MediaNET LAB
Media Network Laboratory

Open Data Platform and FG-DPM Activities

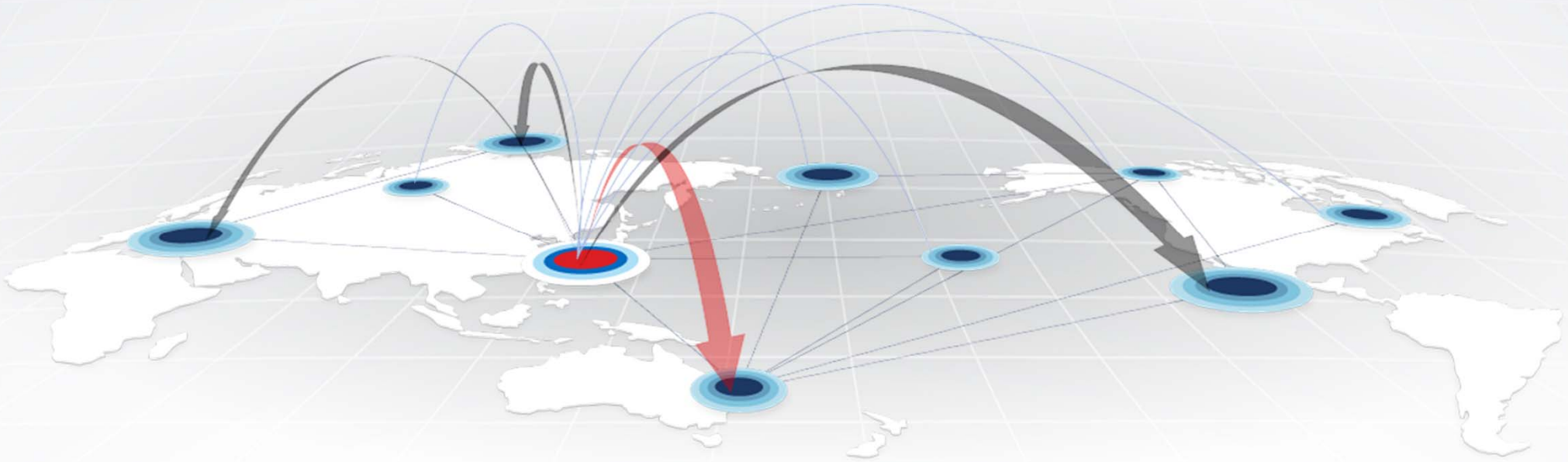
Jun Kyun Choi

jkchoi59@kaist.edu

한국과학기술원

Contents

Toward Open Data Platform



What is the Language of Cyber World ?

Cyber Language	Physical Language	Natural Language
Language for Semantic Knowledge	Human Language	Human Language
Language for Application Platform	Language for Building, Road, Station, Airport, Hospital	Language for Monkey
Language for System and Application	Language for TV, Car, and Airplane	Language for Dolphin, Dog, Pig, Elephant
Operating System (OS)	Language for Hardware and software	Language for Mouse, Chicken, Bird
Machine Language (e.g., device driver)	Language for Components and Sensors	Language for Bee, Ant, Insect

Paradox between Open Data and Data Security - 1

- **(Trust) "Data is safe or trustful" and/or "Data is correct or secure"**
  - Data **Credibility** and **Trustiness** is important
    - Fake news, Clickbait, Yellow journalism, Tabloid journalism, Pseudoscience
  - Who believe Raw Data?
    - Raw image and video can be modified
    - Perceptions of eye and ear are not correct.
  - Modified or enhanced for **better understanding**
    - Only if original data is correct
    - But, may be manipulated intentionally

(note) "The New Digital Age: Reshaping the Future of People, Nations and Business," Google's Eric Schmidt

Paradox between Open Data and Data Security - 2

❒ (Open) Business Opportunities and Threats for Open Data

- **Open data is quite limited**
  - Only for education, health, science, etc. (publically useful data)
    - Business value of public open data is marginal
  - Private data for business is not relevant to open
- **Can create new values only when data is open**
  - Filter and extract valuable portions of knowledge from other related data
    - Data from Nano-technology, bio- and medical technology, and material technology should be open in order to make a process for obtaining additional useful knowledge
- **All the data formats need their own software to make a process!**
  - How to make a process for thousands types of data? → data-driven approaches
    - Software is just a kit to understand and interpret data
  - Who pay for translation and interpretation costs if data is not directly open!
    - Additional conversion and processing costs should be minimized

