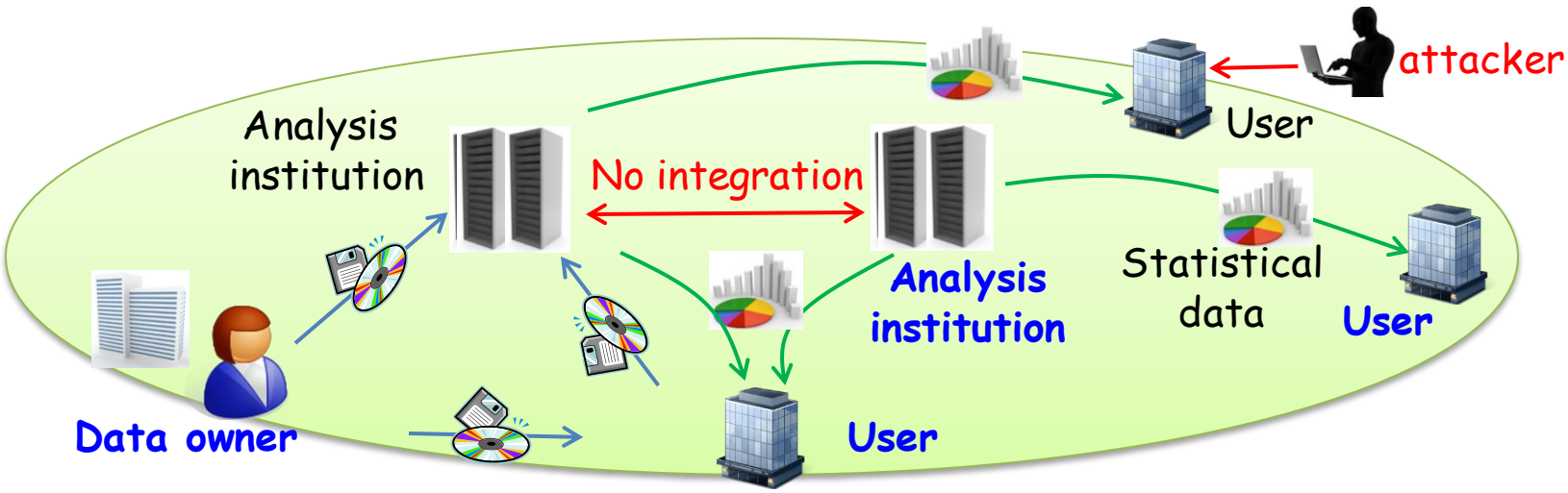
# The Secure and Fair Integration for any entity of Data owner, analysis, & user in Big and reconstructive Data (Surf Bird)

Miyaji, Omote, Futa, Su (JAIST),
Miyake, Kiyomoto, Nakamura (KDDI R&D Labs),
Nishida (AIST), Yamamoto, Tanaka (U. Tokyo)

November 5th, 2014

# Background

- **Social and Economical Problems surrounding big data**
  - Reasonable benefit for data owner to provide their data
  - How to pay for value (P4V) of data with valance of anonymity and risk
- **Too many attacks and errors**
  - Outsider attack: once attacked all data compromised
  - Insider attack: human error, leakage
- **Secure and dynamic integration**
  - Multiple analysis institutions use each privacy policy
    Anonymity level cannot be determined uniquely for any data



# Overview (purpose, target, ripple effect)

- **Purposes**
  - Protect data owners' benefits, establish a secure big-data circulation platform among owner, analysis institute and user.
  - 2 test beds of living safety and medical information.
- **Research target**
  - Attack and error tolerance: robust system for outsider and insider attacks by combining several protections
  - Secure and fair P4V: pay for value (P4V), traceability, risk evaluation
  - Secure dynamic data integration and collection
- **Ripple effect**
  - Establish the big data circulation platform among owners, analysis institution and user under a win-win paradigm circle
  - Establish secure secondary circulation of analysis results of Big Data

# Building Blocks: realize the big-data circulation platform

**Topic 1. Secure big data management**

- (S1) Proof of Retrievability (POR)/Proof of Erasability (POE)
- (S3) privacy policy manager (PPM)
- (S5) risk evaluation for anonymization

**Topic 2. Secure analysis and feedback of big data**

- (S4) privacy set intersection (PSI)
- (S6) secure traceability
- (S7) secure and fair P4V

**Topic 3. Attack and error tolerance**

- (S8) Protection against attacks on TLS protocol
- (S2) leakage detect mechanism

