

Very Large Information Integration and Application Platform's Activity Report

Coordination Program of Science and Technology Projects
"Very Large Information Integration and Application Platform"
Project Director: Shojiro Nishio, Ph. D.

February 4, 2009

Agenda

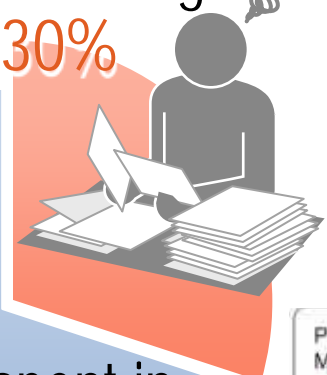
1. Background and Purpose
2. Coordination Program “Very Large Information Integration and Application Platform”
3. Achievements in the Projects of Each Ministry and the Complementary Subject in This Fiscal Year
4. Cooperation during Each Project
5. Future Plans

1. Background and Purpose

Advent of the Age of Information Explosion: Lost in Endless "Searching"

Time spent in searching

30%



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

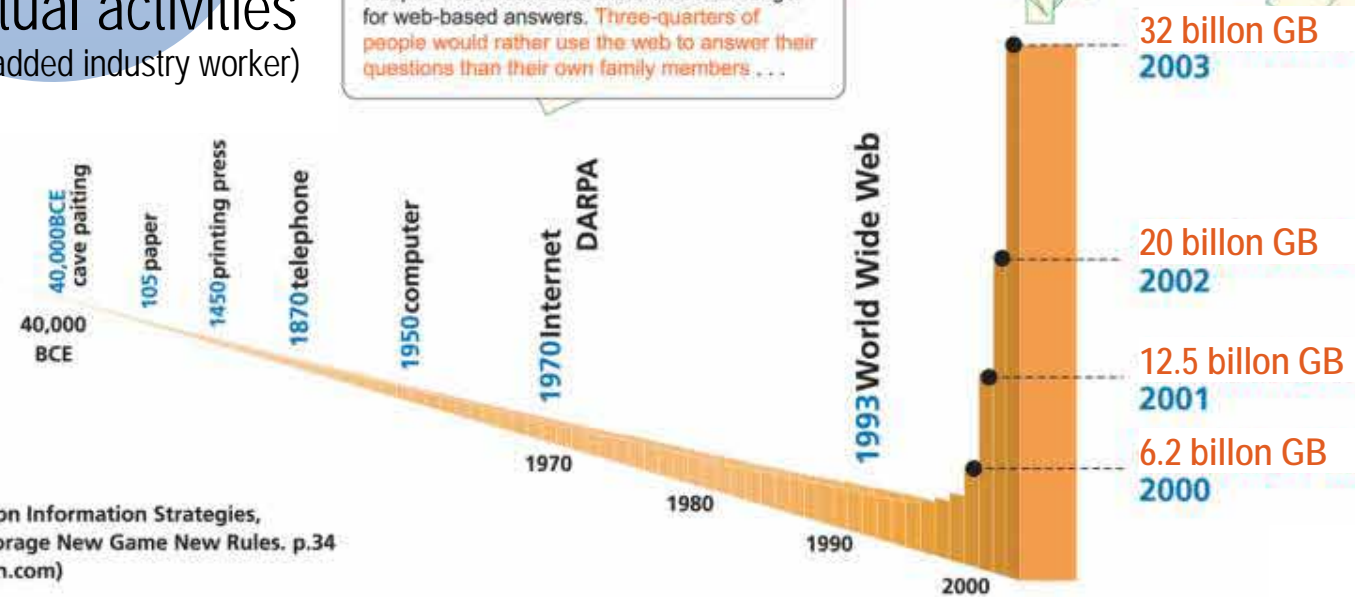


Aiming at Non-1sec response

Fifty percent of "complex" questions are left unsolved. Three people out of four ask web rather than their family.

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...**

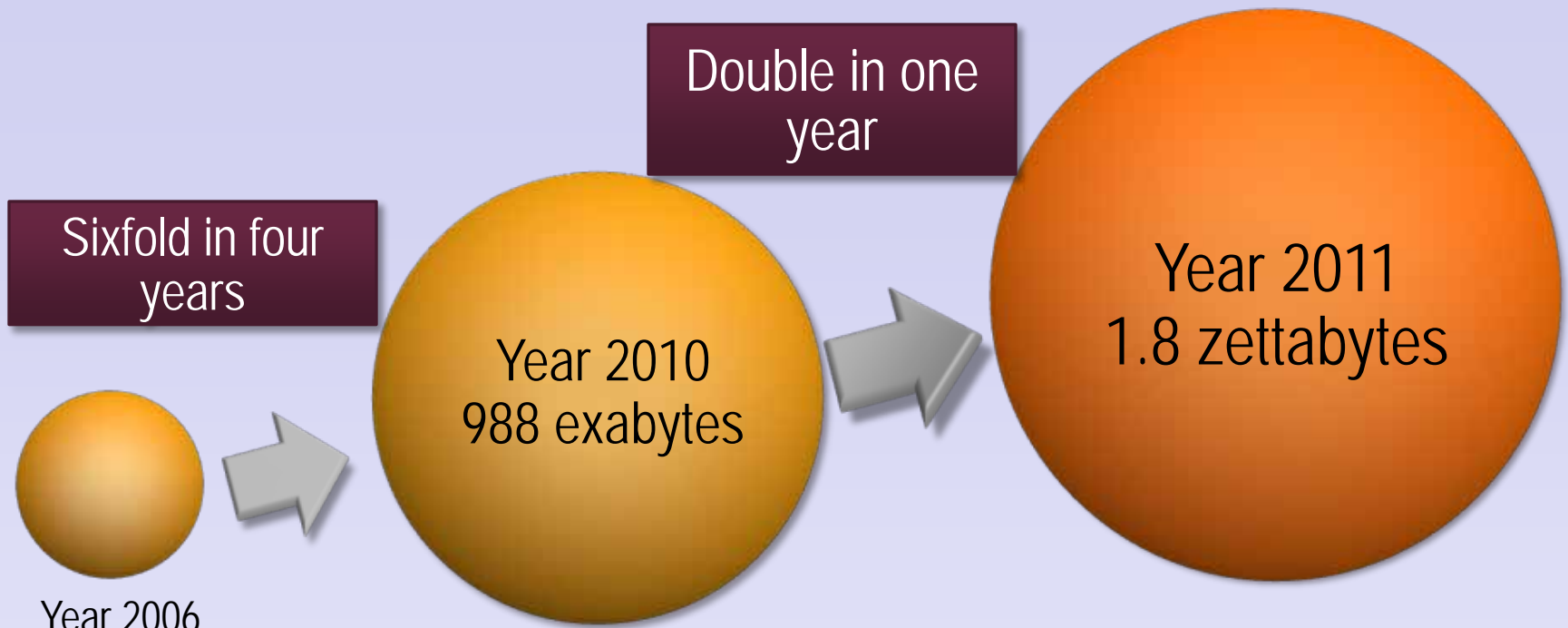
Information explosion



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

Coming Age of Information Explosion

Amount of Information Created, or Acquired and Accumulated
as Electronic Data all over the World