Paradox between Open Data and Data Security - 3

- **(Dilemma) Dilemma between Data Privacy and Big Data**
  - If data for my **personal life** is open,
    - Nobody allows to make a process from his/her personal data
  - But, if additional data processing is not allowed,
    - Right decision for medical treatment is difficult without **personal health records**
  - Big data processing of accumulated health records brings additional values!
- **(Harmony) Harmony between data trust (business value) and data security (cost of technology)**
  - Applicability of open data is tightly relating to the **trust level** of data
  - Acceptable security mechanism is needed according to business application?
    - Security algorithm SHOULD NOT be **perfect**
      - Perfect security algorithm is more dangerous and difficult to handle
    - Security algorithm on tiny sensor devices is difficult to operate and manage
      - Simple and acceptable solutions **both at CAPEX and OPEX** costs are needed.

Paradox between open data and data business

- **Visit ratio of web sites, documents, and TV**
  - Important to evaluate their media impacts
  - Difficult to find valuable information through web sites
  - May not recognize new releases of software and documents
- **How to distribute the contents only for the registered customers confidentially**
  - Payable contents are encrypted with security code to avoid illegal distribution
  - Tracking or certification mechanism are needed
- **Paradox on data sharing and distribution**
  - Public available documents may be used for business
    - Some people add on the values over original open data with security protection
  - Securely protected documents may be open to public
    - Privacy problems may happen when private data are unacceptably open to public
    - Without advertisements, customers feel some difficulty to find the private documents

Strategies for Open Data Standards

- **How to share the data for knowledge accumulation**
  - include metadata like authors, date, genre, and short summaries for easier search and discovery
  - share individual experiences and opinions
- **Metadata and data schema are the key essences for future data eco-society**
  - provide useful information about when data is created, delivered, processed, shared, and consumed by users
  - better understand the actual content of the data
- **New forms of data for development, acquisition, and spread of knowledge**
  - create, collect, accumulate, share, and distribute data with the progress of user interface and human perception

Key Strategies of Data Model - 1

- **Similar to Natural Language**
  - No rule, No grammar, just talk and use the word when people thinks !
  - But, human-understandable and machine-readable logical forms
  - Minimize the preliminary and/or background knowledge
  - Minimize data exchange or summary if both sides have background information
  - Audio/visual signals are also tools to deliver information and knowledge
    - But, decoding software and perception algorithm is prerequisite
- **Data Serialization Model**
  - Used for data delivery via communication network
  - In-band or out-of-band signaling protocol may be needed
  - Most existing protocols should be designed with fixed header format
  - Dynamic header with semantic metadata information can be designed → parsed by input processors
    - Sorting, alignment, classification, and filtering, etc.

Key Strategies of Data Model - 2

- **Entity-Attribute-Value (EAV) Data model**
  - **Well-known data type or model** should be defined
    - Binary, String, boolean, date, day, month, etc.
    - ID, phone, email, SNS, name, repository, product code, on-line sheet, address, etc.
    - DTD, JSON-LD, Microdata, etc.
  - Data model for search, classification and database
    - Catalog, repository, Big file, table, structure/unstructured file, etc.
    - Hierarchical model or hash algorithm
  - Extension of dynamic or active data model
    - Calculate algorithm, external I/O, statistics, query, events and triggering conditions according to situation and contexts,
    - Change or update microdata or Javascript
  - Future flexibility to invite new data model and format
    - Ready for and abstract/clustering of virtual cyber space

Key Strategies of Data Model - 3

- **Data Model according to**
  - Data creation and collection (device type, protocol, etc.)
  - Data Storage and Processing (database, R, Hadoop, Excel)
  - Data Delivery (protocol)
  - Data layout (web, speaker, screen, touch, peripheral, etc.)
- **How to avoid dependency of platform, OS, and device**
  - Minimize the commonality → minimum interoperability
    - GPU, famous OS, and cloud platform as a common ?
  - Review the existing and on-going data models for their own applications
    - First, at A/V and telecommunication application as well as on-line media
    - Secondly, Geolocation, Game, Transport, Energy, Health
  - Identify the benefits of web-based data model for applications
    - XML/RDF, and HTML