**Test Bed 1. Big data circulation platform for Living safety**

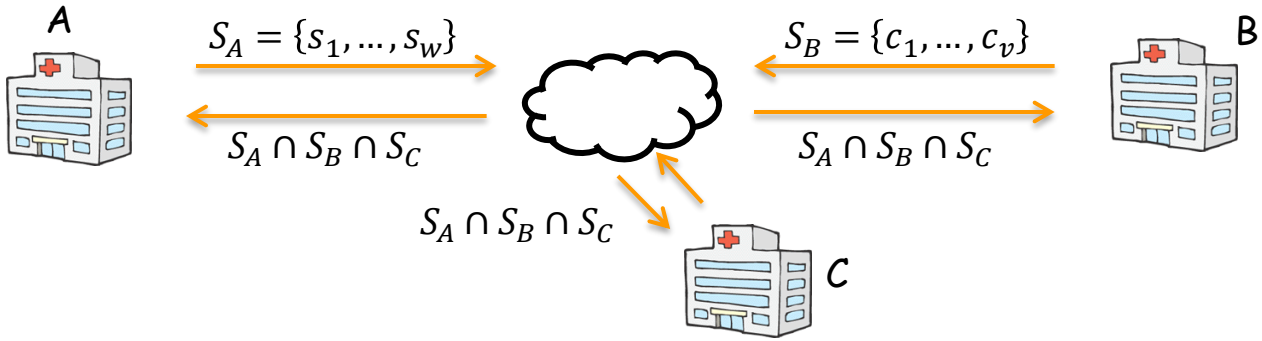
- (T1-1) multi-organization data modeling with dynamic privacy policy

**Test Bed 2. Big data distribution platform for medical info**

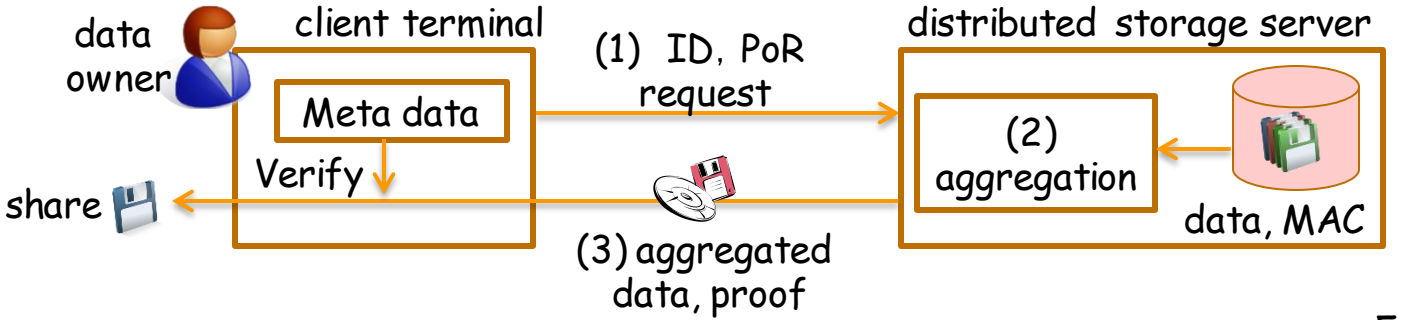
- (T2-1) ID linkage techniques
- (T2-2) empirical medical data integration and utilization

# Secure Data Analysis & Management

- **PSI(Private Set Intersection)**
  - Extract information, while protecting each entities' privacy
  - Not achieved yet to extract information of multiple entities

