

# **Very Large Information Integration and Application Platform**

CSTP Coordination Program of Science and Technology Projects

Project Director: Shojiro Nishio, Ph. D. (Osaka University, Japan)

September, 2008

# Agenda

- 1. Background and Purpose**
- 2. Very Large Information Integration and Application Platform**
- 3. Four Major Projects of Ministries (METI, MIC, MEXT)**
- 4. Collaboration among Four Projects**
- 5. Conclusions and Future Work**

METI: Ministry of Economy, Trade and Industry

MIC: Ministry of Internal Affairs and Communications

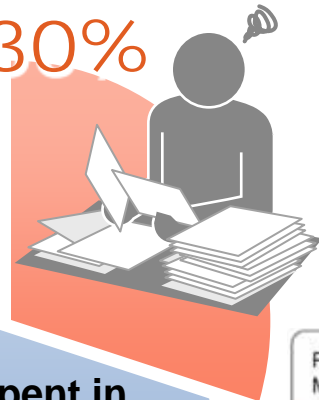
MEXT: Ministry of Education, Culture, Sports, Science and Technology

# 1. Background and Purpose

# Background-1: Information Explosion Era

Time spent in **searching**

30%



Time spent in **intellectual activities**  
(high-value-added industry worker)



Plenty of information  
MSN's own research found **50% of complex queries go unanswered**, while even more simple searches take an average of 11 minutes. Despite this it found that there was real hunger for web-based answers. **Three-quarters of people would rather use the web to answer their questions than their own family members . . .**

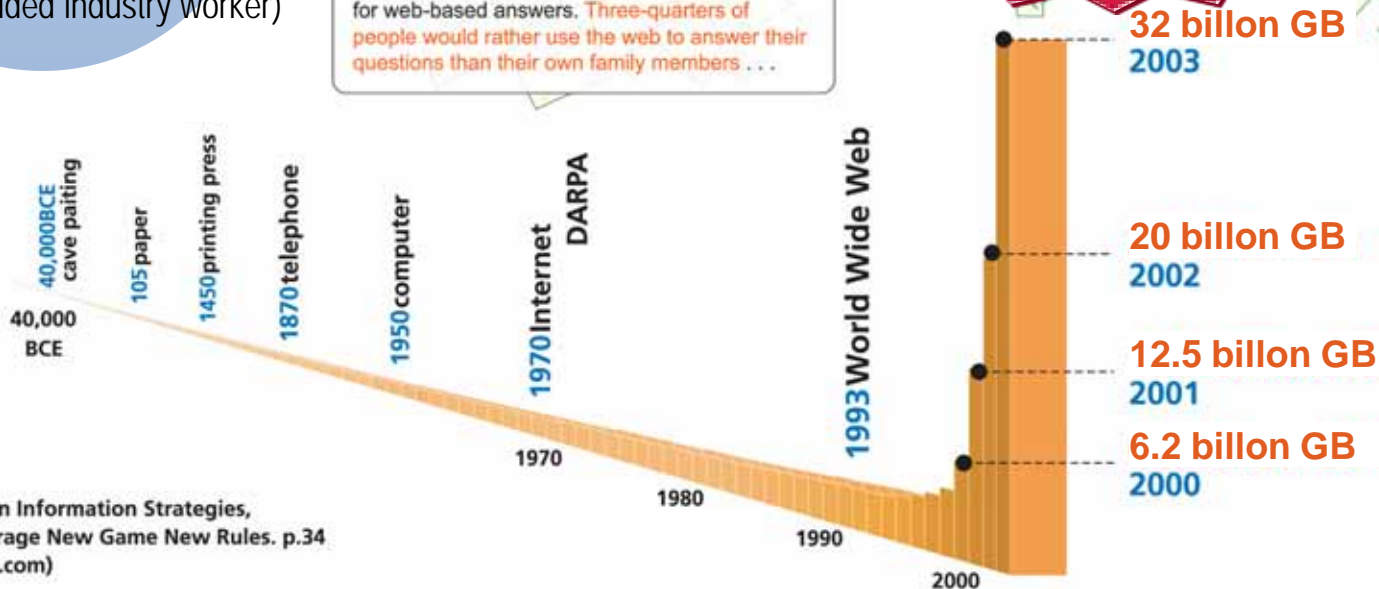
161 billion GB (2006)

988 billion GB (2010)

IDC, "Expanding Digital Universe" (Mar., 2007)

- n Growth of Web
- n Rich Information
- n Emerging sensor technologies

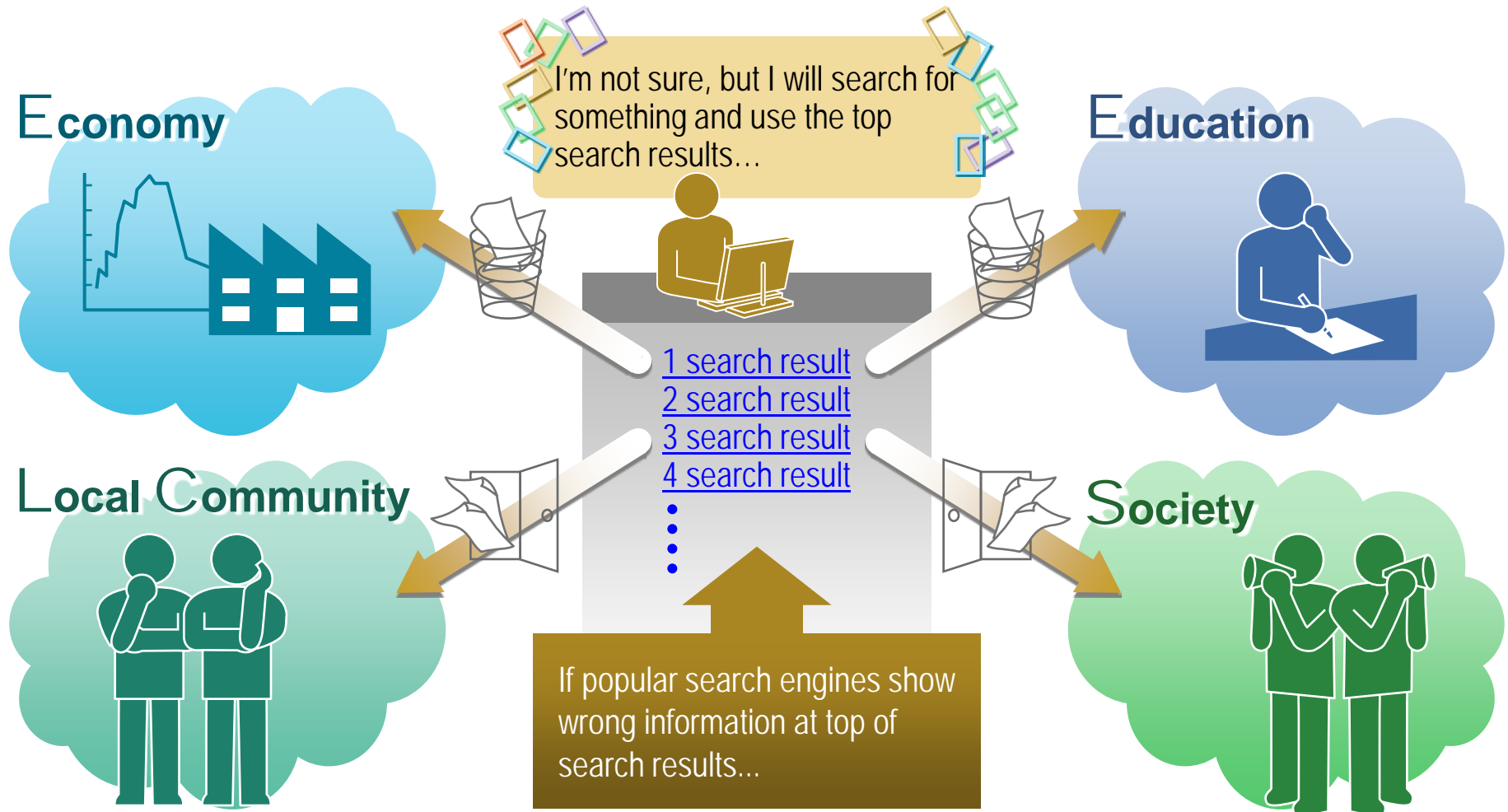
**Information Explosion !!**



Source: Horison Information Strategies, cited from Storage New Game New Rules. p.34 (www.horison.com)

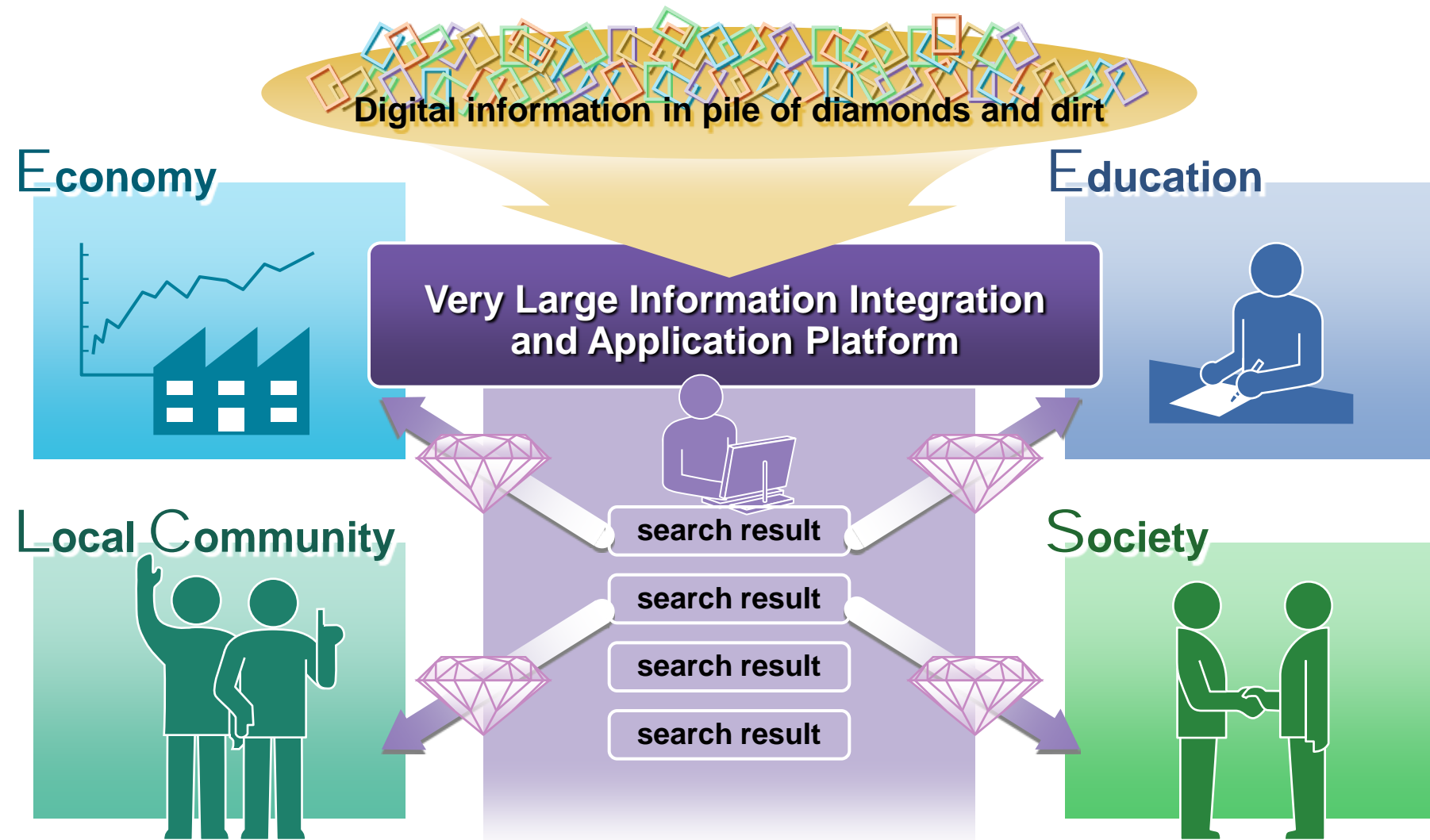
## Background-2: Problems with Information Retrieval Engines