Key Strategies of Data Model - 4

- **Metadata model**
  - Depending on usage and business application → at least keep minimum interoperability
    - Same data contain different metadata for delivery, processing, search, visualization
  - Relationship between Data and Metadata
    - Same format, but, metadata is contextual data that describe data or its characteristics
- **Data model for authentication and authorization data model**
  - Data encryption and dynamic access control by using login procedure, converting table (i.e., hash table) and priority, etc.
- **Web-based data model**
  - For rendering and screen display
    - But, minimize transformation overheads of XML/RDF data to HTML
    - Limitation of HTML parser → Extension of CSS and Javascript
    - Synchronize multiple screen at separate location!
  - Web UI/UX for touch screen and peripherals including mouse and joystick

Technical Issues of New Data Format

- **HTML5-based UI/UX to help human perception**
  - Device API, RESTful interface, Javascript, etc.
- **Data and Metadata Together**
  - Media object and media resource model
  - Media Ontology, Media Annotation
  - File Format, Microformat, ATOM/RSS
  - Device Data Format for Mashup
- **Data format for Web-based open Platform**
  - Open, Auto-configurable, and future flexible
  - But, Securable and manageable is in question ?

Web-based IoT Data Format

❒ XML/RDF Schema

- Well-known data format ?
  - JSON, Microformat (e.g., vCard, hCalendar), ATOM/RSS
- DTD syntax, schema, and semantics, etc.
  - Tag, Index, Summary, Thumbnail, Preview, etc.

❒ IoT Data Format for future life and business

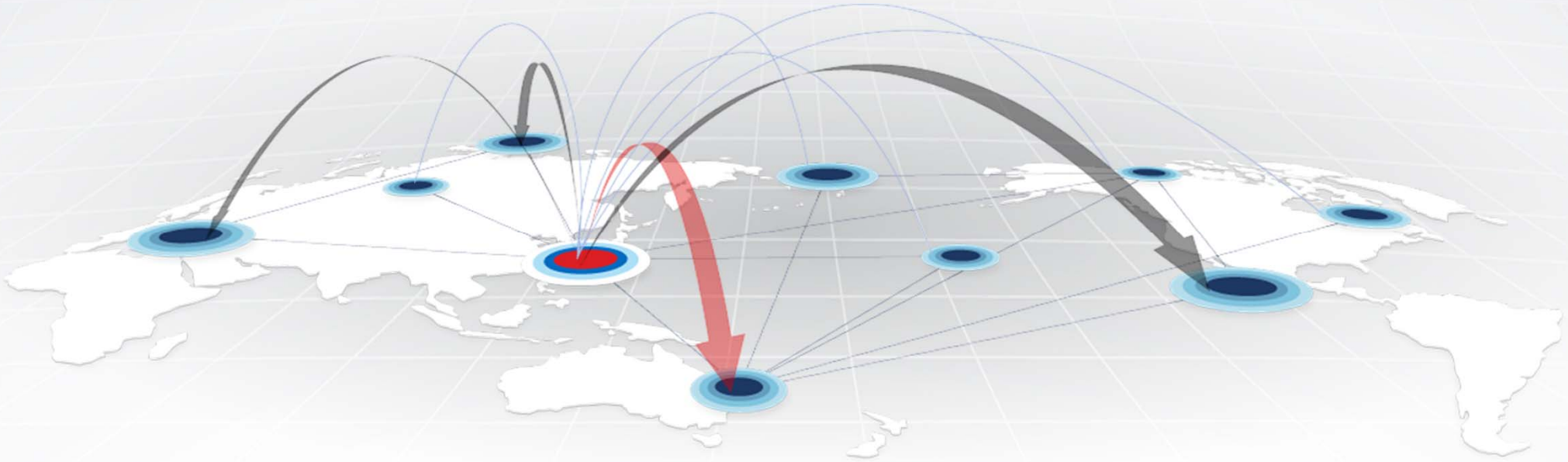
- Continue to URI/URL/URN-based ?
- Microservice, microdata for semantics
- Mashup format for heterogeneous data applications

KAIST 한국과학기술원 Korea Advanced Institute of Science and Technology
jkchoi59@kaist.ac.kr