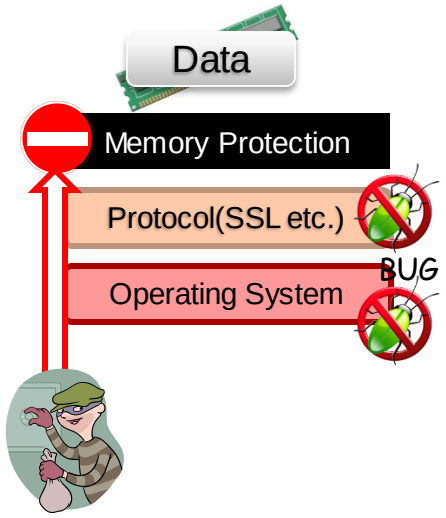


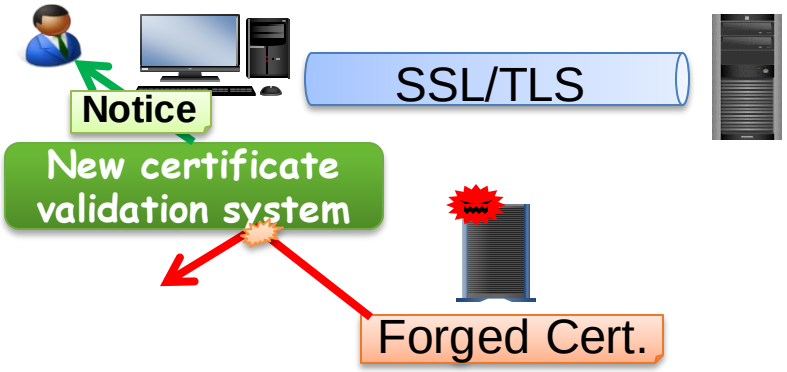

- **POR(Proof of Retrievability)**
  - Check the integrity of big data with robustness
  - Repair of corrupted data on a server while still decoding.



# Attack and Error Tolerance

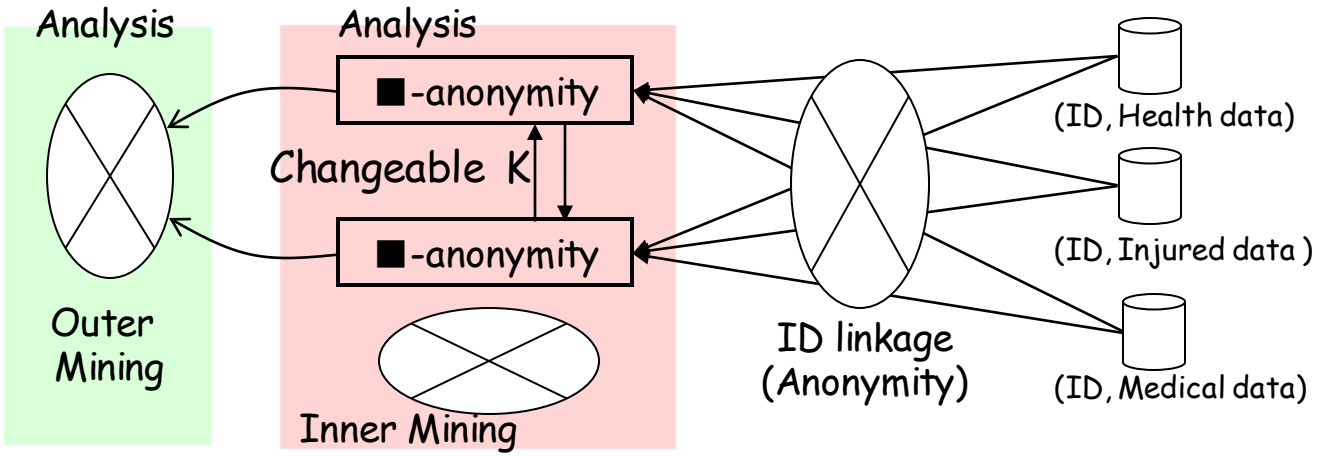
- **Protection of Data on Physical Memory**
  - Data leakage prevention on a vulnerable environments (Bug Free Concept)
  - Practical solution (Performance and Security)
- **Leakage Detect mechanism**
  - Embed watermarks using k-Anonymization algorithms
- **Protection against attacks on TLS protocol**
  - Countermeasures against forged certificate issued by corrupted Certificate Authority
    - Eliminating the forgery risk by validation of certificate format
  - Security analysis of encryption schemes in TLS



# Test-bed related technologies

- **Data integration and ID-linkage in multi-organization**
  - Prevent the accidental privacy violations
- **Dynamic privacy level adjustment fit for the data size**
  - Dynamic privacy level fit for mining after anonymization
- **Application to Home care**
  - Secure data collect various sensed data for home care
  - Secure storage and analysis and utilization for sensor data



# Approach to Global Tendency of Privacy/Security for Big Data

## Global Issue

There are several research projects of Privacy Information Management.
→No yet: secure Big Data infrastructures with PDP, traceability, P4V, and dynamic integration of analysis and management.

## Standardization activities ← Feedback our research results

ISO/IEC JTC1/SC27/WG5
- Just started, not enough discussion: Privacy/Personal information management system, Privacy Impact Assessment, Privacy Architecture Framework

ISO TC215 (Health informatics)/WG4
- Long discussion, except privacy ← few usable privacy-preserving technologies: Privacy protection on personal health information

## Research activities ← Feedback our research results

ABC4Trust (Attribute-based Credentials for Trust) (EU FP7)
- Platform for information exchange with privacy-preserving
- Neither integrated infrastructure nor enough experiment with real-world Big Data

→ New research project on privacy of big data is planned for EU Horizon 2020.