- Search engines do not analyze and present information from user's perspective
- Garbage and harmful information may be presented at top of search results



# Purpose: New Intelligent Information Retrieval Platform

n Achieving a safe society by using an intelligent platform



## **2. Very Large Information Integration and Application Platform**

2.1 Promotion Scheme

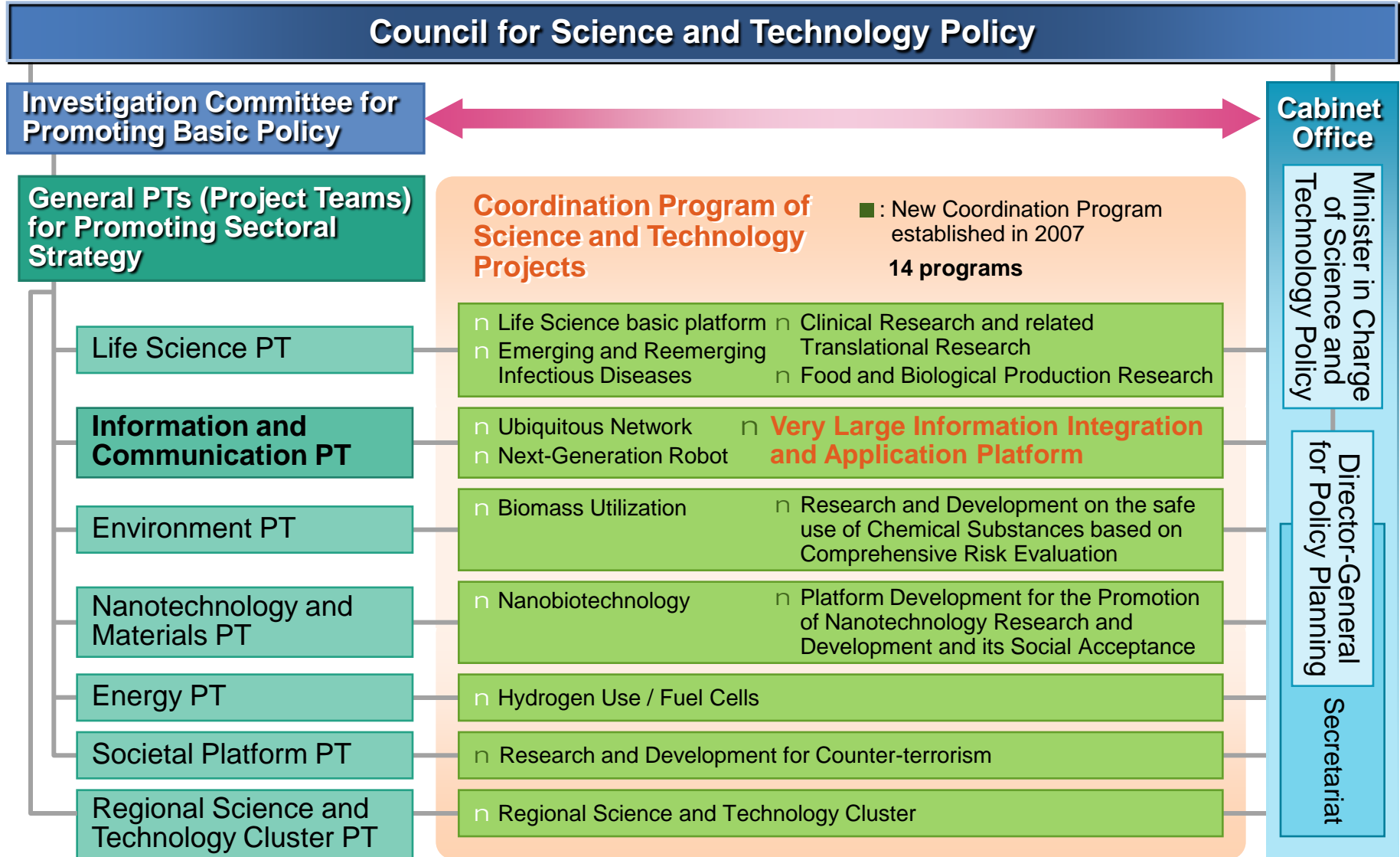
2.2 Overview

2.3 Positioning

2.4 Goal

2.5 Schedule

# 2.1 Promotion Scheme -1 : Coordination Program of Science and Technology Projects





## 2.1 Promotion Scheme -2:

### What is Coordination Program of Science and Technology Projects?

#### Coordination Program of Science and Technology Projects

A framework mandated to

- n Maximize results from projects of governmental organizations
- n Strengthen projects by assigning coordinators to lead them
- n Make needed adjustments, e.g. eliminating project overlaps
- n Utilizing Grants in Aid of Special Coordination Funds for Promoting Science and Technology.

## 2.2 Overview: New Coordination Program in 2007

### Very Large Information Integration and Application Platform

#### Overview

Develop **platform technology for intelligent, next-generation information use**

- n Safe, appropriate, and secure information systems
  - n Collection, analysis, and management of various information
- ➔ Providing unique information services