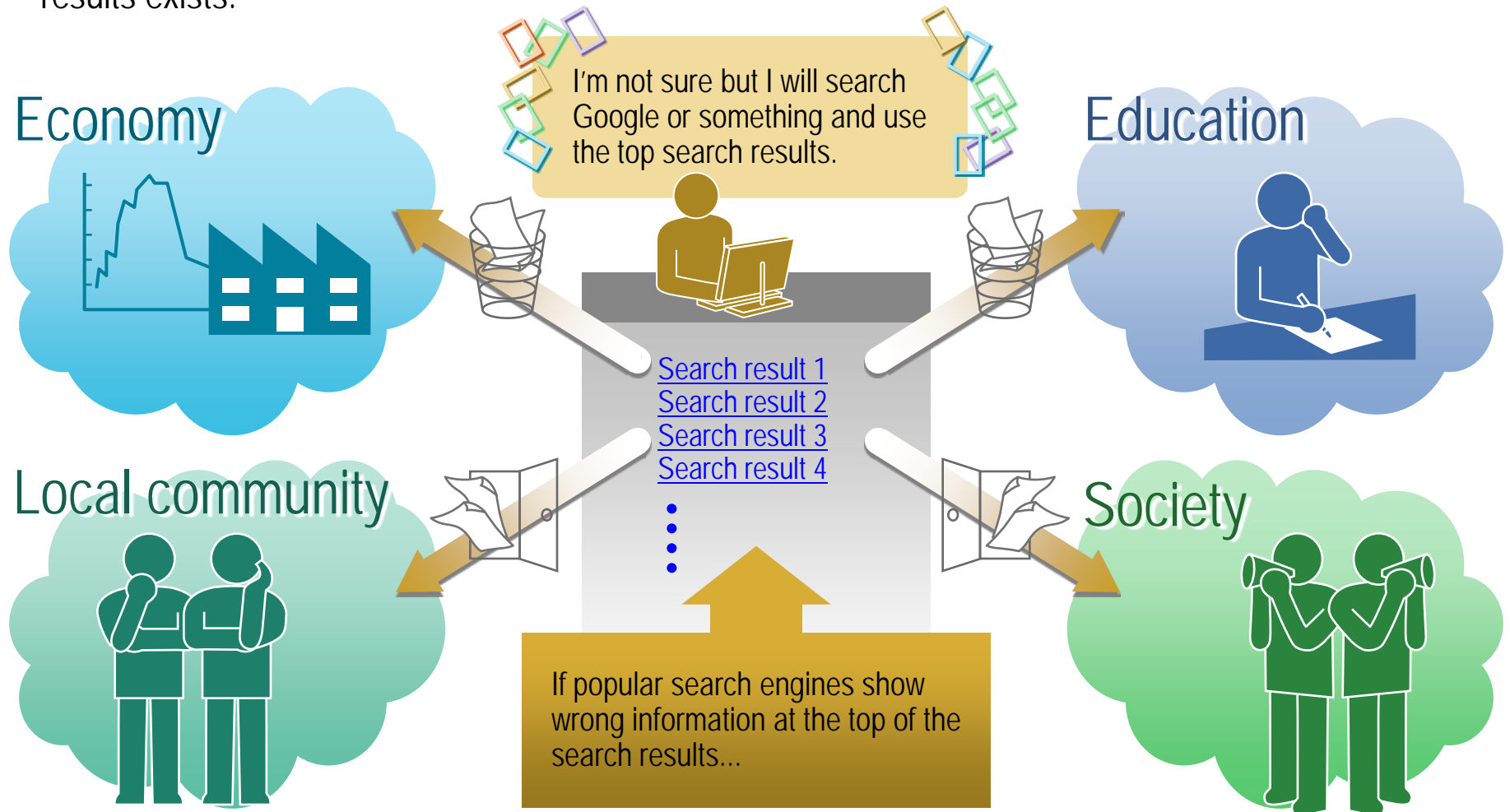


1 exabyte = 10^6 trillion bytes

1 zettabyte = 10^9 trillion bytes

Examples of Societal Problems Faced by Existing Search Services

- Current search engines does not analyze and present information from the user's perspective.
- As a result, the possibility of garbage and harmful information being presented at the top of search results exists.



Visage of a Society Actively Using an Intelligent Information Access Platform

Target: Building society through the use of safe and intelligent services

Mixture of good and bad digital information

Economy



Education

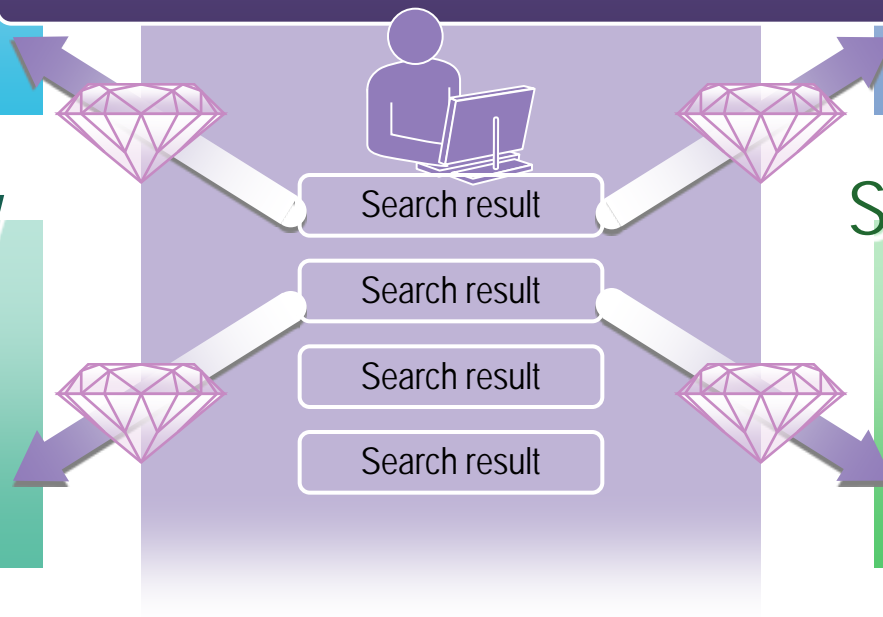


Intelligent Information Access Technology
Platform Ver.2011

Local community

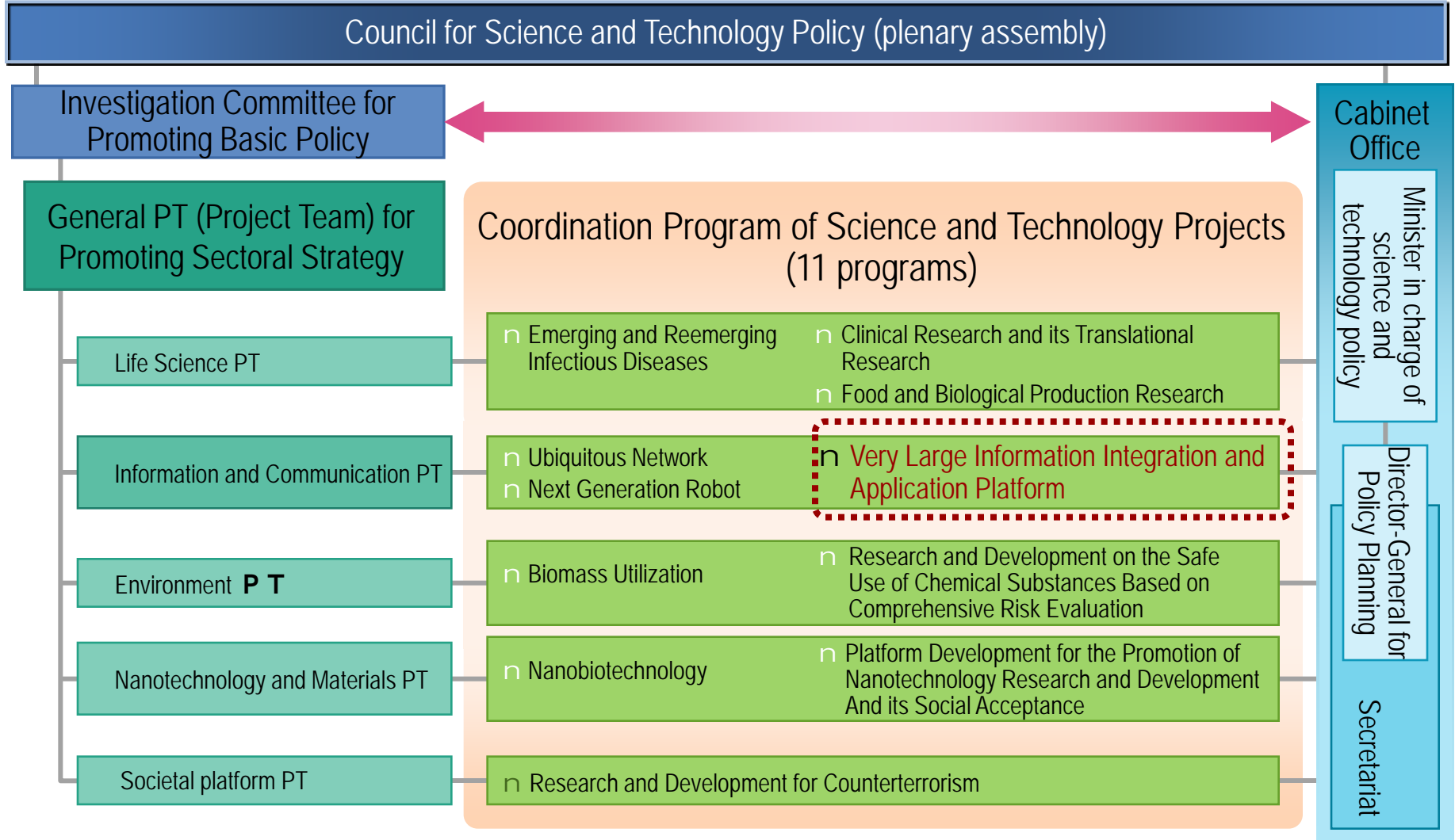


Society



2. Coordination Program “Very Large Information Integration and Application Platform”

Promotion Scheme for Coordination Program of Science and Technology Projects

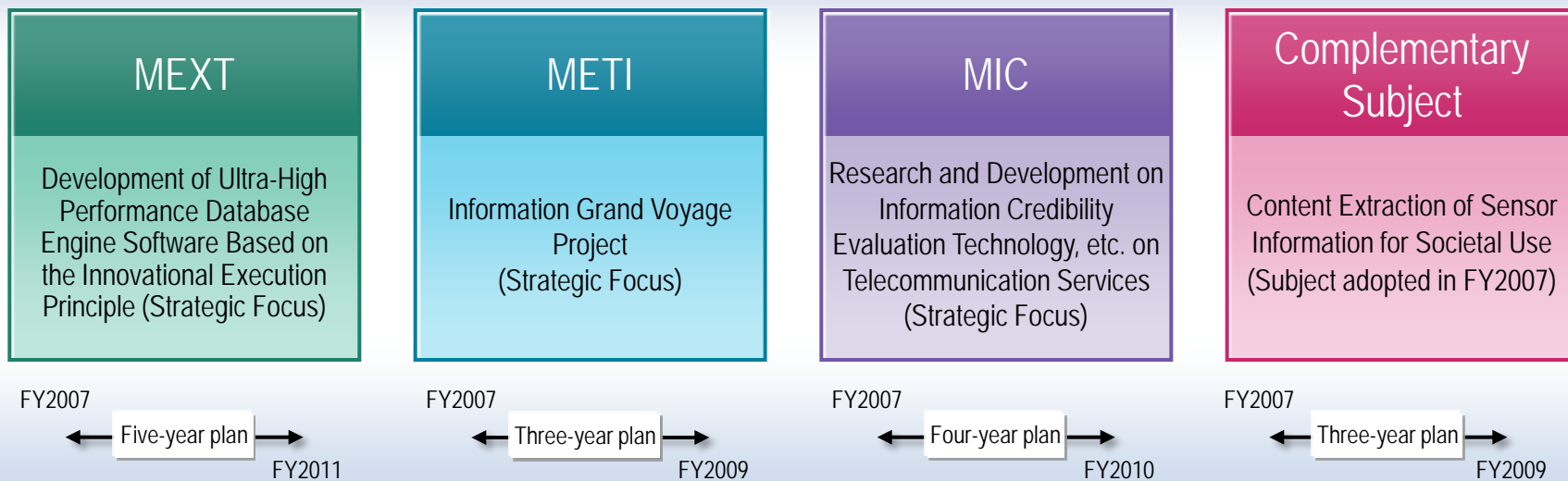


Coordination Program "Very Large Information Integration and Application Platform"