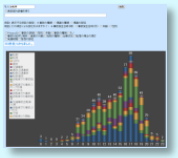
# Toward Two Test Beds using Real data

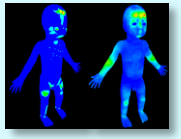
- Utilize securely integrated data of various organizations
- Collect securely data of various organizations while matching

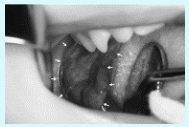
secure traceability

secure and fair P4V

Data
- living safety
- medical info

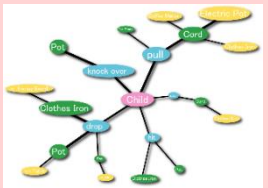




data owner

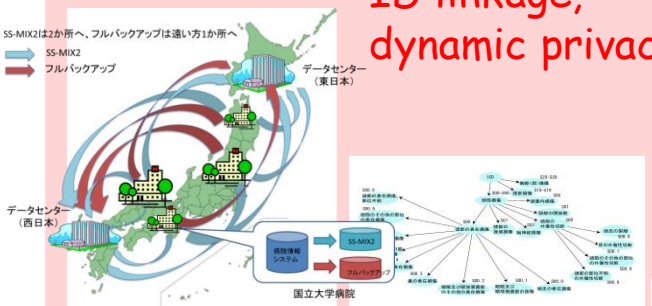
Collection

secure data management

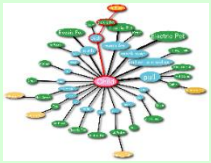



dynamic, integration

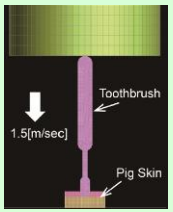
ID linkage, dynamic privacy policy




Modeling



attack and error tolerance


analysis institution

user

# Goals until March, 2017

manufacturers

AIST (living safety)

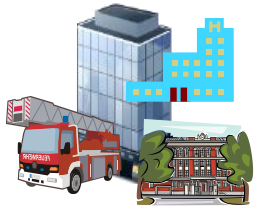
Develop & evaluate utilization of database: ontology (ICD, ICF, DenverII), dimension operating, and probabilistic modeling

Ontology
R static

U. Tokyo (medical information)

De-identified by minimal random swapping
Still preserve basic static values
Establish models for home care

database



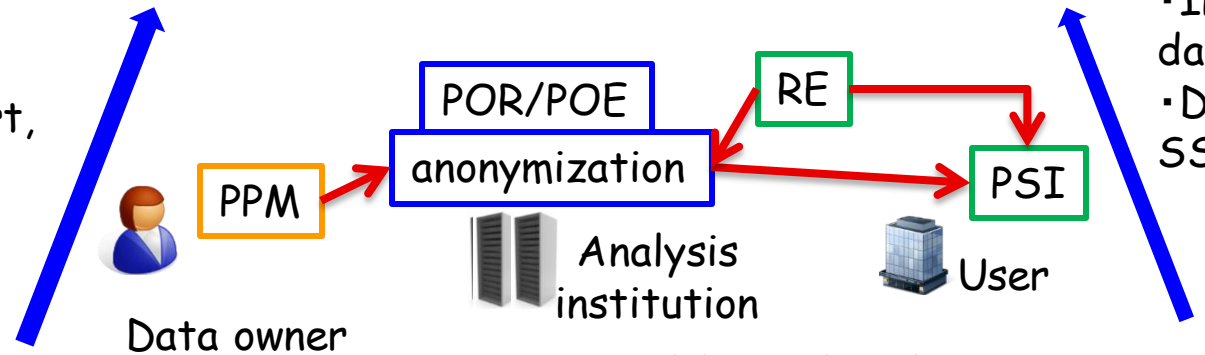
- hospital-based injury surveillance,
- first-aid transport,
- school-based insurance

integrated security

POR/POE/PSI+PPM, PSI/PoR/POE +RE

database

- Insurance claim data,
- DPC related data, SS-MIX2




JAIST

- Secure-data management : POR/POE
- Attack-tolerant method: secure SSL
- Secure Analysis: PSI

KDDI R&D Laboratories

- Secure-data management : semi-automatic PPM/ Risk Evaluation (RE)
- Attack-tolerant method: Leakage detect mechanism, secure memory