#### Four Targeted Projects

##### MEXT

R&D of Ultra-High  
Performance  
Database Engine

##### METI

Information Grand  
Voyage Project

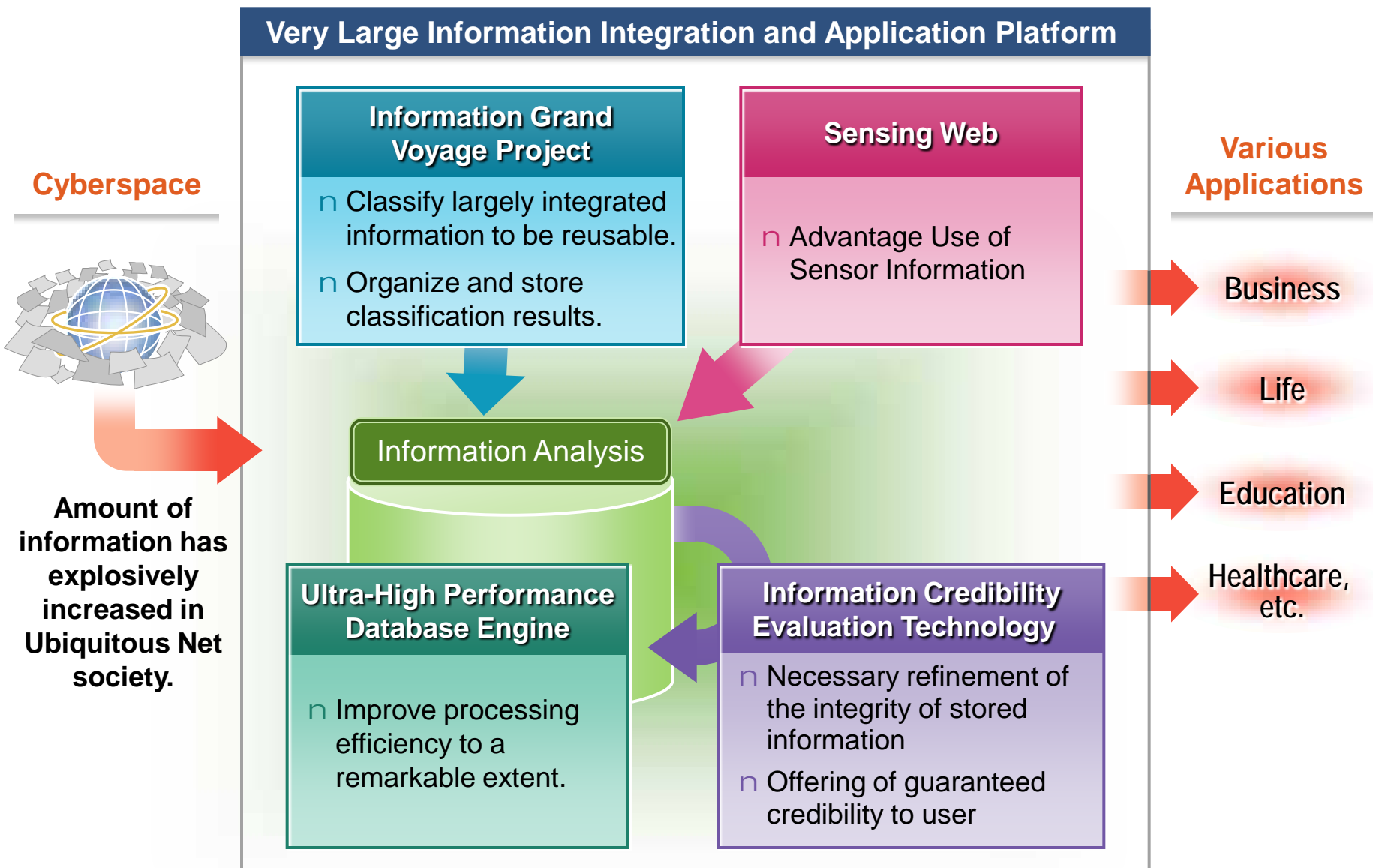
##### MIC

R&D of Information  
Credibility  
Evaluation  
Technology

##### Complementary Project

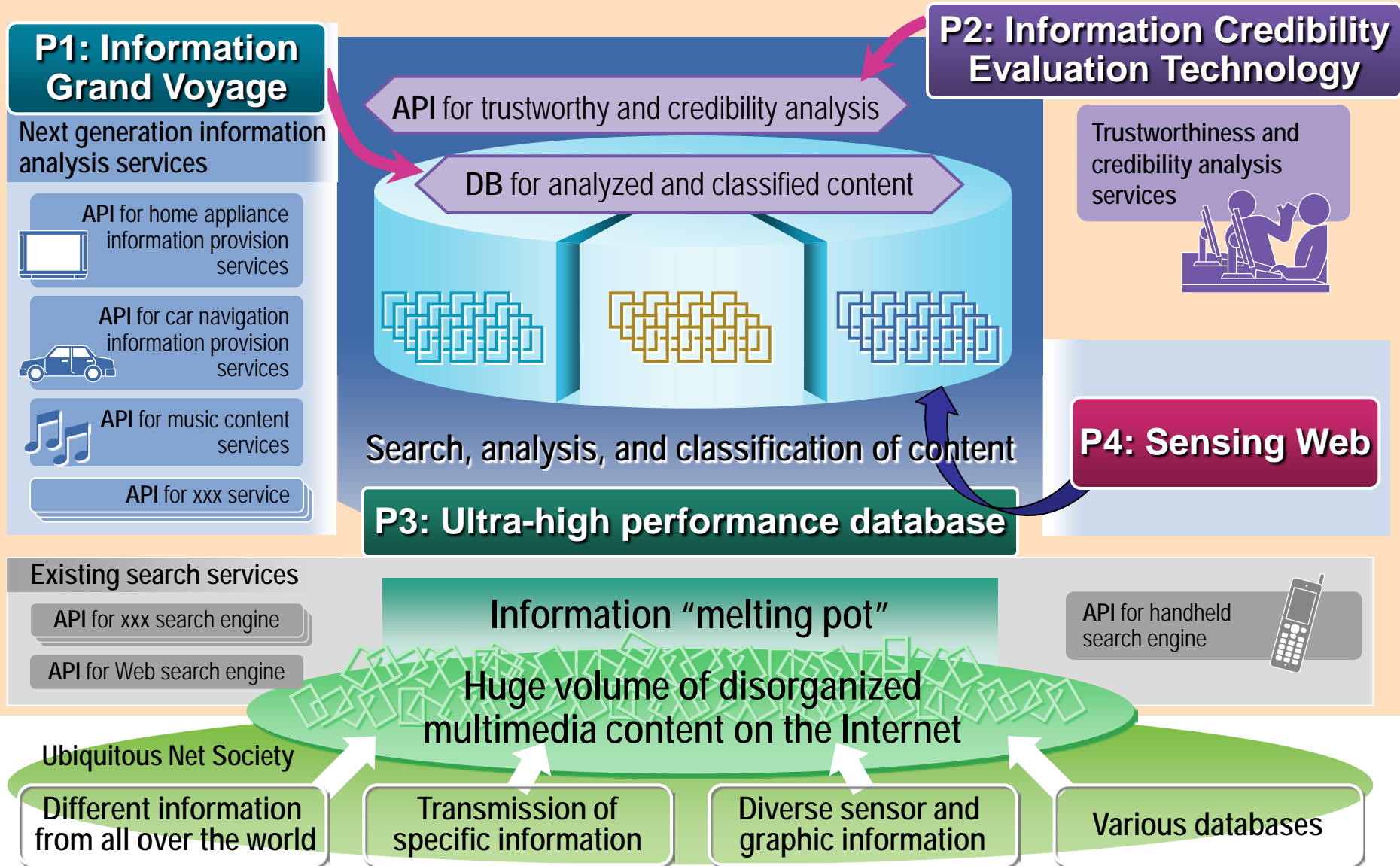
Sensing Web:  
Advanced Use of  
Sensor Information

## 2.3 Positioning



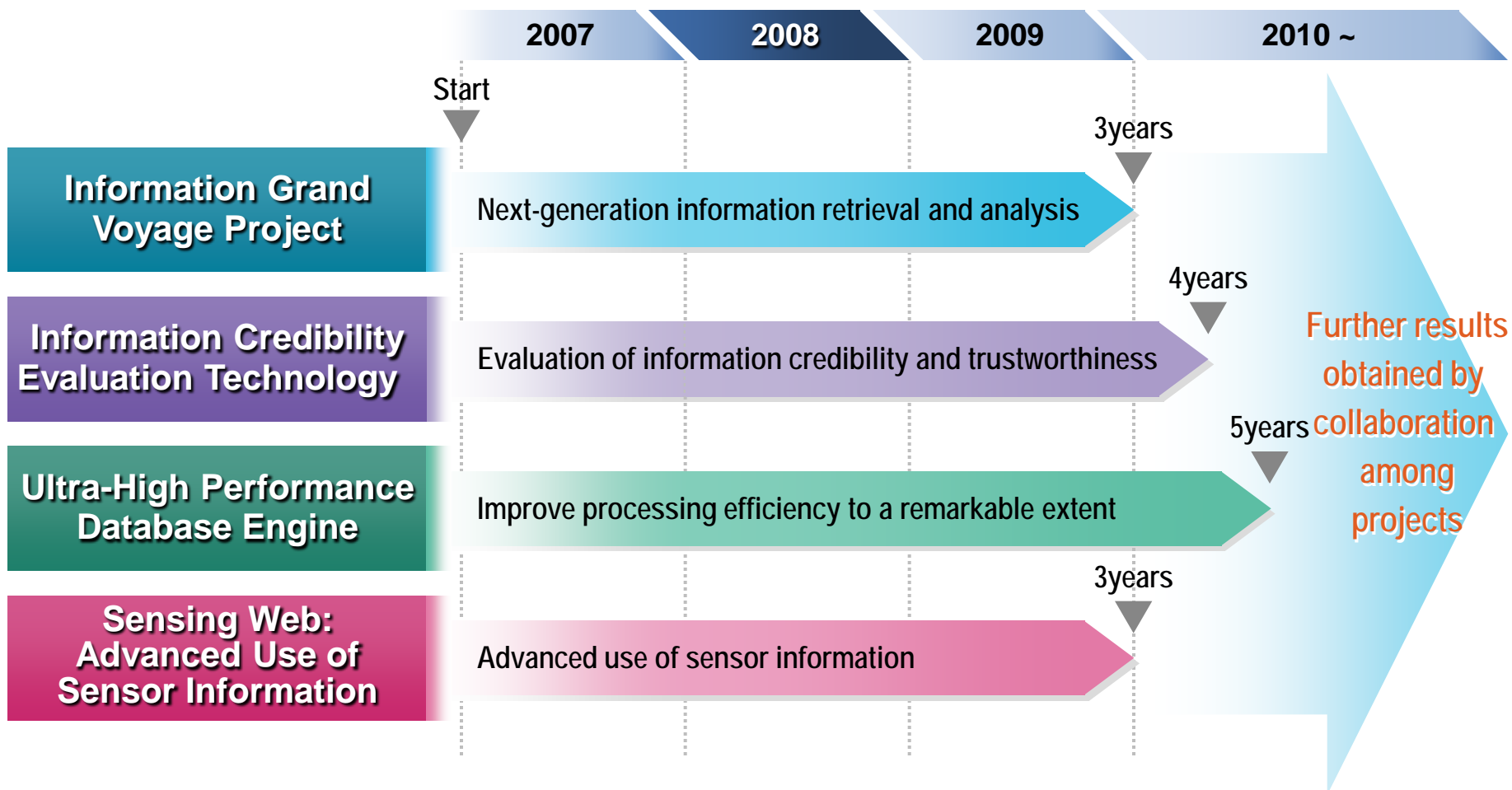
## 2.4 Goal:

# Fiscal 2011 “ Very Large Information Integration and Application Platform ”



## 2.5 Schedule

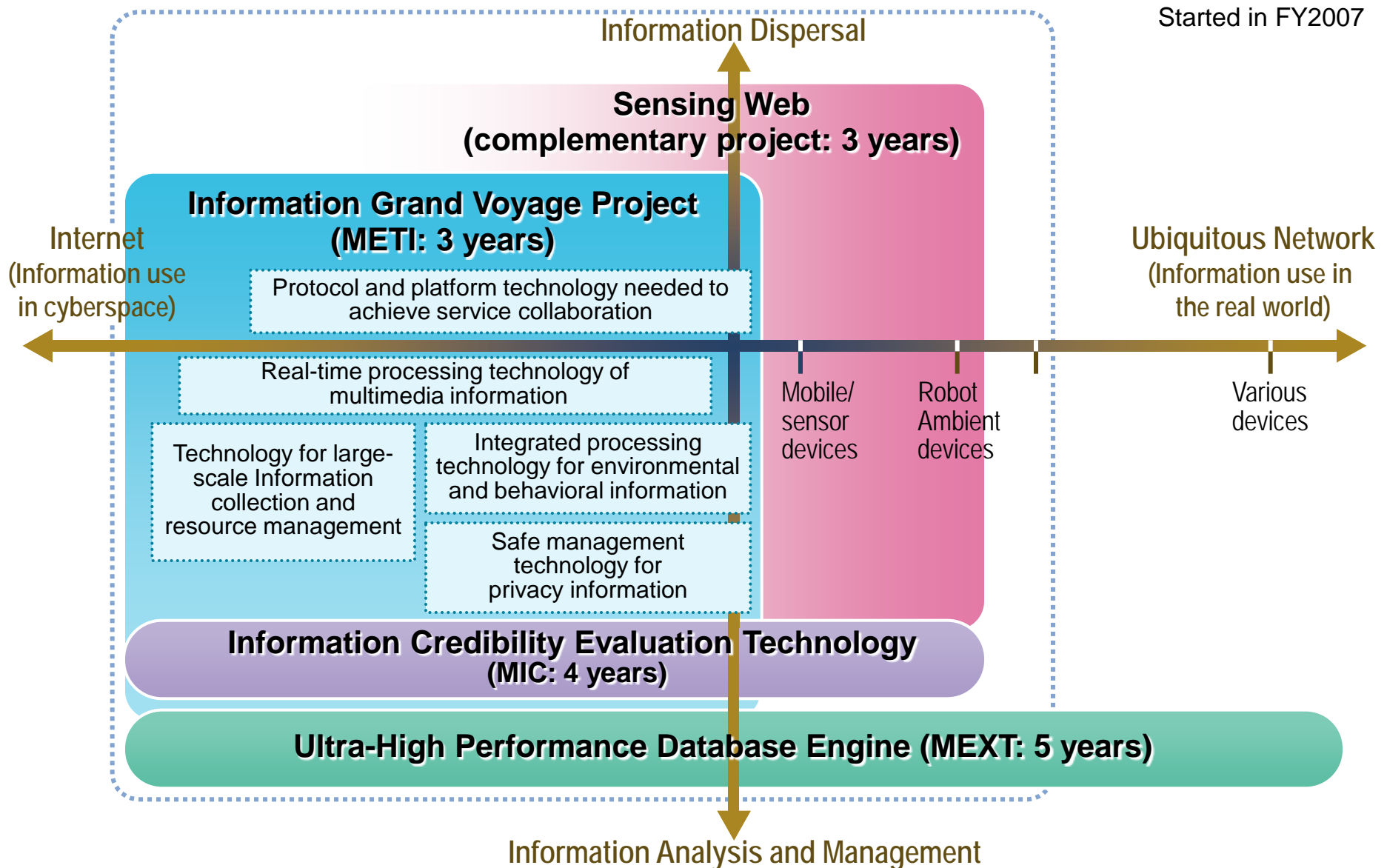
# Very Large Information Integration and Application Platform