Objectives


To develop the platform technology for intelligent information application of next generation, for safe, appropriate, and secure information collection, analysis, and management of various information (content) to provide unique information service

Coordination of Each Ministry's Strategic Project




Project Positioning

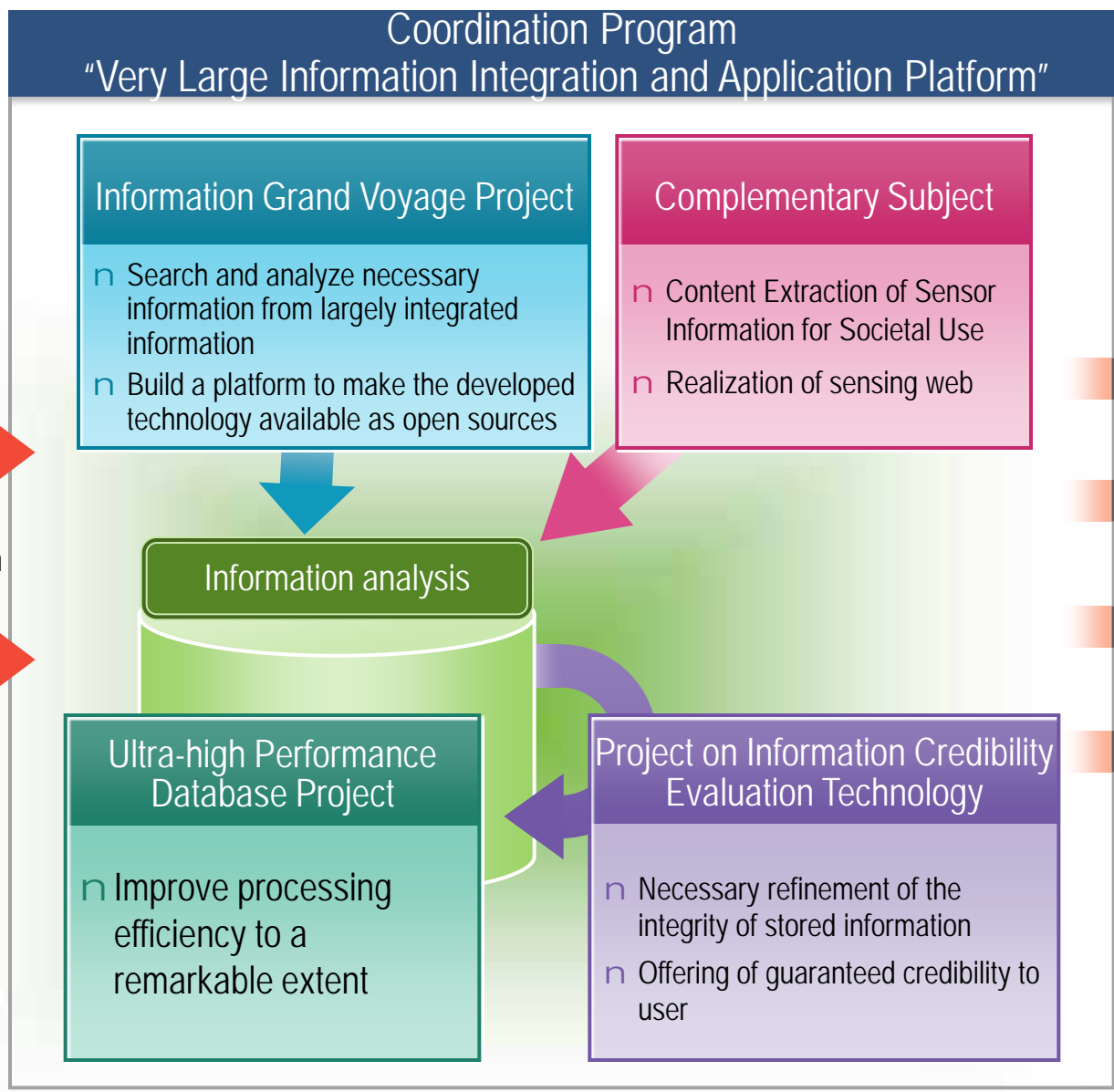
Cyberspace




The volume of information has explosively increased in the Ubiquitous Net society.



Real world

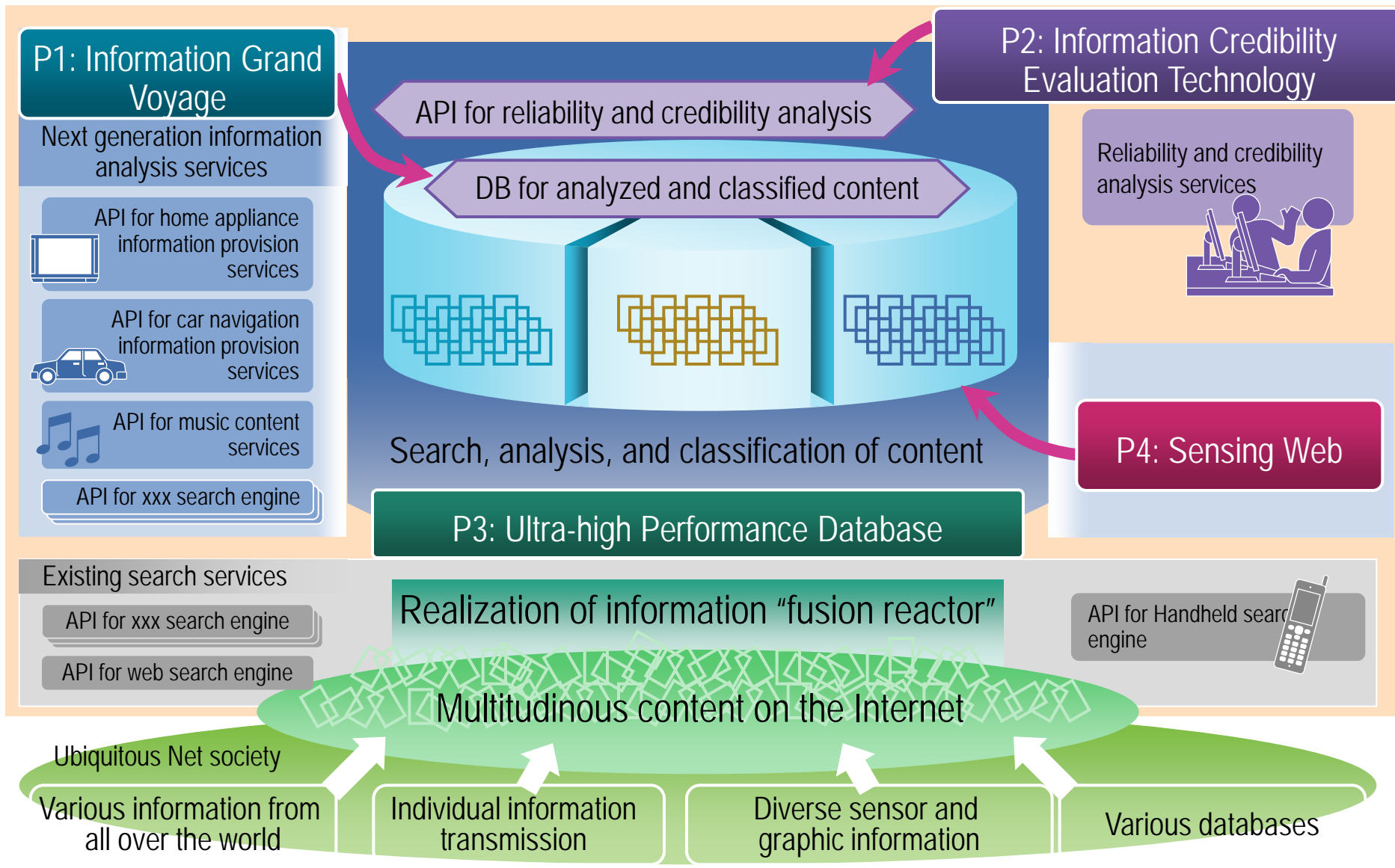


Various Applications



- Business
- Life
- Education
- Healthcare, etc.