# Schedule and relations research results

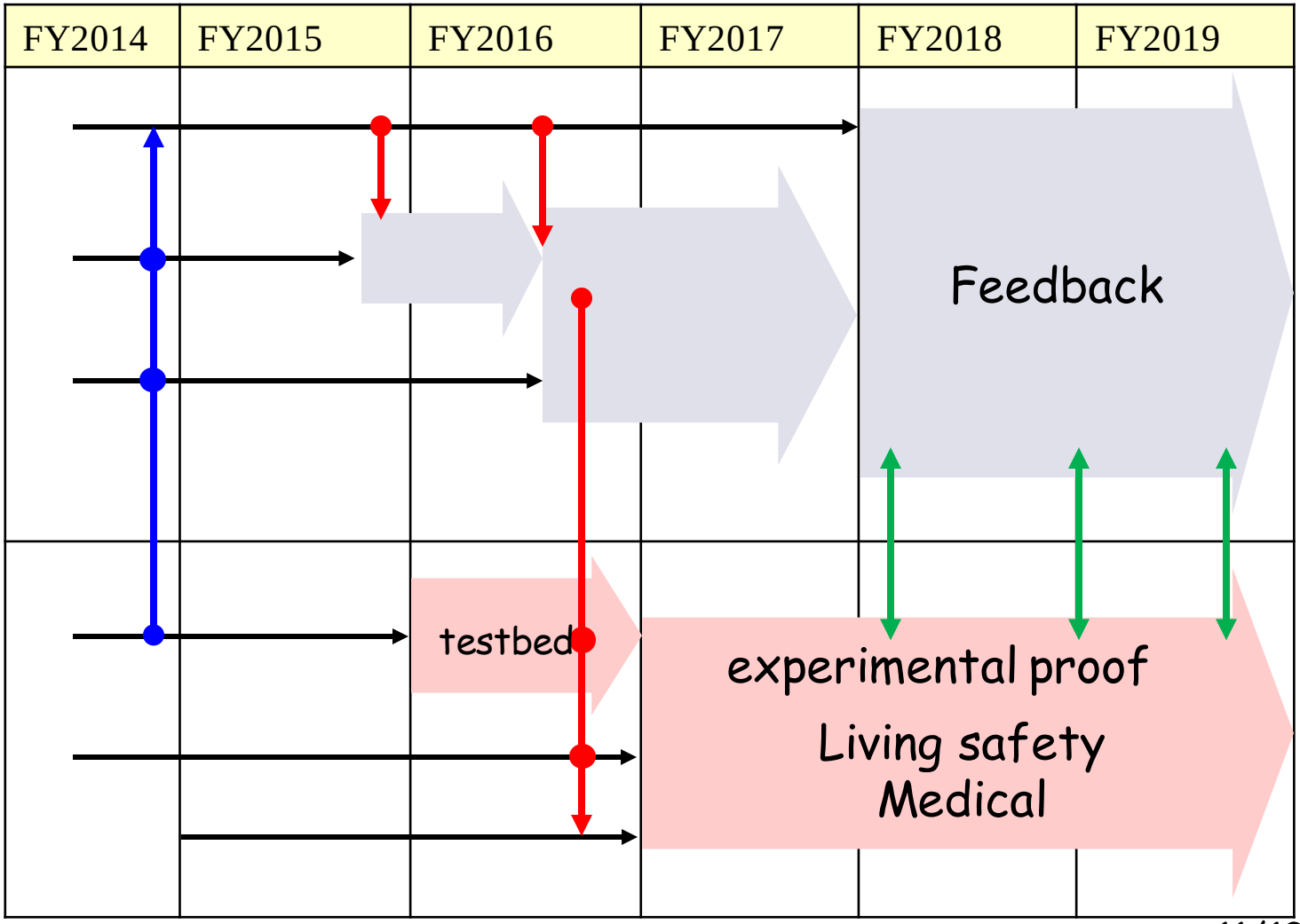
(JAIST/KDDI)

- Attack-tolerant methods
- Secure data management
- Secure analysis & Feedback

(AIST/U. Tokyo)

- Database Specifications
- Dynamic data modeling

ID linkage



# Future prospects of our research

1. **Establish technologies for a secure big-data circulation platform**
   - Attack and error tolerance
   - Secure and fare P4V: traceability, risk evaluation
   - Secure dynamic data integration

2. **The first big-data circulation platform experiments for living safety and medical information**

3. **International standardization and leadership of the next generation in the information industry**

4. **Social Impact**

   Solution Economy: solve problems by cooperation of industries.

# Question and Comment