## 3. Four Major Projects of Ministries (METI, MIC, MEXT)

- 3.1 Overview
- 3.2 Information Grand Voyage Project
- 3.3 Information Credibility Evaluation Technology
- 3.4 Ultra-High Performance Database Engine
- 3.5 Sensing Web: Advanced Use of Sensor Information

# 3.1 Overview



## 3.2 Information Grand Voyage Project

METI: Ministry of Economy, Trade and Industry

<http://www.meti.go.jp/english/index.html>

[http://www.igvpj.jp/index\\_en/](http://www.igvpj.jp/index_en/)



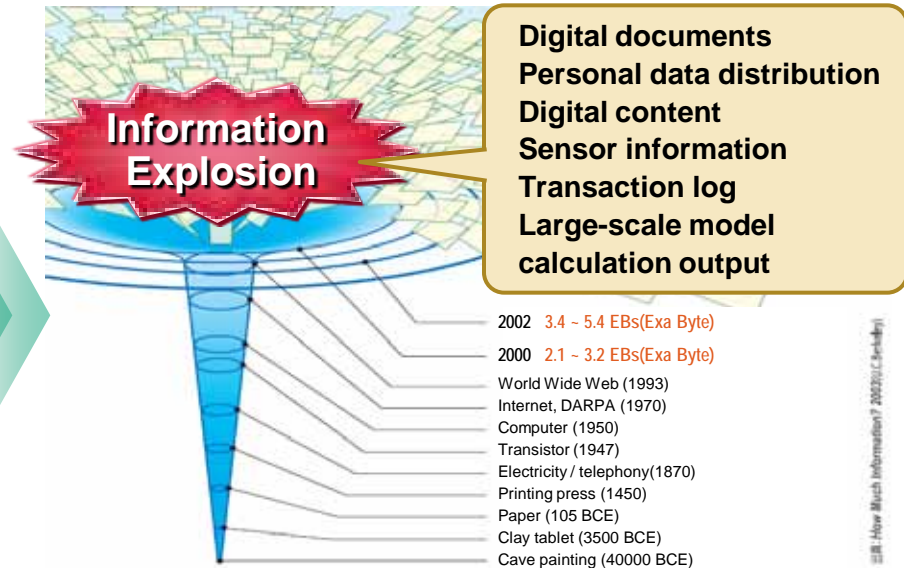
## 3.2 Information Grand Voyage Project (METI)

### Background

Growth of Web

Rich Information

Emerging sensor technologies



### Objectives: Info-plosion driven aspirations

Applying “Information Explosion” to achieve a richer social environment:

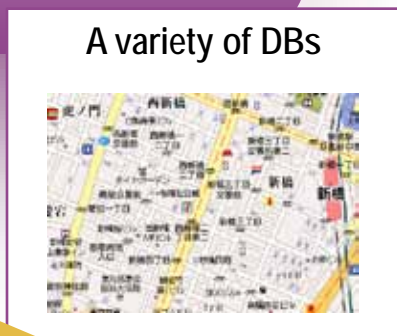
- n Build the infrastructure for next-generation information retrieval / analysis technology to help users gather and analyze a huge quantity and variety of information on the Web and elsewhere with ease, convenience, and success.
- n By fiscal 2011 , make it possible to strategically exploit mega data, central to the era of information explosion.

# Information Fusion by Information Grand Voyage Project



**Web Information**

**Non-Web information**



**Next-generation information retrieval / analysis technology**

**Information Fusion**

**Greater innovation and Richer info-based society**

# Commitment of the Information Grand Voyage Project

## Triune Approach

### Demonstration with model services \*

- n Verify effectiveness and feasibility of the next-generation information retrieval / analysis technology using 10 model services

\* More detail description next slide

Establish mechanism for independent innovation through practical use of information

**Industrial implementation and strengthening competitive power**

### Establish and revise regulations and environment

- n Establish and revise regulations on privacy and copyright protection.
- n Create mechanism for smooth distribution of intellectual property
- n Prepare an environment for development and demonstration

### Technology Development \*

- n Develop next-generation information retrieval / analysis technology
- n Achieve versatile and common retrieval / analysis technology (provide a common technology)

# Model Services in FY2007

## A Future-oriented personal services considering privacy

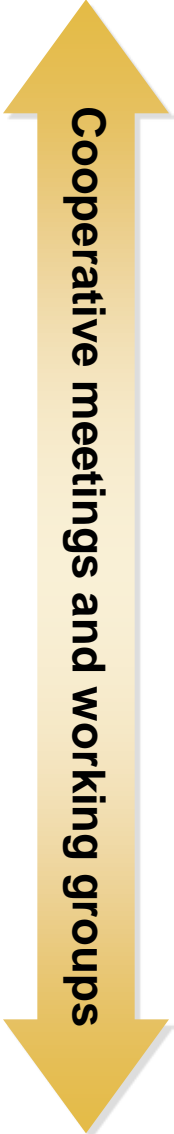
- n A-1 **My life assist service** (Behavior-based chain search using cell phone)
- n A-2 **Cooperative service using contactless IC card for traffic** (Sharing sensor information with PASMO)
- n A-3 **Profile passport** (Application of behavior history using game machines)

## B Next-generation Web services giving new values

- n B-1 **Laddering search service** (Interactive search)
- n B-2 **Sagool TV** (Internet movie search)
- n B-3 **View search Hokkaido** (Next-generation image search)

## C IT services based on new social infrastructure

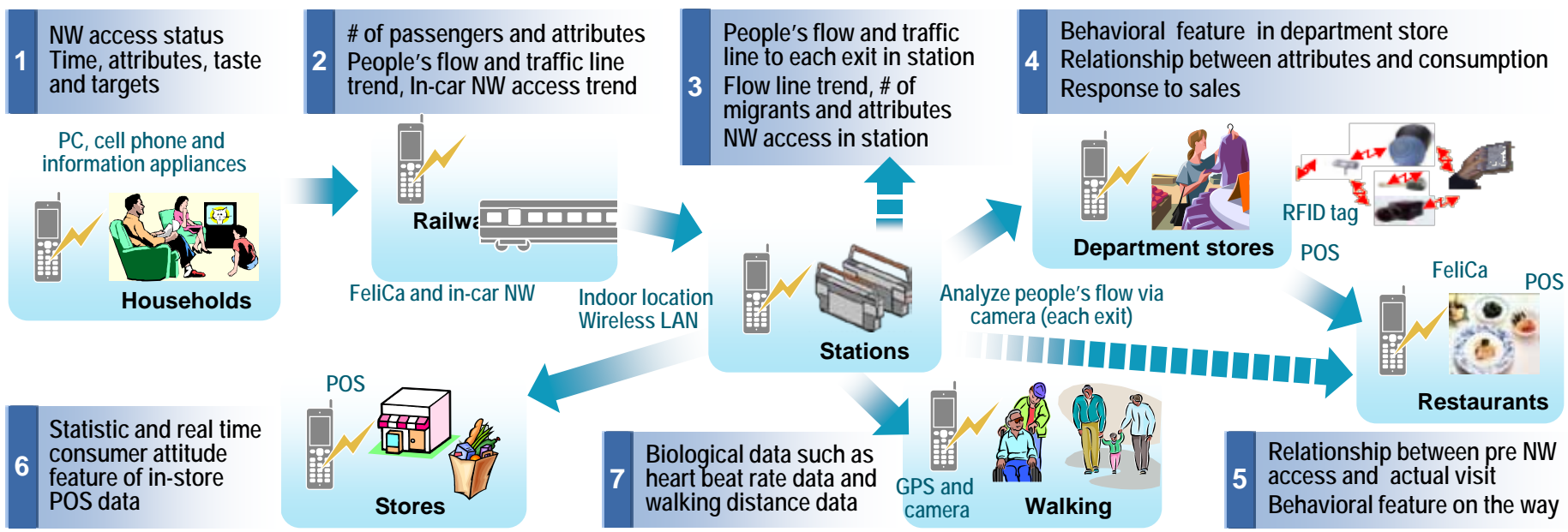
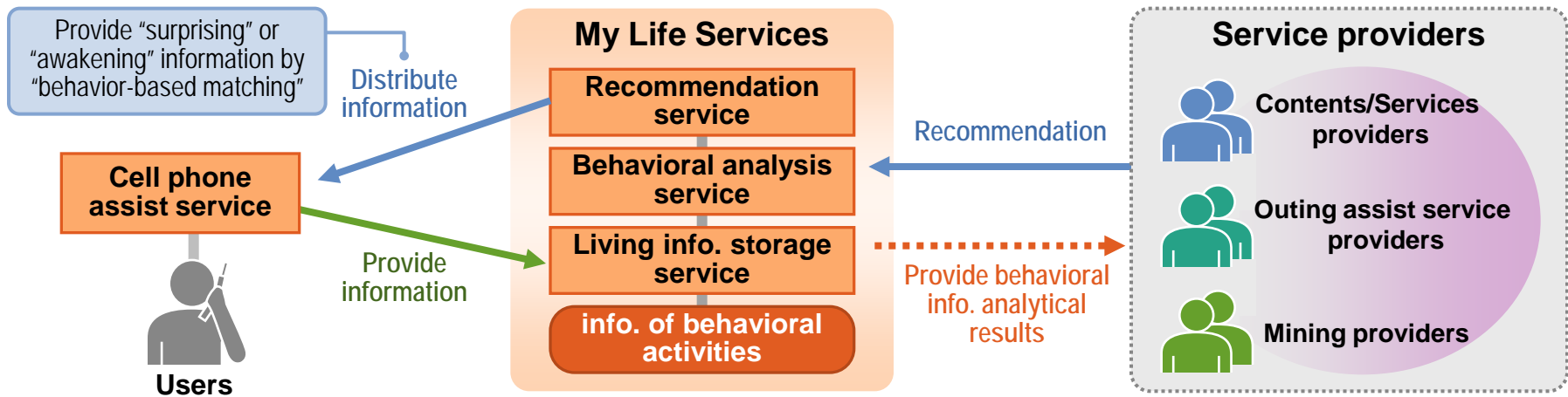
- n C-1 **Time-space information mining service** (Real-time information retrieval)
- n C-2 **Health Monitoring and assistance Service** (Medical sensor)
- n C-3 **New integrated safe flight support system** (Collection and analysis of safety information)
- n C-4 **Mega research** (Interview-type questionnaire)



Cooperative meetings and working groups

# Example of Model Service: A-1 My Life Assist Service (1)

## Cell phone assists you in real life!



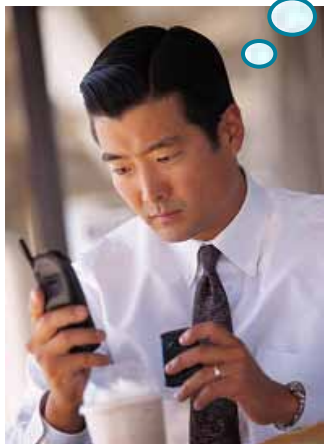
# Example of Model Service: A-1 My Life Assist Service (2)

## Preview channel

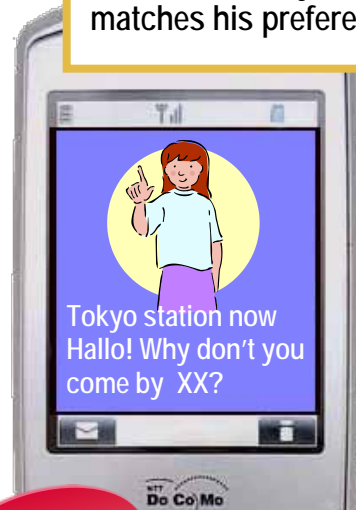
You can get contents related to your everyday behavior and your potential desires  
(It starts by opening the folding cell phone)

One day, I got off work on time and opened my cell phone on the way home ...