IoT Data Standardizations for Smart City

- IoT data cloud system for Smart City
- Real-time data analytics for Smart City
- IoT metadata for Smart City
  - Syntax, semantics, context-aware for unstructured IoT data
- Data security and trust for Smart City
- Data applications for Smart City (including visualization)
- Interoperability and Deployment scenarios for IoT-based Smart City

FG-DPM Activities



Terms of Reference of FG-DPM

- To study, review and survey existing technologies, platforms, guidelines and standards for data processing and management including data format in support of the mandate of SG20;
- To recognize and highlight the various perspectives for the future of data driven eco-environments;
- To promote security and trust within data management frameworks;
- To identify and study data protection techniques;
- To facilitate cross-cutting data interoperability issues and point the way to the development of efficient and scalable approaches to managing systems data;
- To study meta-data;
- To study trust in data management frameworks including digital identification and certification;
- To investigate the role of emerging technologies to support data management and emerging trends including blockchain;
- To identify challenges in the standardization activities for data processing and management.
- To establish liaisons and relationships with other organizations which could contribute to the standardization activities for data processing and management.

Focus Group Structure

- WG1 - Use Cases, Requirements and Applications/Services;
- WG2 - DPM Framework, Architectures and Core Components;
- WG3 - Data sharing, Interoperability and Blockchain,
- WG4 - Security, Privacy and Trust including Governance
- WG5 – Data Economy, commercialization and monetization

Work on Deliverables

D0.1: DPM Terms and Definitions, Taxonomies
D1.1: Use Cases Analysis and General Requirements for DPM
D1.2: IoT and SC&C Applications and Services using DPM

D2.1: DPM Framework for Data-driven IoT and SC&C
D2.2: DPM Functional Architectures
D2.3: Data Modeling and Formats Specification for DPM
D2.4: Technical Report on Data Analytics
D2.5: Technical Report on Data Visualization
D2.6: Framework of Meta Data Management
D2.7: Framework of Data Life Cycle Management

D3.1: Framework of Open/Private Data
D3.2: Technical Enablers for Open Data Platform
D3.3: Technical Report on Data Interoperability
D3.4: Technical Report on Data sets Interoperability
D3.5: Overview of IoT and Blockchain
D3.6: Bockchain-based Data Exchange and Sharing Technology
D3.7: Using blockchain to improve data management

D4.1: Framework of Security and Privacy in DPM
D4.2: Privacy Management in Smart Cities
D4.3: Technical Enablers for Trusted Data
D4.4: Data Quality Management for Trusted Data
D4.5: Data Governance Framework for IoT and SC&C

D5.1: Modeling of Data Economy for value creation and Stakeholders identification
D5.2: Business models, commercialization and monetization to support data economy
D5.3: Data economy impact assessment, policy and sustainability implications
D5.4: Data economy regulatory framework
D0.2: DPM Standards Roadmap

Output Documents of 2nd meeting

❒ Location: Geneva, 20 - 25 October 2017

Deliverable	Title
D1.1	Draft Technical Report on “Use Cases Analysis and General Requirements for DPM”
D2.1	Draft Technical Report on “Data Processing and Management Framework for Data-driven IoT and Smart Cities and Communities”
D2.3.1	Draft Technical Report on “Data format in IoT and smart city”
D2.3.2	Draft Technical Report on Web based Microdata formats for IoT and Smart city
D2.3.3	Draft Technical Report on “Metadata format in IoT and smart city”
D3.2	Draft Technical Specifications on “SensorThings API – Sensing, a cross-domain IoT data model and RESTful API”
D3.3	Draft Technical Report on “Framework to support data interoperability in IoT environments”
D3.5	Draft Technical Report on TR.IoT-BC-overview “Overview of IoT and Blockchain”
D3.6	Draft Technical Report on “Bockchain-based Data Exchange and Sharing Technology”
D3.7	Draft Technical Report on TR.IoT-BC-DM “Blockchain Based Data Management”
D4.1	Draft Technical Report on “Framework of Security and Privacy in DPM”
D4.2	Draft Technical Report on “Privacy Management in Smart Cities”
D5.3	Draft Technical Report on “Data Economy, Commercialization, and Monetization”

Next FG-DPM Meeting

- **Brussels, Belgium from 19 to 23 February 2018 including a one-day Workshop**
  - Hosted by Open & Agile Smart Cities (OASC)


Thank you!