussions

<table>
<thead>
<tr>
<th>Issue Date</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov 2011</td>
<td>[SC32N2181] ISO/IEC JTC 1/SC 32 N2181, “Resolutions and topics from the recent JTC 1 meeting of particular interest to SC 32 participants”, SC32 Chair – Jim Melton</td>
</tr>
<tr>
<td></td>
<td>→ SC 6 Telecommunications and information exchange between systems</td>
</tr>
<tr>
<td></td>
<td>→ SC 32 Data management and interchange</td>
</tr>
<tr>
<td></td>
<td>→ SC 39 Sustainability for and by Information Technology</td>
</tr>
<tr>
<td>Jun 2012</td>
<td>[SC32N2241] Ad-hoc on “Next gen analytics” - Keith Hare - Chair</td>
</tr>
<tr>
<td>May 2013</td>
<td>Posted several documents</td>
</tr>
<tr>
<td></td>
<td>- 32N2383-Big_data_&amp;_Next_generation_analytics</td>
</tr>
<tr>
<td></td>
<td>- 32N2386-Reference Model for Big Data</td>
</tr>
<tr>
<td></td>
<td>- 32N2387-thoughts_on_big_data-20130526-farance</td>
</tr>
<tr>
<td></td>
<td>- 32N2388b-report_SG_big_data_analytics</td>
</tr>
<tr>
<td>Nov 2013</td>
<td>SC32 Chairman presented the report of SG to JTC1 Plenary, the reference model is included in the report.(ISO/IEC JTC 1 N11819)</td>
</tr>
</tbody>
</table>
Procedure for developing a reference model for Big Data

1. Eliciting requirements and analyzing the environment of Big Data

2. Establishing visions and strategies for achieving the goal of Big Data

3. Defining a concept model / a reference model / a framework for Big Data

4. Deriving use-cases for applying the Big Data
A lifecycle of Big Data

- Collection/Identification
- Repository/Registry
- Semantic Intellectualization
- Integration

- Analytics / Prediction
- Visualization

- Data Curation
- Data Scientist
- Data Engineer

- Workflow
- Data Quality

32N2386-Reference Model for Big Data
Reference Model for Big Data

* A Reference Model for Big Data

![Diagram of Reference Model for Big Data]

- **Data Layer**
  - Data Collection
  - Data Registry
  - Data Repository

- **Platform Layer**
  - Data Integration
  - Data Semantic Intellectualization

- **Service Layer**
  - Analysis & Prediction
  - Workflow Management
  - Data Quality Management
  - Data Visualization

- **Service Support Layer**
  - Security
  - Data Curation

- **Big Data Management**

32N2386-Reference Model for Big Data
Reference Model for Big Data

A Reference Model for Big Data

- **Data Layer**
  - Data Collection
  - Data Registry
  - Data Repository

- **Platform Layer**
  - Data Semantic Intellectualization
  - Data Integration

- **Service Layer**
  - Analysis & Prediction
  - Workflow Management
  - Data Quality Management
  - Data Visualization

- **Service Support Layer**
  - Data Curation
  - Security

- **Service Management**
Concept for Big Data Reference model

- Reference model might cover whole processes of Big Data from planning to discovering.
- Reference model is not just for harvesting other standards, it might be able to be implemented by user requirement as Big Data services and technologies.

Big Data Reference Model

- Planning
- Preparation
- Analysis
- Visualization

User Requirement
- Operation and Management
- Preprocessing
- Analysis
- Visualization

Big Data Service & Technology
discovery
Concept for Big Data Reference model

Planning Reference Model

Operation and Management

Preprocessing → Analysis → Visualization

Management Reference Model

Planning

Technical Reference Model
Uses of the Big Data reference model

- To enhance Big Data industry, Big Data reference model can be a guidance for Big Data consultation services and development solutions

Stakeholder’s requirement
(data provider, data creator, data analyzer, Service provider, service user, and so on)
Possible Structure of parts

- Part 1 Overview
- Part 2 Planning Reference Model
- Part 3 Management Reference Model
- Part 4 Technical Reference Model
Issues

• How can WG2 contribute to Big Data Standardization?
• ...