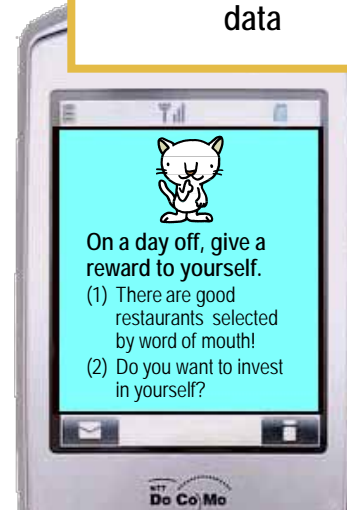
Well, I have worked overtime and came home late for several days ...



Surprising proposal that hasn't appeared in his recent activity and matches his preference



Propose a new activity by predicting her next behavior from her past data



I did a good job this week. I'll give myself a reward!



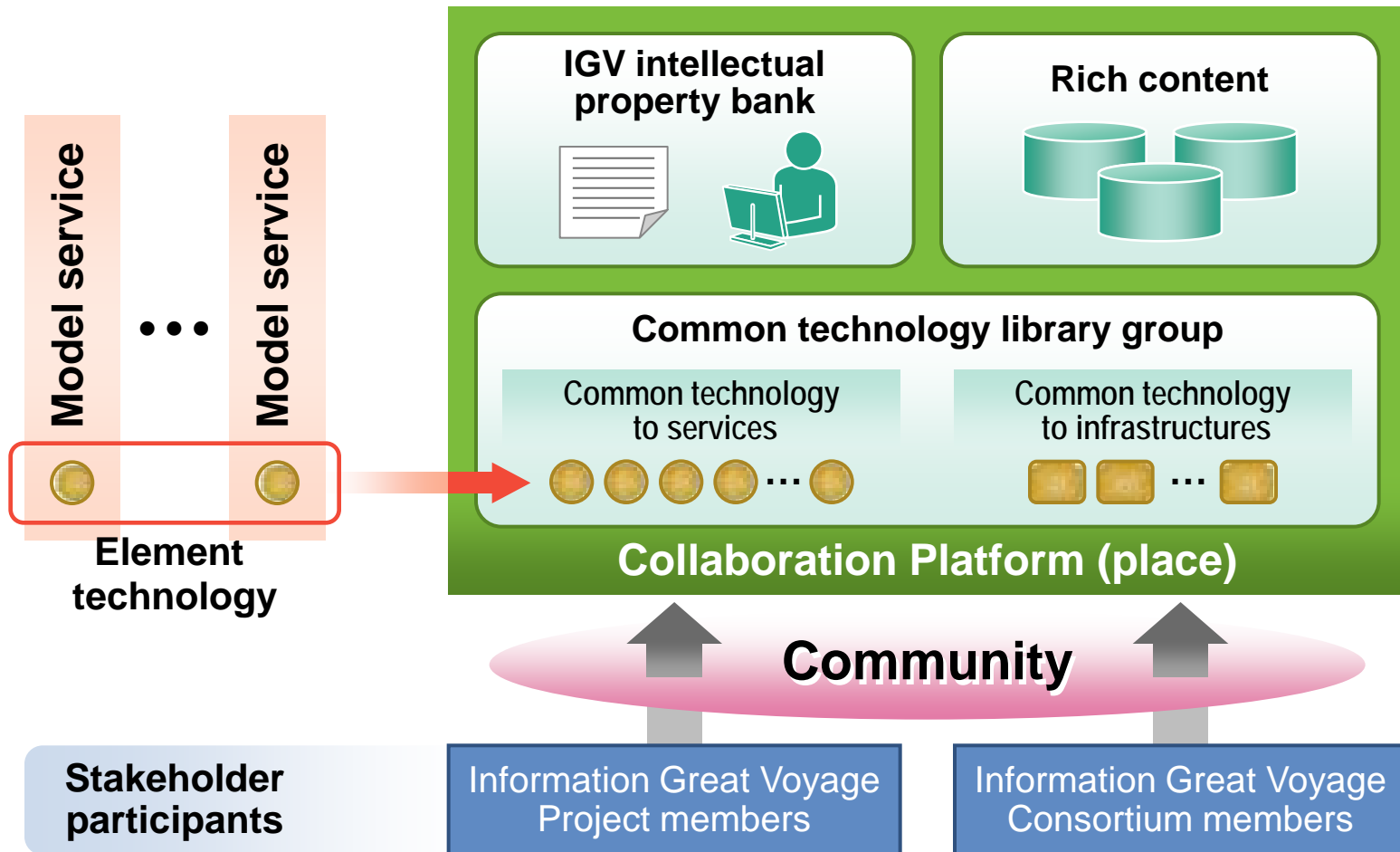
**Click!**

- n List of restaurants in Yokohama
- For relaxation
- For stress release
- For healing ...
- Linked to information

# Technology Development: Preparation of Collaboration Platform

## Definition

- (1) Place to share achievements of Information Grand Voyage Project
- (2) Platform to accelerate R&D of the next-generation information-access technology and enhance competitive edge



## 3.3 Information Credibility Evaluation Technology

MIC: Ministry of Internal Affairs and Communications

<http://www.soumu.go.jp/english/index.html>



## 3.3 R&D of Information Credibility Evaluation Technology (MIC)

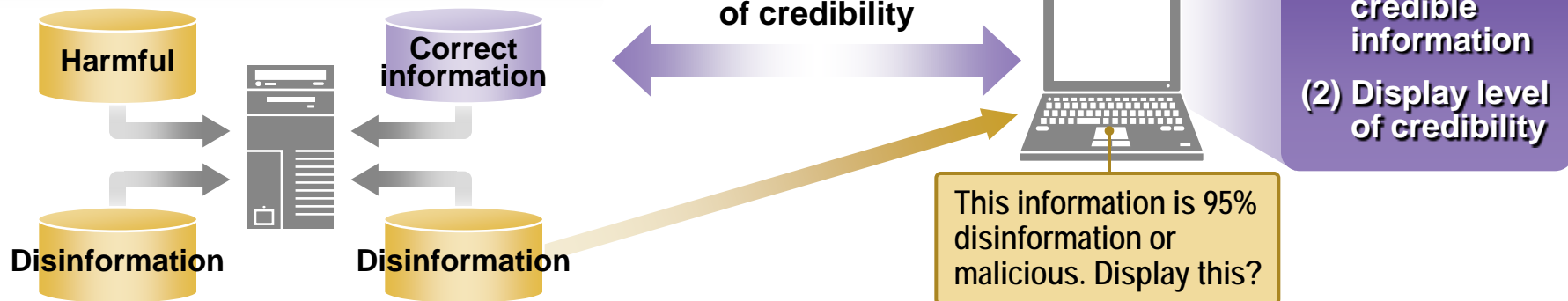
### Current Problems

- n Judging credibility of information is left to users based on experience, intuition, etc.
- n In the large volume of information flowing through the Internet, valuable info is mixed into unknown, harmful, and incorrect info: obtaining needed info is difficult!
- n Phishing and various disinformation scams cause real damage!

### Project Goal

To create an environment where anyone can manage and use text, audio, and graphic information on the Internet with ease and security by establishing technologies to detect false and deceptive information and thus supply credible information.

Understand the meaning of information over the Web and analyze its relationship with other information

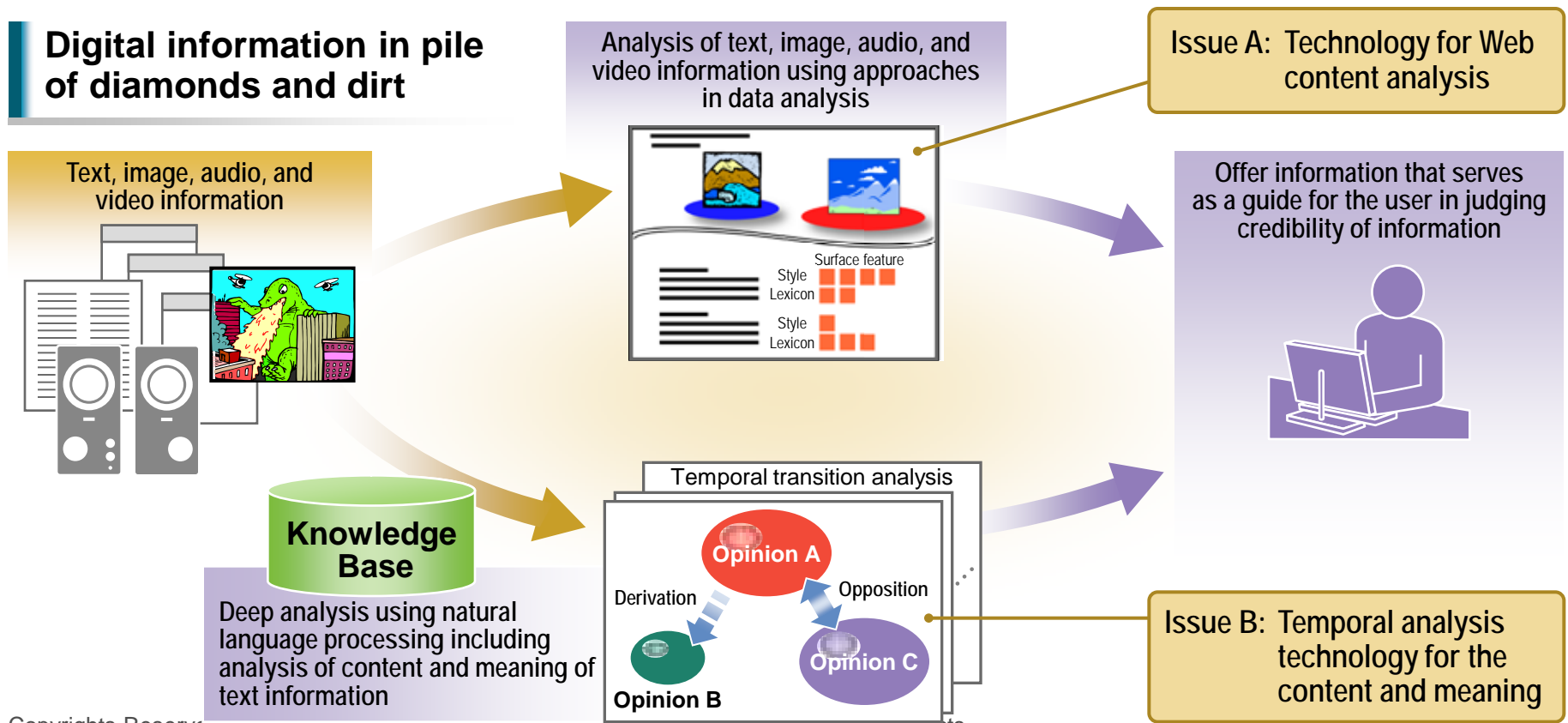


# Overview

## Research and Development on Evaluation Technologies for Information Credibility

- n Develop technologies to analyze information on the web and to predict the probability of its credibility
- n Offer analysis results so users can evaluate information as “good” or “bad” based on the user’s own judgment criteria