"Intelligent Information Access Technology Platform" in Year 2011 and After



3. Achievements in the Projects of Each Ministry and the Complementary Subject in This Fiscal Year

Information Grand Voyage Project

Ministry of Economy, Trade and Industry

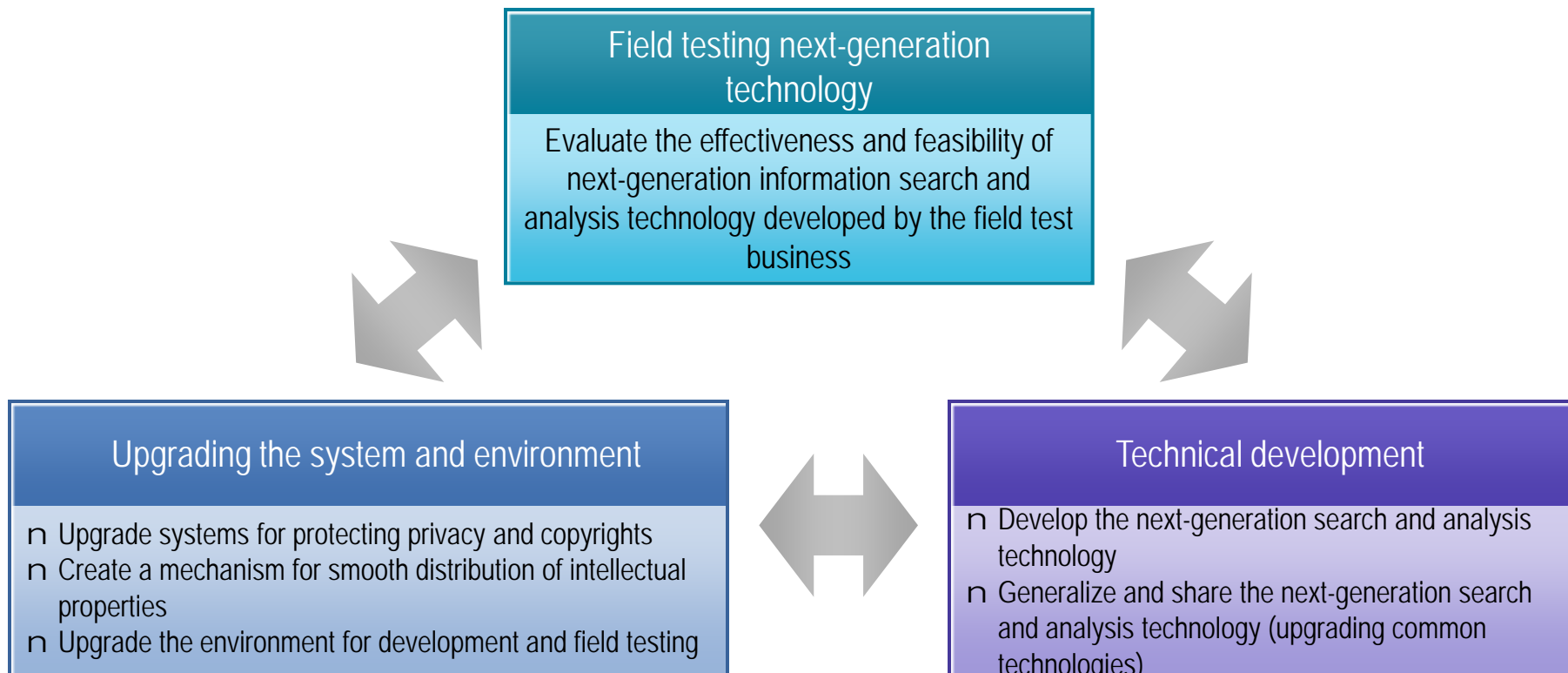
<http://www.igvpj.jp/index/>

http://www.igvpj.jp/index_en/

Information Grand Voyage Project

Purpose

To promote further maturity of the "institution and environment" and further growth of "technology" with the success of the next-generation technology-related business as a trigger, and to establish an innovation creation mechanism leading to the creation of business, including more advanced public activity, in order to activate industry and strengthen competitiveness



Major Achievements: Next Generation Field-test Business

The field testing of next-generation technology

Ten field-tests were performed from the following three directions, and ideas for creating innovations were obtained.

A. Future-oriented personal service considering privacy

Provide commodities, services, and information suited for the lifestyle of individual users by safely collecting and accumulating various information, such as user profiles and action logs, while considering their privacy, and matching various services and contents.

By safely collecting and analyzing personal information (profiles, action histories, and preferences), user convenience and information-collecting capabilities can be improved and new services can be created.

B. Next-generation web services that generate new value

Generate a next-generation web service that allows users to find new content value by developing intuitive access to the contents that could not be obtained with conventional text-based search, along with access to unknown contents by searching images and videos and using interactive and emotion-related searches.

By solving problems caused by ambiguous and abstract user intention during information searches and by utilizing rich content, the improvement and expansion of market value can be expected.

C. IT services of the new social infrastructure

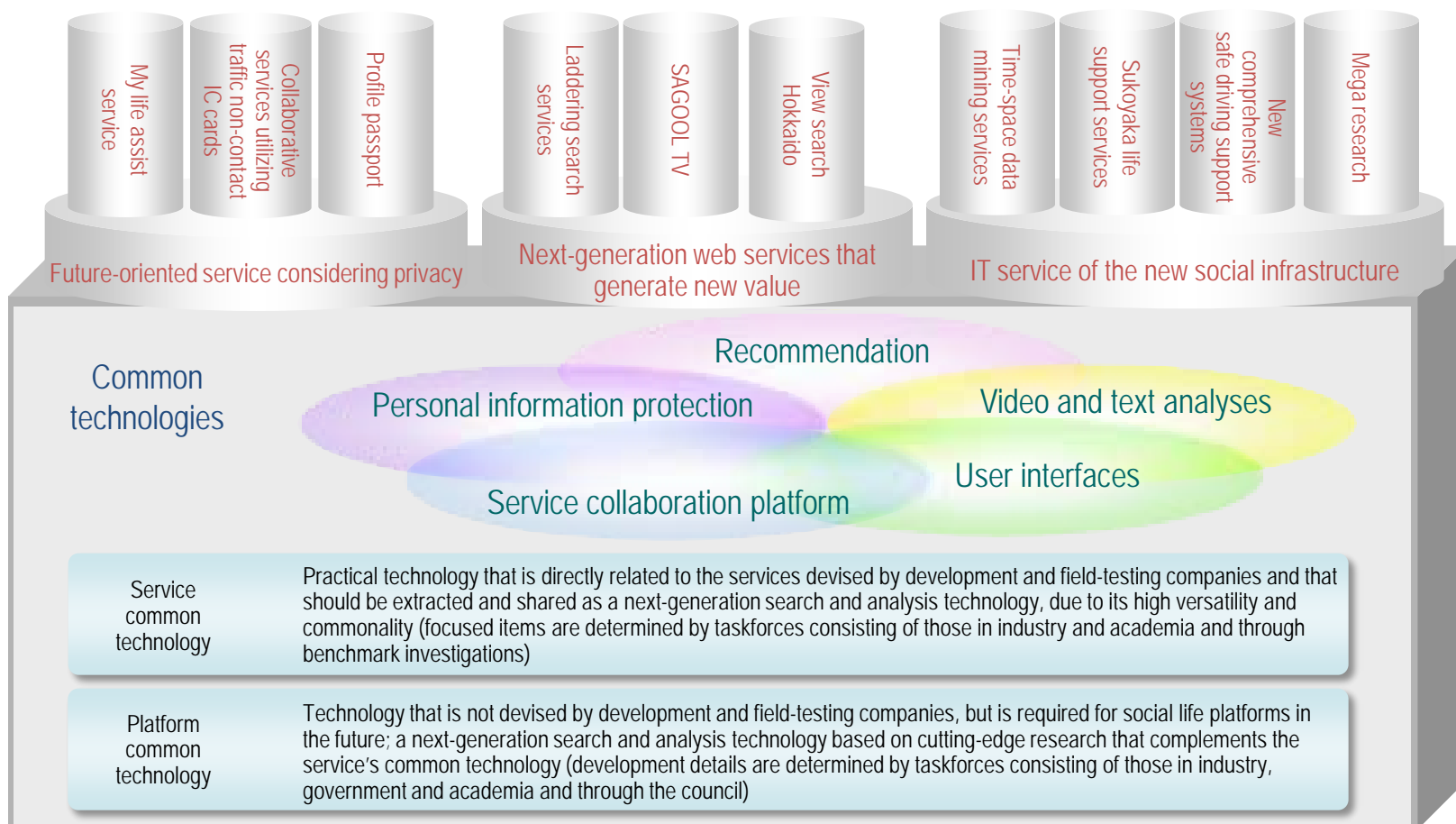
Make high level use of IT, such as countermeasures against enterprise risks in various social infrastructures and provision of advanced medical services through the utilization of medical information, by analyzing various information, including real-time information in a comprehensive manner, in order to realize a safer and more secure society.

By constructing a mechanism of detecting the possibility of accidents and diseases in advance and by taking appropriate countermeasures against them through the utilization of next-generation search and analysis technology, various risks can be avoided or reduced.

Major Achievements: Common Technologies

Common technologies

In collaboration with taskforces consisting of those of industry, government and academia, focused items and development details were formulated. As next-generation search and analysis technologies with high versatility and commonality, **practical technologies that are directly related to services** and **cutting-edge basic technologies** that complement them were developed.



Future Plans

Field testing of next-generation technology

- rn In the field-test business, evaluate the possibility of the next-generation services utilizing the collaboration platform and promote the creation of new common technologies at which collaboration is begun.

Common technologies

- rn Formulate a common technology development road map in each field, an international standardization action plan, and formulate a revised version of the common technology architecture.
- rn For collaboration platform, review the operation rule for application and formulate more convenient measures. Upgrade the smooth application environment with the improvement of CP.

Systems

- rn Develop linkage between the technologies being developed, such as personal information anonymization technology, and the proposals of systems and measures. With this linkage, formulate personal information and recommendation guidelines.
- rn Make case studies concerning intellectual property processing and prepare an upgrading plan of the intellectual property processing organizations. In addition, prepare an action plan for the operation of an intellectual property bank.