### Digital information in pile of diamonds and dirt



# Issue A: Technology for Web content analysis

Develop technologies to analyze the credibility of large volumes of web-based text, graphic, audio, and video information using methods in data analysis.

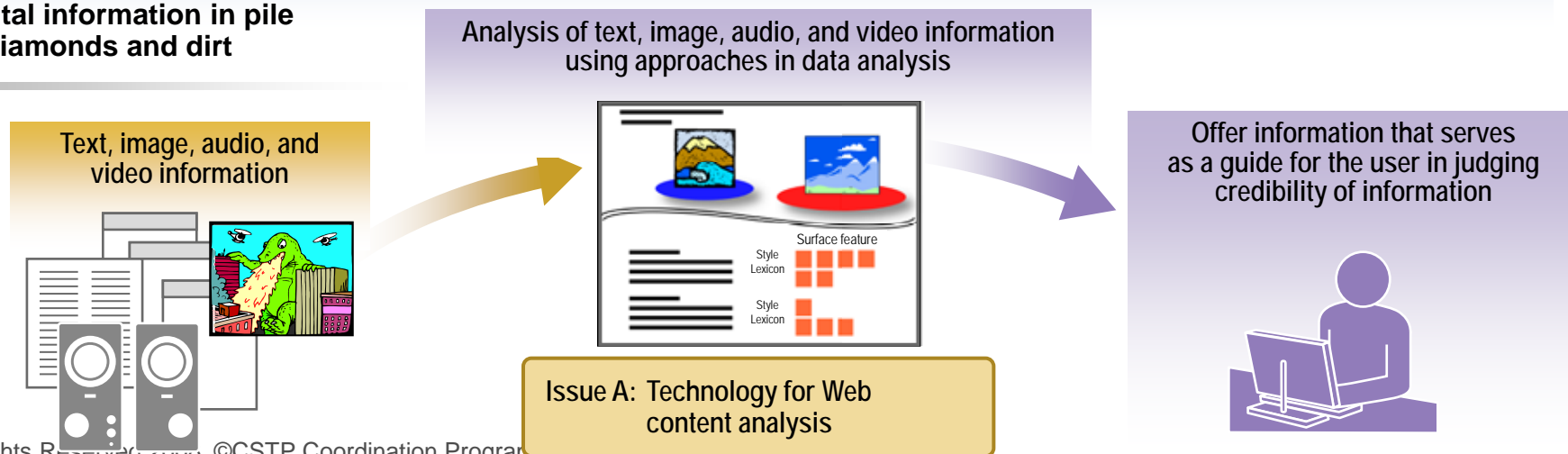
## Technology to analyze graphic, audio, and video information

- Develop a **cross-media information analysis engine** to verify inconsistencies in graphic, audio, and video information, and their explanatory text information

## Technology to analyze text information

- Develop a **surface-feature analysis engine** to analyze the quality of text from the surface features of web text such as blogs; such features include typographical errors, omitted letters, lexical diversity, and sentence readability.
- Develop a **sender analysis engine** to analyze the features of the topics most frequently raised by the sender based on sending history, etc.

### Digital information in pile of diamonds and dirt



## Issue B-1: Temporal analysis technology for meaning and content (analysis of content and social evaluation of information)

Natural Language Understanding and Evaluation technology to extract opinions latent in the collection of text on the Web and logical relationships between opinions

### Natural Language Understanding and Evaluation technology

- n **Develop the technology for building large-scale lexical knowledge** to use in capturing the logical relationships between opinions
- n **Develop a logical relationship-recognition engine** to identify the logical relationship between any two opinions
- n **Develop an opinion-analysis engine** to analyze and sort information related to fields and viewpoints of interest to the user is interested and to present logical relationships between opinions as well as their importance
- n **Develop a summary-creation engine** to present a general view of the opinions in the fields of interest to the user

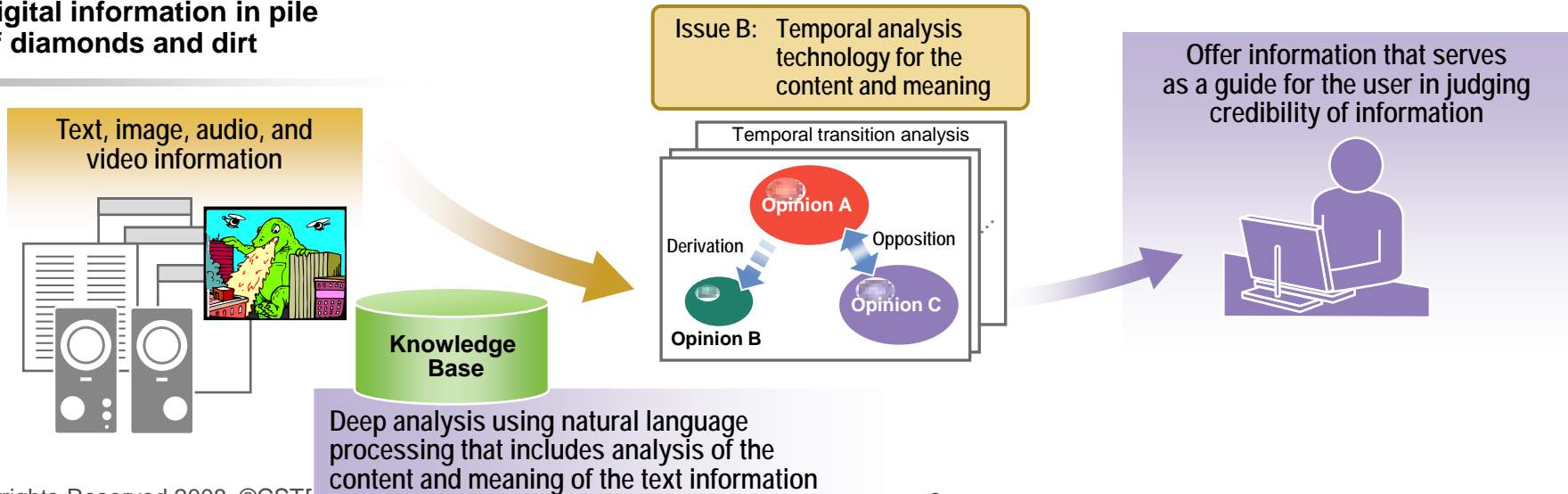
# Issue B-2: Temporal analysis technology for the content and meaning (analysis of content and social evaluation of information)

Temporal analysis technology to analyze the temporal transition in opinions, their mutual relationships, and process used to transmit the information

## Temporal analysis technology

- n **Establish a formula to handle opinion transition quantitatively**, and, **establish a way to express the transitions**
- n **Establish a temporal clustering method** to output opinion transition and relationship type relevant to needed information. **Develop information propagation phase-judgment engine** to determine the phase of the clustered information in the information propagation process: Whether the opinion is established, has spread out rapidly to many persons, or is already out of date

## Digital information in pile of diamonds and dirt



## 3.4 Ultra-High-Performance Database Engine

MEXT: Ministry of Education, Culture, Sports, Science and Technology

<http://www.mext.go.jp/english/index.htm>

## 3.4 R&D of Ultra-High-Performance Database Engine (MEXT)

### Background: Towards Database Engine for Info-plosion Era

- n Growth of amount of Data: so called 'Info-plosion' (Information Explosion)
- n Digital shadow is becoming substantial large
- n Traceability system introduced for various applications: PLM, GAP, Green IT, etc.
- n Our target: huge, not mid-range database

### Change

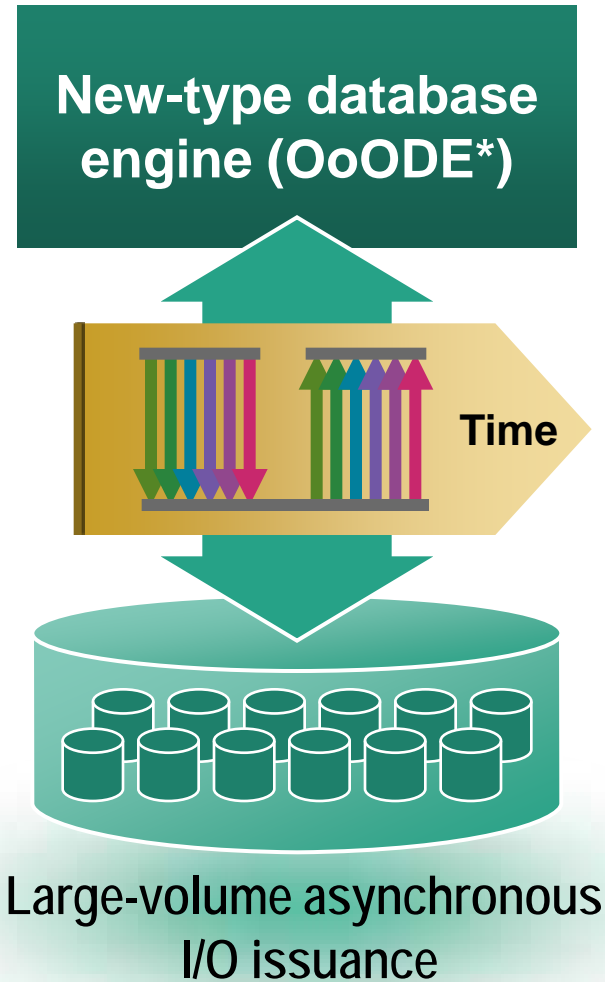
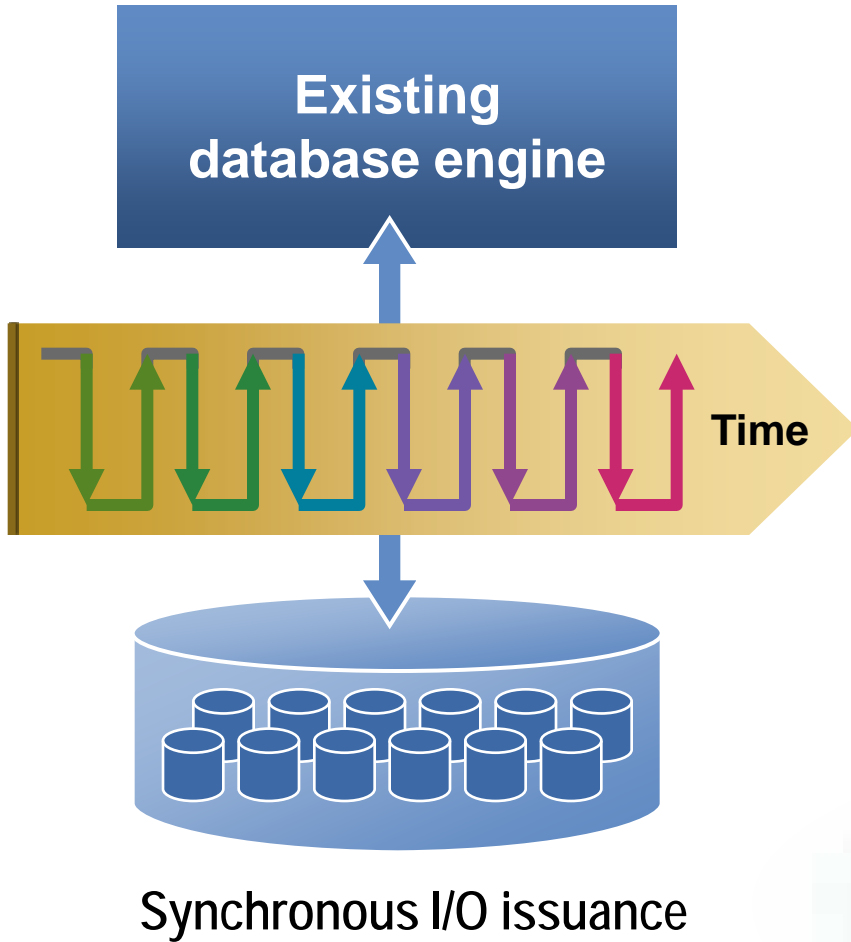
- n Computing Environment drastically changing
- n Increasing number of cores: New Moore's Law
- n Storage system revolutionized by SAN technology
- n Database Engine has only slightly changed

### Project Goals

- n Creation of technology providing ultra-high performance advantage for large-scale database management systems:
  - (1) Establish innovative execution principle "Out of Order Database Execution Principle"
  - (2) Design and implement database platform software based this principle
  - (3) Verify its effectiveness through actual use

# Performance Comparison: Existing and OoODE Database Engines

Drastic improvement in performance through Out of Order Execution



\* OoODE: Out of Order Database Engine



# OoODE (Out of Order Database Engine)

## Development Component

Application is constant

**Out of Order DBMS**

Implementation of new function

OS

Storage

Improvement of existing function

## Performance Prospects

Relative performance

100 times of performance improvement is planned

10 times of performance improvement is planned

100

10

1

2007

2008

2009

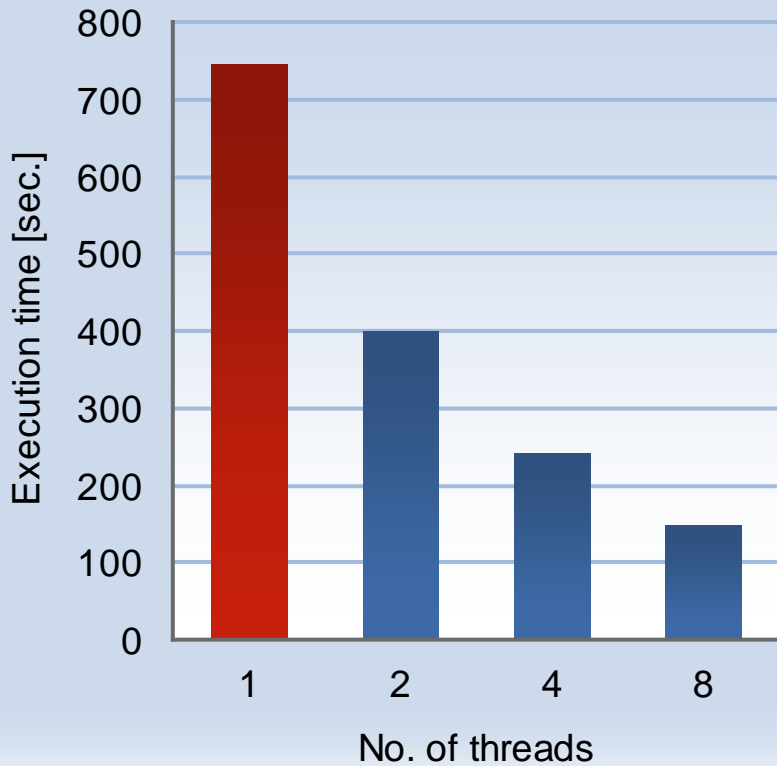
2010

2011

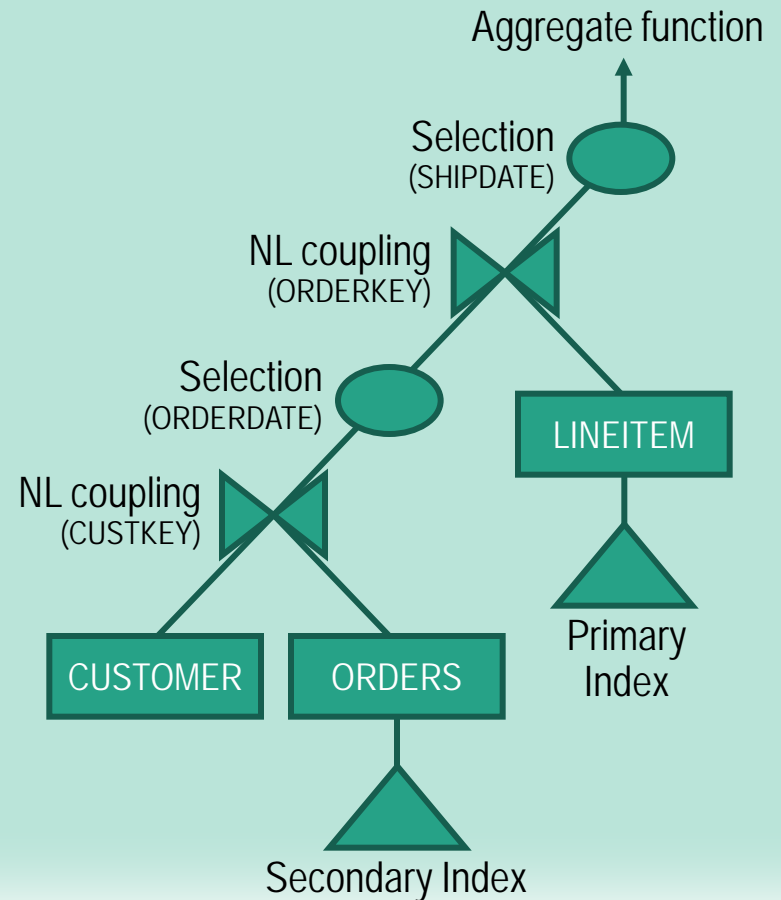
# Initial Experiment: Improvement of Performance through Large-Volume I/O Issuance Mechanism

## Results of Initial Experiment

TPC-H benchmark Q3 MySQL trial environment (1CPU)



## Example of execution tree in Out of Order Execution



## **3.5 Sensing Web: Advanced Use of Sensor Information**

Complementary Project: Kyoto University

<http://www.kyoto-u.ac.jp/en>

<http://www.mm.media.kyoto-u.ac.jp/sweb/index.html> (in Japanese)

## 3.5 Sensing Web: Advanced Use of Sensor Information

### Background

- n Web has grown worldwide: now a huge knowledge base
- n Sensor network, however, are not utilized widely:
  - | **Problem of protecting privacy in information:**  
Main target of sensor networks has been non-personal, such as weather monitoring
  - | **Variety of sensors in use:** Different procedures required for operation