Research and Development on Information Credibility Evaluation Technology, etc. on Telecommunication Services

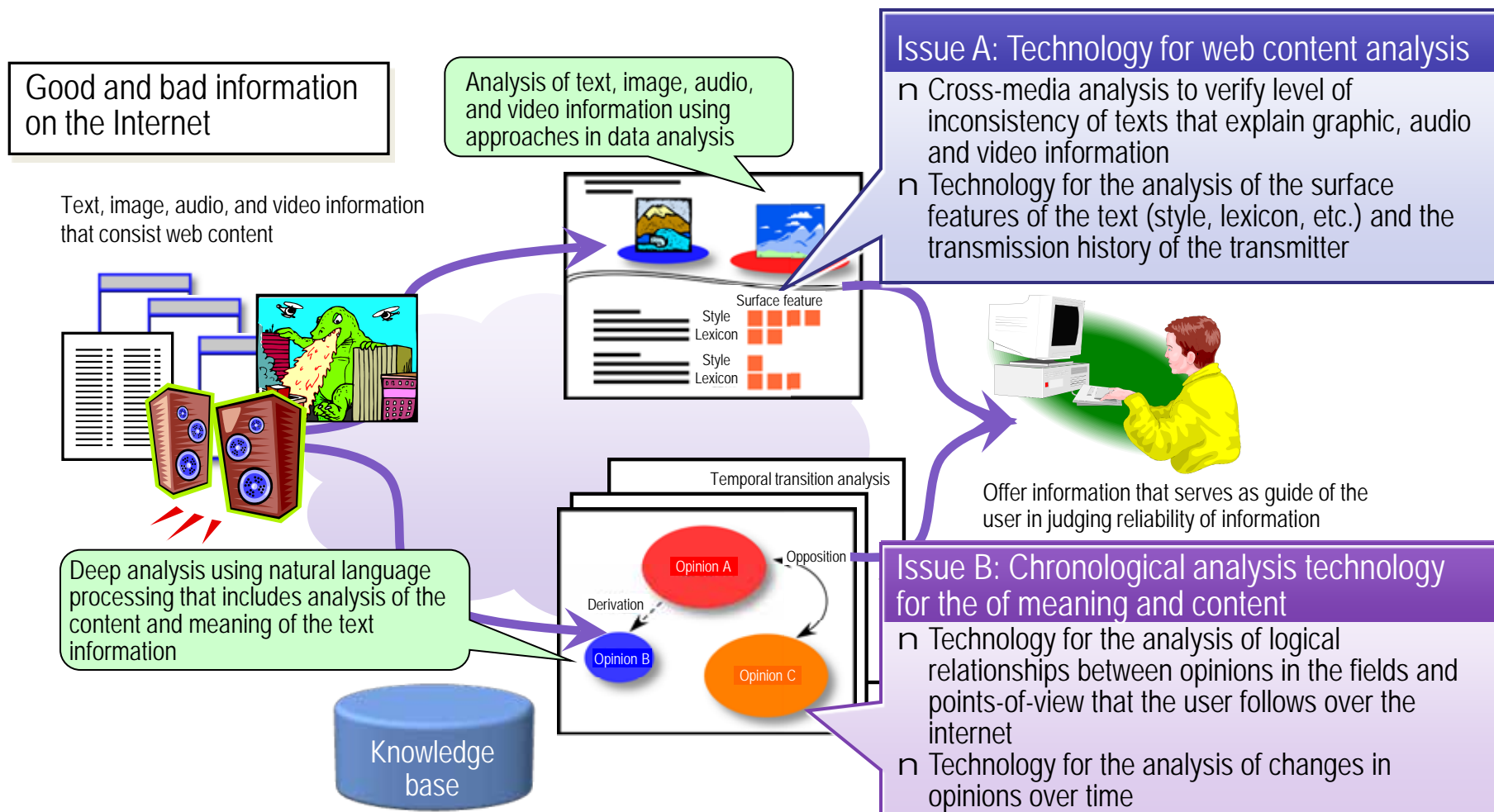
Ministry of Internal Affairs and Communications

<http://kc.nict.go.jp/project1/>

Information Credibility Evaluation Technology

Information credibility evaluation technology

Develop technologies that enable users to use of web content to their benefit easily by **analyzing and presenting** good and bad information that flows through the Internet based on the user's judgment criteria.



Major Achievements: "Technology for Web Content Analysis"

Technology for web content analysis

- 1) Reliability evaluation technology of multimedia information on web
- 2) Reliability evaluation technology of web textual information

The above items were developed, and basic technologies, such as reliability evaluation technology by the analysis of textual information, collected from peripheral contents and other information sources, and the image feature quantity analysis were evaluated.

1) Reliability evaluation technology of multimedia information on the web



Analysis of the typicality, excess, and deficiency, and consistency of web images



Analysis of video and audio biases and analysis of audience evaluation information



Analysis of the percentages of incorrect or missing descriptions in electronic maps using web archives

2) Reliability evaluation technology of web textual information



Analysis of the popularity of the contents, topic coverage, and social acceptability of the searched page

Analysis of the transmitter sentiment and familiarity

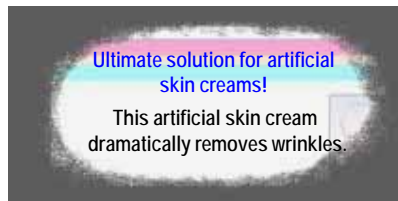


Major Achievements: "Technology for the Chronological Analysis of Meaning and Content"

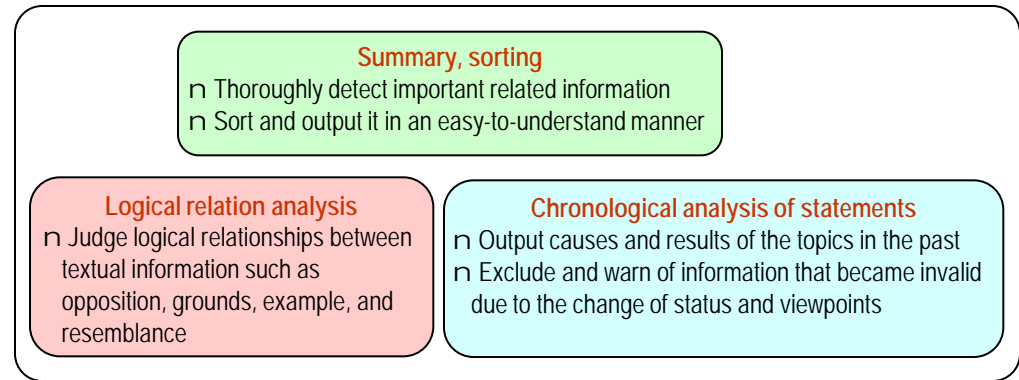
Technology for the chronological analysis of meaning and content

We embarked on research and development of basic components towards the construction of **an integrated system that supports the user's panoramic reliability judgment by exhaustively presenting logically and chronologically related information** to the noted textual information.

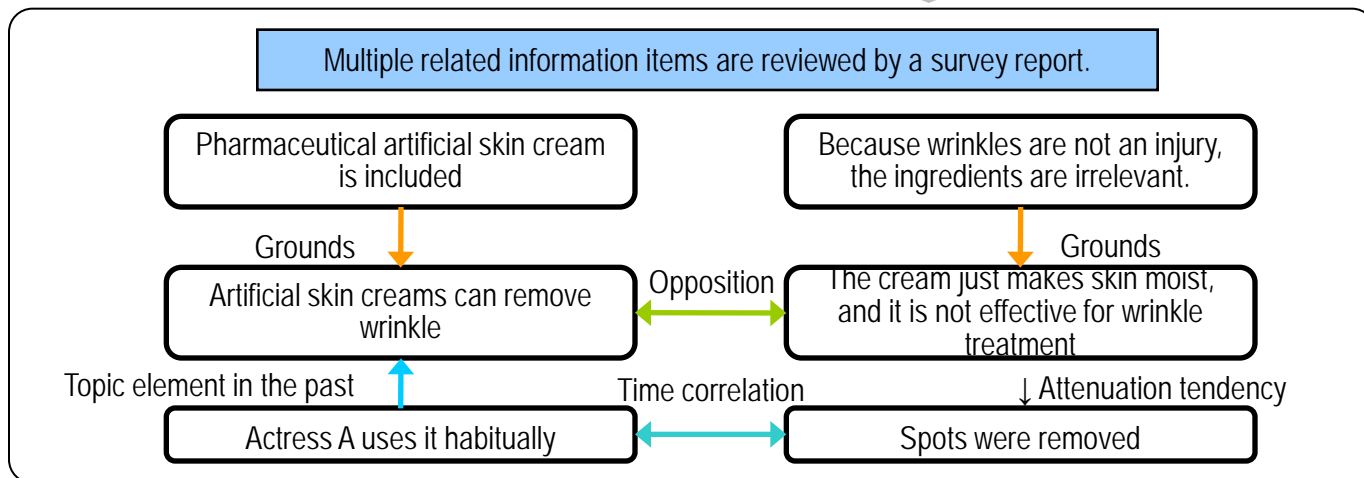
State of narrowed visual field that views only convenient information