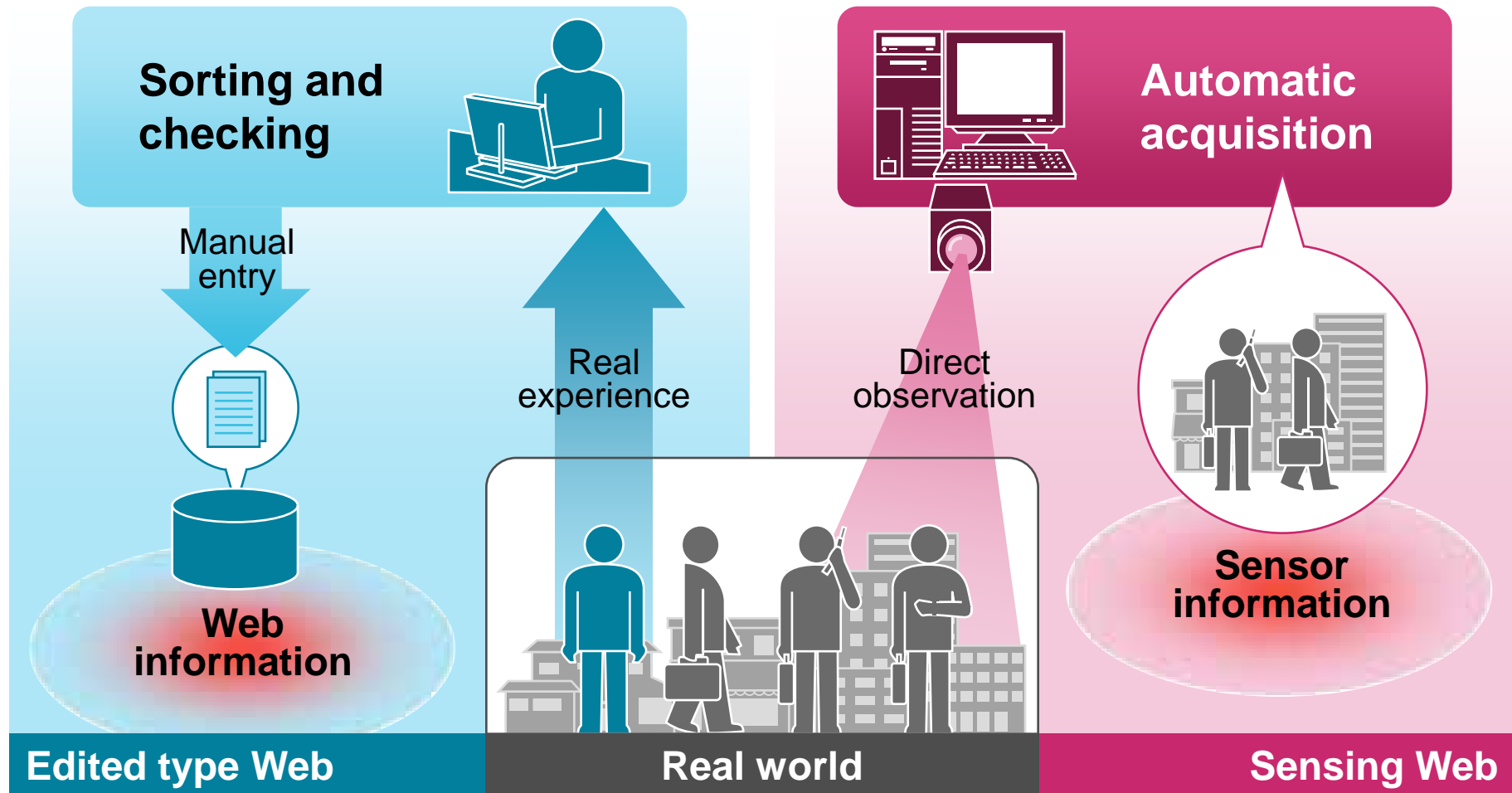
### Purpose

#### Creating “Sensing Web”

- n Sensing Web offers a Web-like environment for freely using sensor information by applying **privacy-protection technology** (e.g. filtering)
- n Real observed information can be used to create new multimedia services

# Difference between Web Information and Sensor Information

- n Web information requires human effort for information input
- n Sensor Information uses direct observation for automatic input



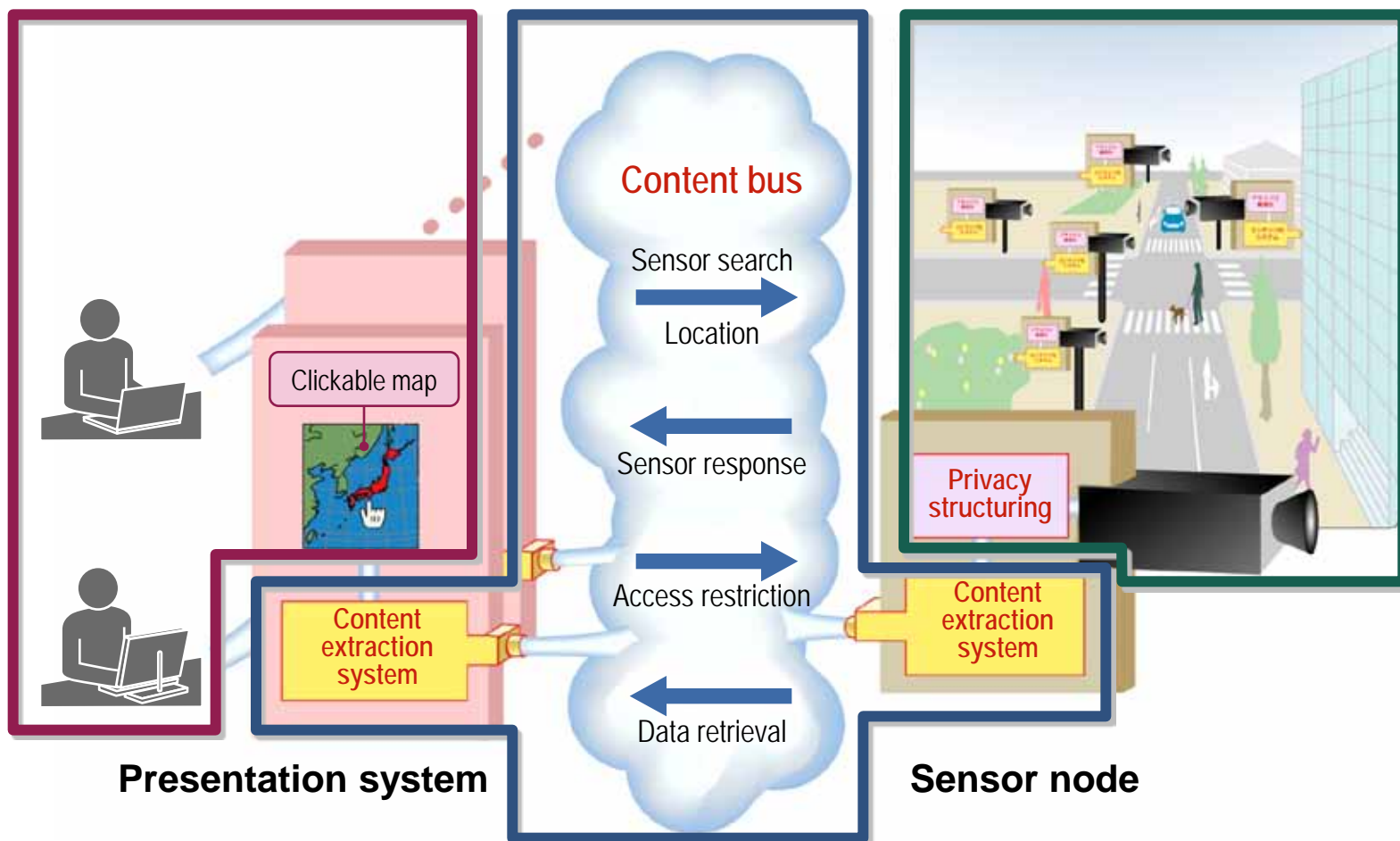
# Content Extraction of Sensor Information for Societal Use

## Sensing Web

**3** Observation real-world content-presentation technology (Utilization of Information)

**2** Technology for content extraction of sensor information (Information Sharing)

**1** Technology for privacy information management (Access Management)



# Framework of the project

## 1 Technology for privacy information management (Access Management)

### Responding to problems unique to sensor information

- n Privacy structuring and access management
- n Extraction method of information from streaming data

## 2 Technology for content extraction of sensor information (Information Sharing)

### Mechanism for sharing sensor information

- n Description method of sensor information including time-space context
- n Description method of requirements from user
- n Search and detection methods of required sensor

## 3 Observation real-world content-presentation technology (Utilization of Information)

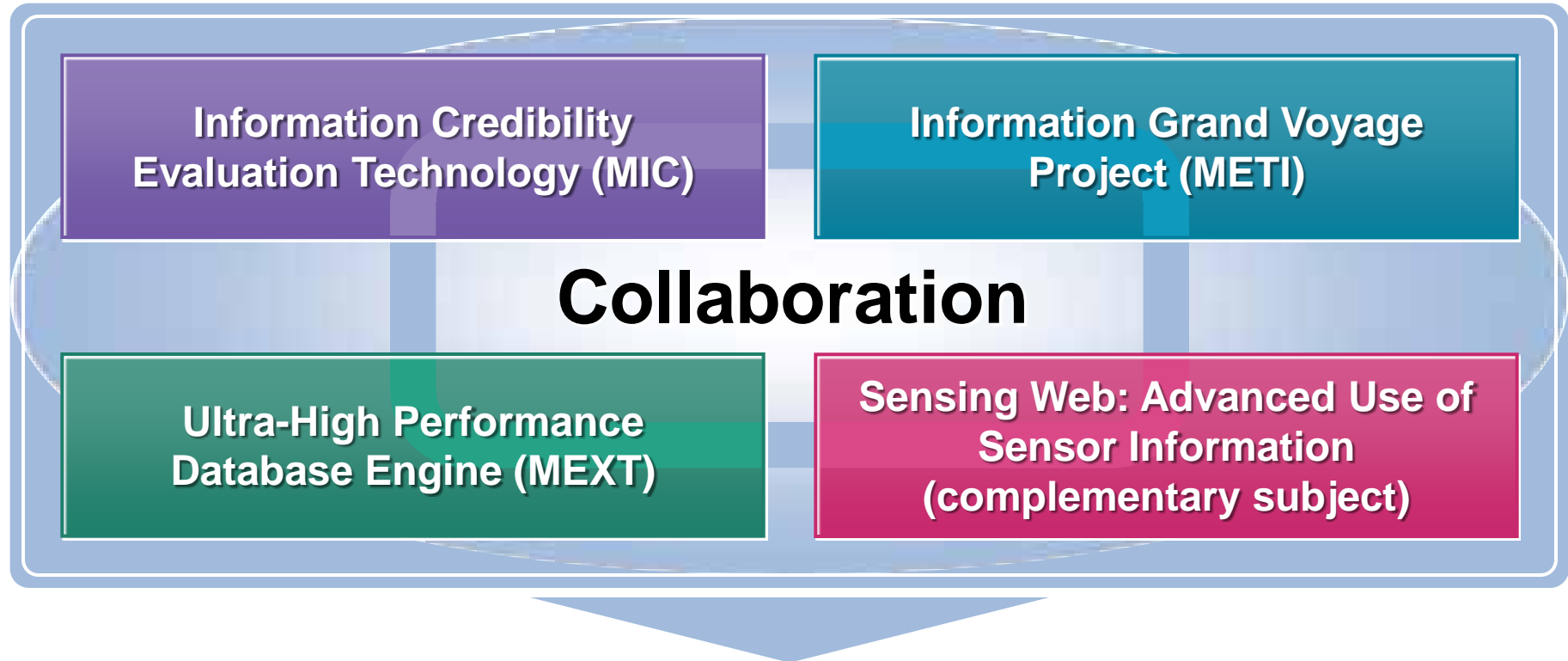
### Search and presentation of distributed sensor information

- n Collection method of information from largely distributed sensors
- n Information processing from sparse observation environment over a large area

## 4. Collaboration among Four Projects



# Collaboration among Four Projects



## Goals

- n To **build a next-generation information retrieval / analysis platform** enabling users to collect and analyze needed information from diverse sources inside and outside the Web easily, effectively, and reliably.
- n To **develop technologies for strategically using super-large-volume data** in this age of information explosion by fiscal 2011.

# 5. Conclusions and Future Work

## 5. Conclusions and Future Work: How do we advance this Coordination Program?

### Moving the coordination program into the future

#### Pioneering new fields in an age of information explosion requires application of information.

- n Contribute to an innovative environment by **establishing a platform technology that can be applied over the Web** using next-generation information retrieval / analysis technology focusing on media content
- n Increase the values of various projects by collaboration among them and adoption of existing technologies like Google to create new services and expand markets

#### Discussions among specialists in information retrieval, database, and language processing

- n A **“highly sophisticated” information retrieval / analysis technology** must be established within three-year development period of Information Grand Voyage Project
- n Opportunities must be provided for information exchange such as symposia and discussions with experts