Noted textual information



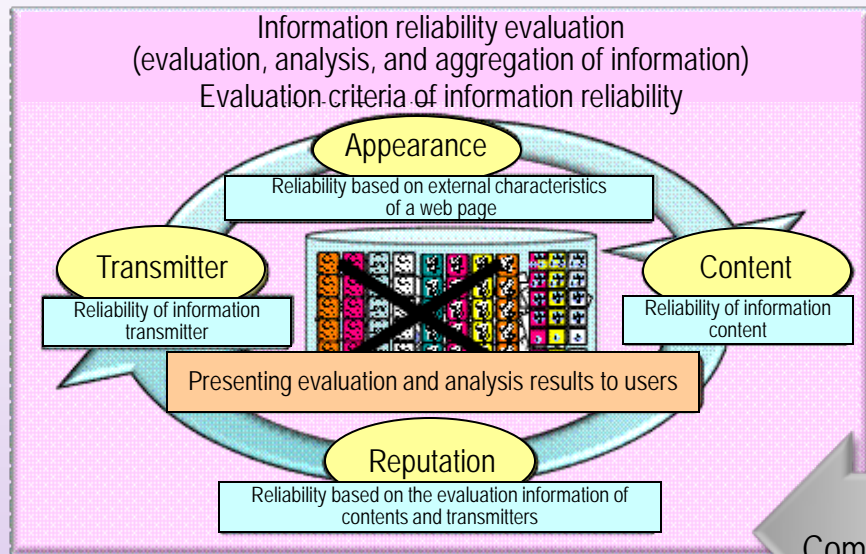
Survey report



Future plans: Constructing an Effective Collaboration System

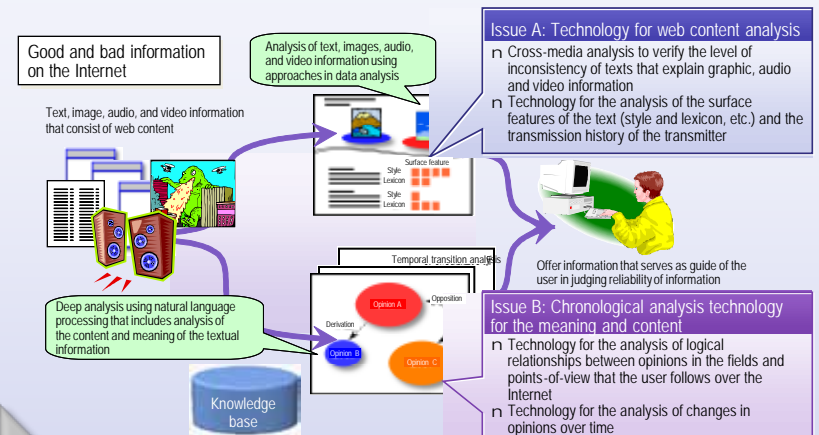
Establish a system for collaboration between the research conducted by the National Institute of Information and Communications Technology (NICT) itself and the research assigned by NICT, and then strongly promote the Information Credibility Evaluation Technology to allow anyone to use web content in a safe and secure manner with the utilization of the basic technology of the Information Grand Voyage brought into view.

NICT research



- From a large amount of web content, information with which users can judge reliability is provided with four types of information analysis methods, based on natural language processing technology.

Assigned research



- Analysis technology of text, images, audio, and video constituting web contents using the data analysis approach
- Analysis technology of logical relationships among information items using deep analysis with natural language processing and temporal transitions

Mutual Complementarity

Development of Ultra-High Performance Database Engine Software Based on the Innovational Execution Principle

Ministry of Education, Culture, Sports, Science and Technology

<http://cif.iis.u-tokyo.ac.jp/OoODE/#>

Ultra-High Performance Database Engine Based on the Out of Order Execution Principle

Purpose

Develop “ultra-high performance database engine software” that enables “strategic application of super large-volume data”, in the age of the information explosion, such as sensor network information analysis and distribution traceability systems.

Issues

1. The “strategic application of super large-volume data” is a key technology as a source of national power that ensures social platforms supporting the safety and security of the nation and enables the creation of various new industries.
2. The “strategic application of super large-volume data” in the age of the information explosion cannot be implemented just by improving existing database platforms.

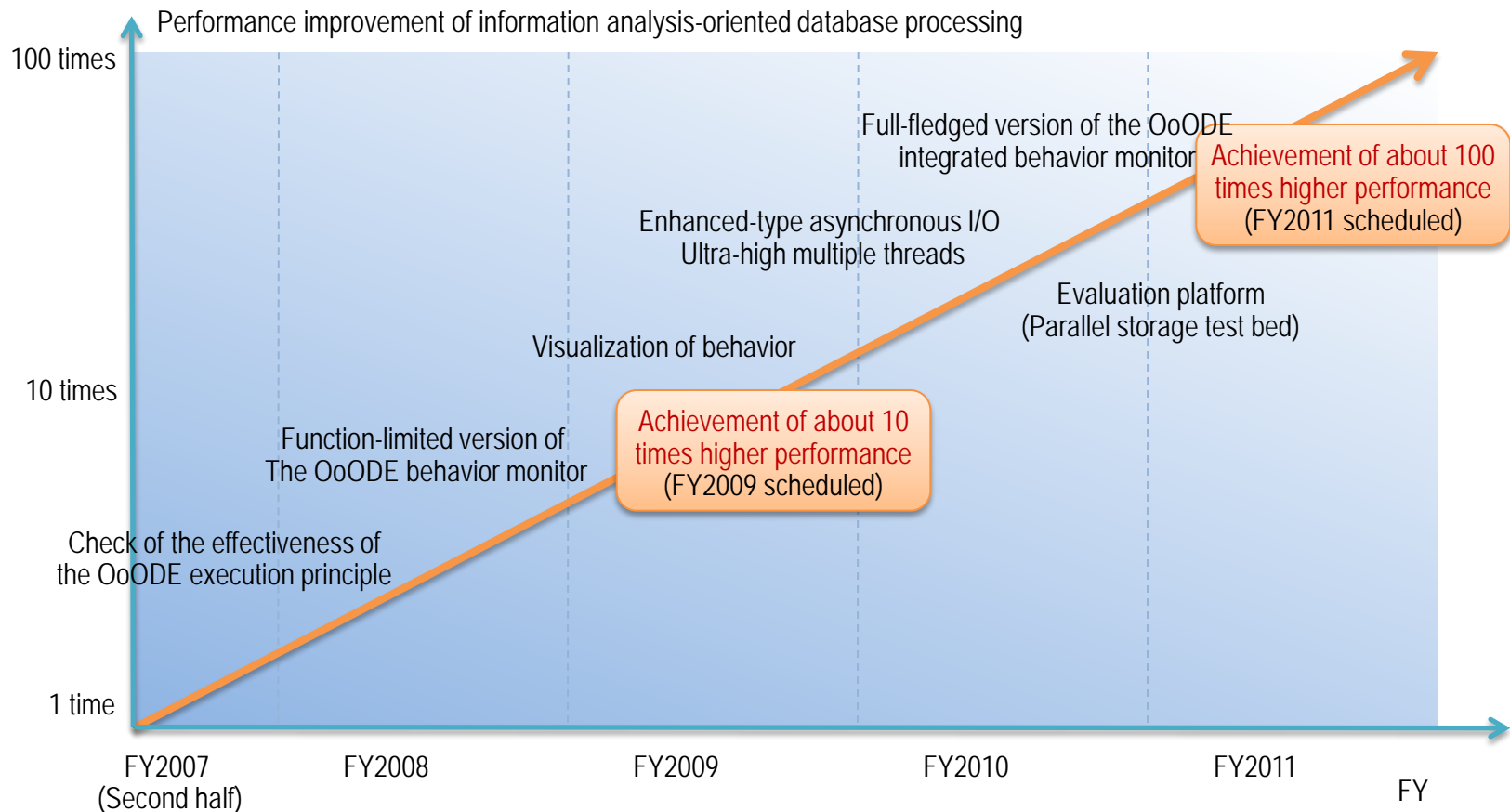
Performance breakthroughs with the innovation of the execution principle is required.

Content of the research and development

- n In order to develop an ultra-high database engine based on the out-of-order execution principle, conduct the following research and development.
- (1) Establish an innovative execution principle, the “out-of-order database execution principle”.
 - (2) Design and implement database platform software based on (1) above.
 - (3) Verify the effectiveness through actual application use.

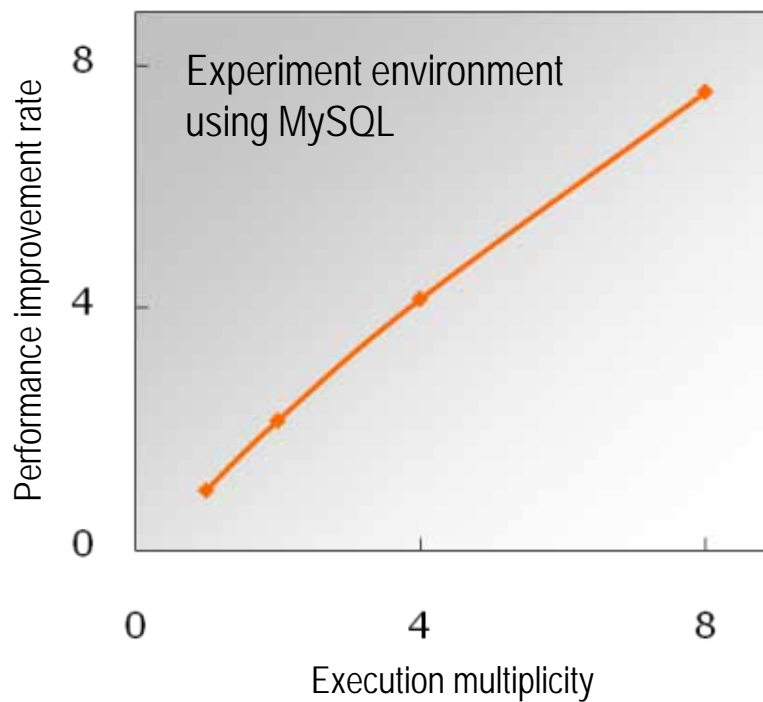
Achievement Goals

- By the end of FY2009, develop a database engine that applies the out-of-order execution principle to some database operations to achieve a tenfold performance improvement of information analysis-oriented database processing, compared to existing technology.
- By the end of FY2011, develop a database engine that applies the out-of-order execution principle in earnest to achieve a hundredfold performance improvement of information analysis-oriented database processing, compared to existing technology.

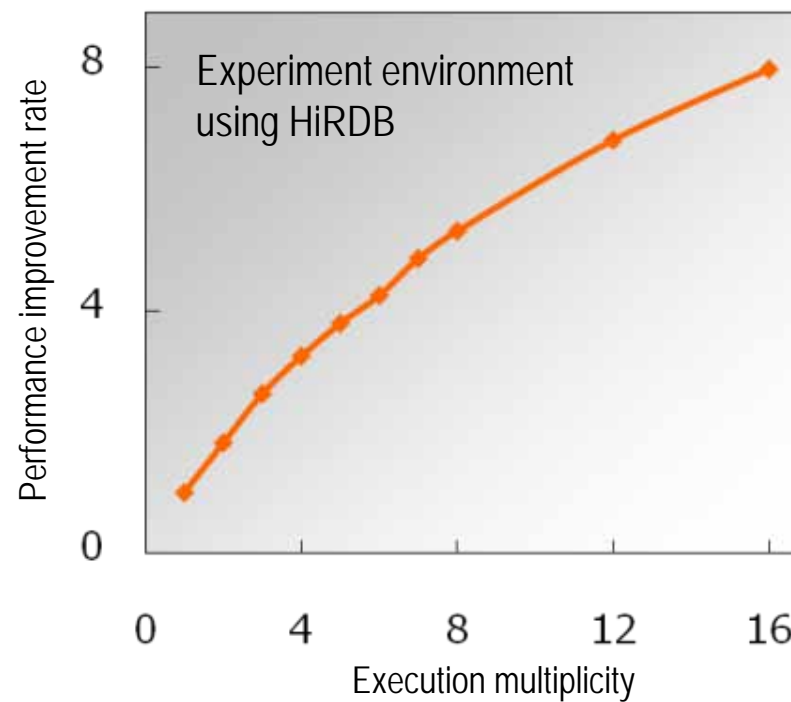


Major Achievements: Experiment Results

- n Small-scale experiments were conducted continuously by using an existing database system, and steady progress was made towards tenfold performance improvement, which is the interim goal in FY2009.
 - For the open source database system (left figure), an approximate eightfold performance improvement was verified for execution multiplicity 8.
 - For the commercial database system (right figure), an approximate eightfold performance improvement was verified for execution multiplicity 16.



TPC-H data set (SF=4.0; 4.0GB) Q.3
Itanium2 1.67GHz * 8, 30 HDDs, RedHat Linux AS3

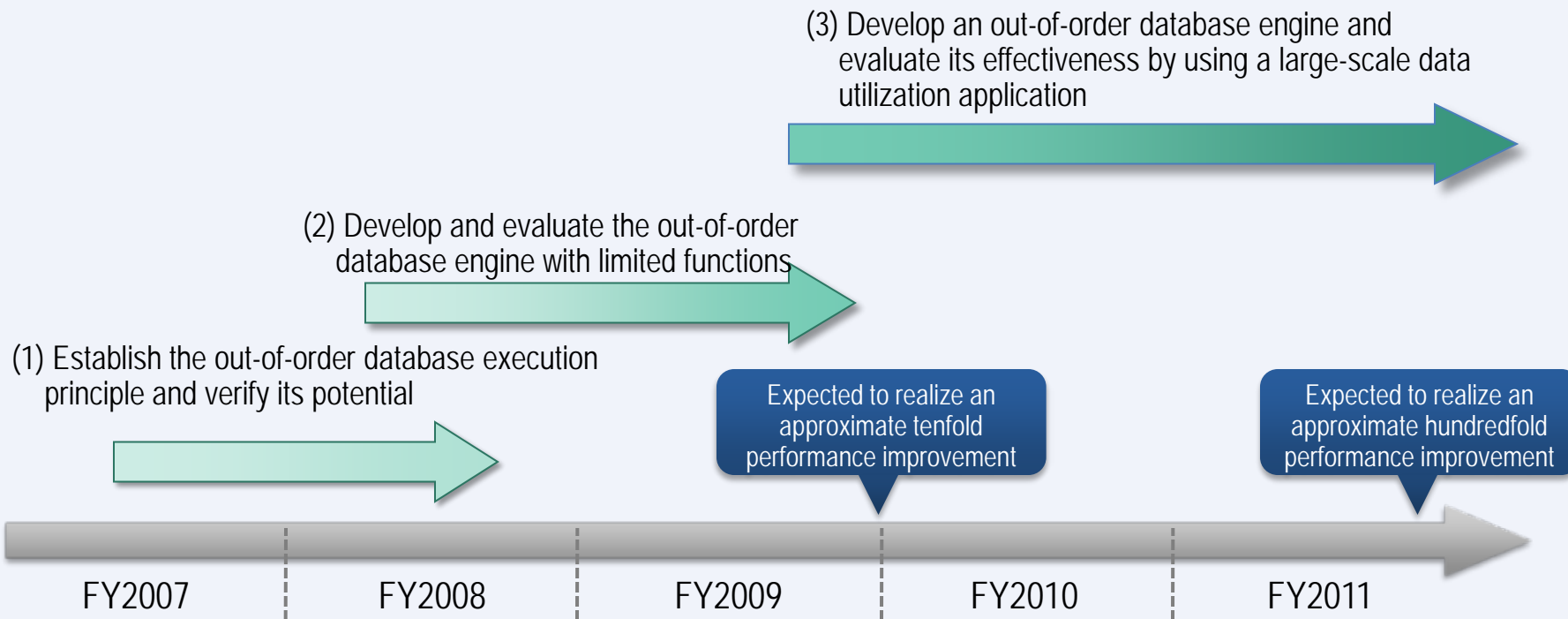


TPC-H data set (SF=30; 30GB) Q.8
Xeon 2.40GHz, 1GB RAM, 10 HDDs, Fedora Core 4

Future Plans: Towards Performance Improvement

Development and evaluation of an out-of-order database engine with limited functions

- n Design and mount the out-of-order database engine with limited functions
- n Evaluate the performance improvement of the out-of-order database through a small-scale experiment
- n Design the I/O control mechanism that enables ultra-high multiple asynchronous I/O processing
- n Design the integrated monitoring mechanism of the database engine, OS, and storage
- n Design the field-testing evaluation platform system and begin its construction



The Sensing Web: Content Extraction of Sensor Information for Societal Use

Complementary Subject
(Representative Institution: Kyoto University)

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

Background and Purpose of Research

Background

- n The Internet is developing on a global scale and is becoming an enormous knowledge base.
- n Sensor networks are closed on purpose and are not widely utilized.
- n Research to develop sensor information into contents that can be used in a society is essential.

Purpose

Research and develop the following functions to realize sensor information applications (sensing web).

- n Mechanism for sharing sensor information (information sharing)
- n Responding to the problems unique to sensor information (access management)
- n Search and presentation of distributed sensor information (information application)

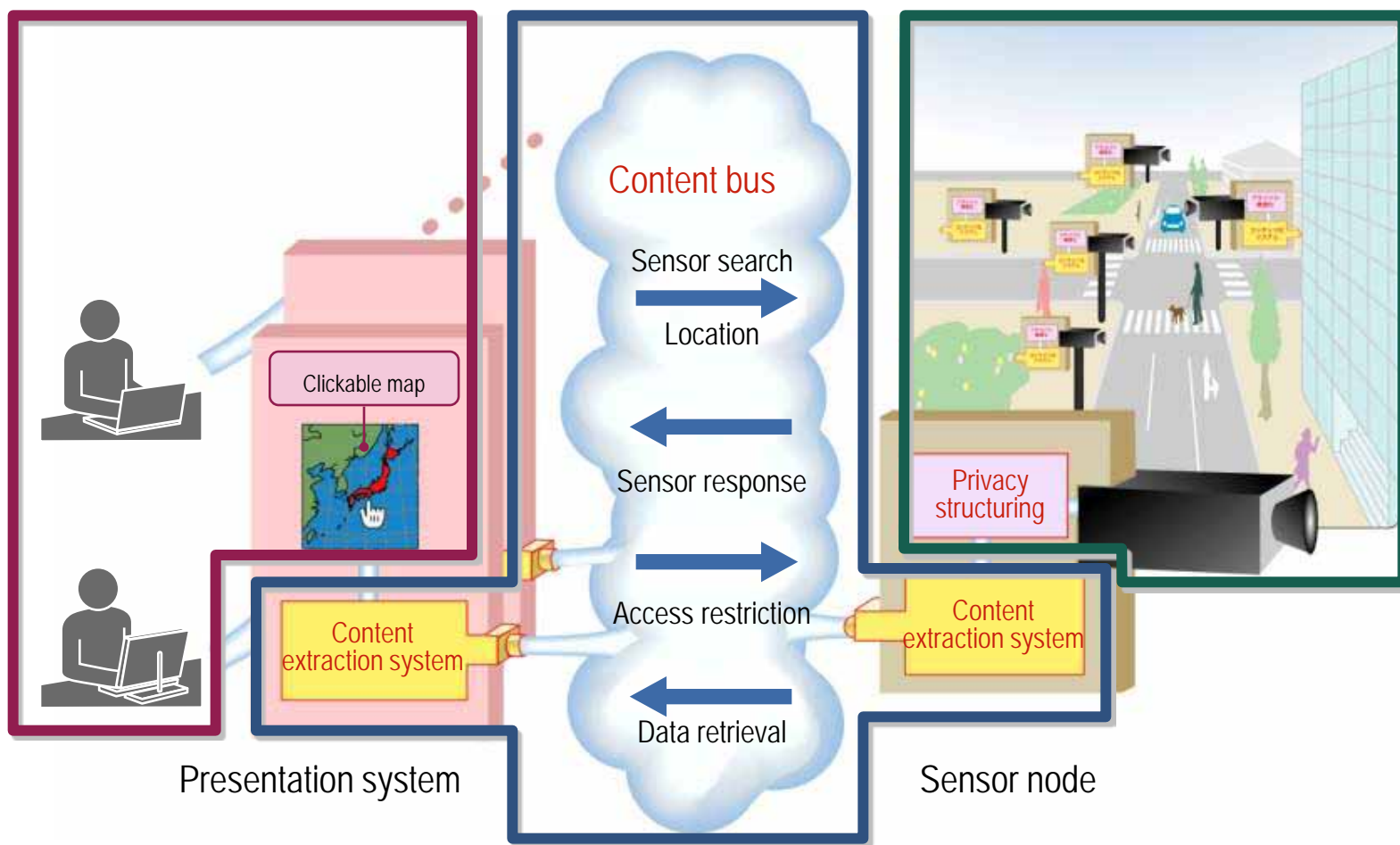
Content of the Research and Development

Sensing Web

3 Observation-type real world content presentation technology (information application)

2 Technology for content extraction of sensor information (information sharing)

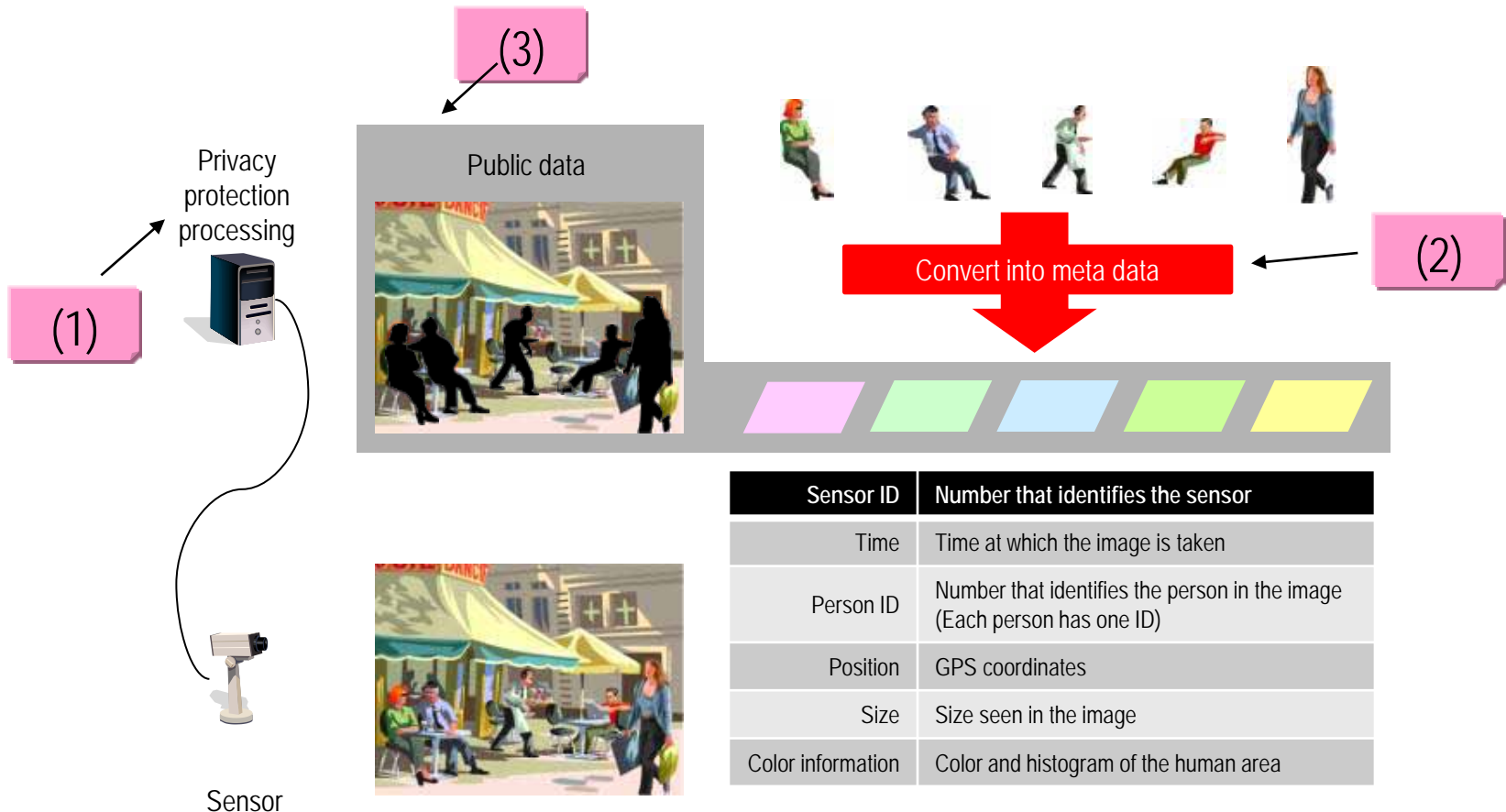
1 Technology for privacy information management (access management)



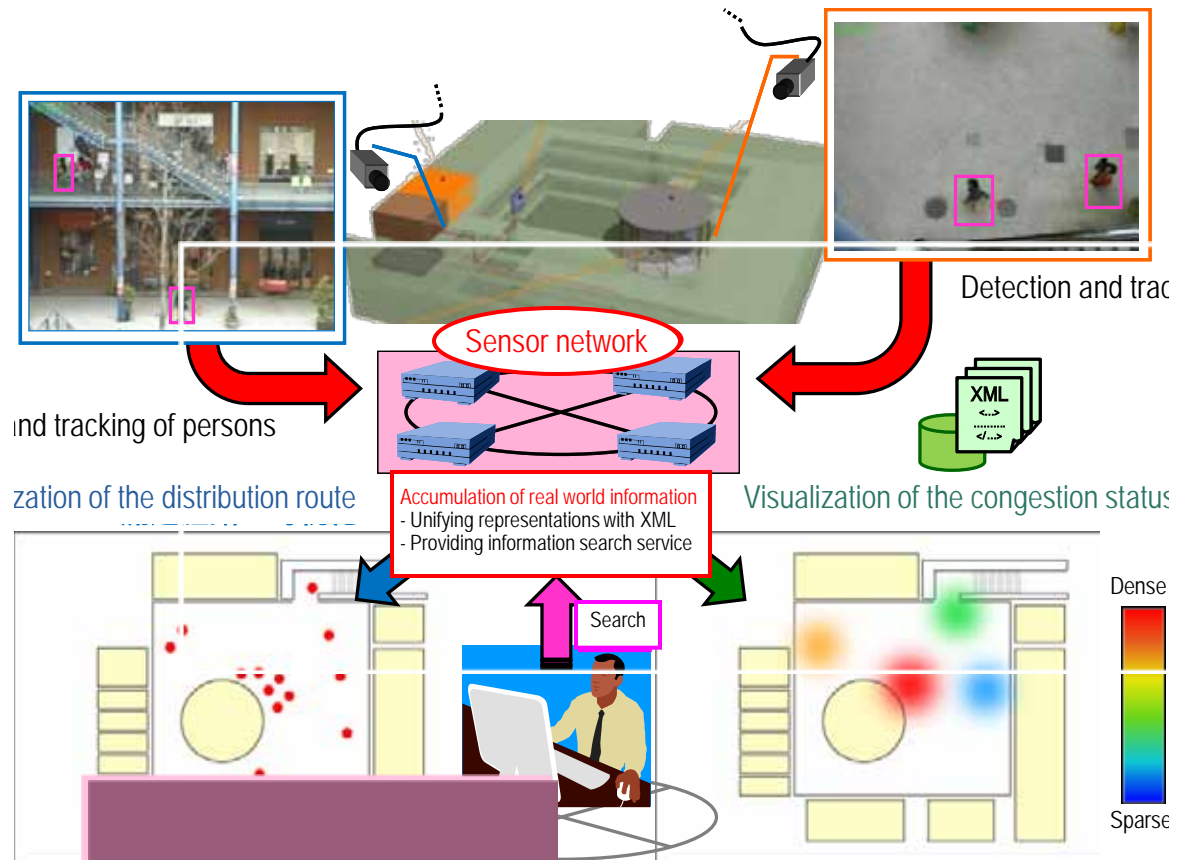
Major Achievements: Sensing Web Technical Development

The basic functions of web-sensing technology was developed.

- (1) Technology for removing the private data of observed persons in sensor information before releasing it publicly
- (2) Technology for accepting various information demands regardless of the sensor type and installation status
- (3) Technology for integrating data provided from sensors and presenting it in an easy-to-understand manner



Future Plans: Field-test Experiment



Summary of the Projects of Each Ministry and the Complementary Subject: List of Major Achievements

List of Major Achievements

	List of major achievements	Information transmission
Information Grand Voyage	<ul style="list-style-type: none"> n Next-generation field-test experiments: Conducted for 10 services n Common technologies: 55 basic technologies, including video and text analysis technologies, were developed. (http://www.igvpj.jp/index/cp-info/000/) n System: Guidelines were prepared (personal information protection and copyright, etc.) 	<ul style="list-style-type: none"> n Exhibition in Japan, Symposium n International exhibition n Media coverage: 34 articles
Information Credibility Evaluation Technology	<ul style="list-style-type: none"> n The reliability of web information (images, movies, audio content, and text were developed): A field-test experiment was conducted. n Analysis of the degree of reliability of web transmitters (speech map creation systems, summarization and sorting of textual information, and development of chronological analysis) 	<ul style="list-style-type: none"> n Symposium in Japan n Hosting international conference n Media coverage
Ultra-High Performance Database Engine	<ul style="list-style-type: none"> n A small-scale experiment was conducted for the out-of-order execution principle and an approximate eightfold performance improvement was verified (in the real world, such as an OSS and commercially available DB) <ul style="list-style-type: none"> - A tenfold improvement is the interim goal in FY2009. 	<ul style="list-style-type: none"> n Symposium in Japan
Sensing Web	<ul style="list-style-type: none"> n A sensor information description method was developed. <ul style="list-style-type: none"> - Structuring of privacy, buffering processing n Method of publicizing personal tracking data 	<ul style="list-style-type: none"> n Symposium in Japan n Lecture at an international conference n Introduced in the feature pages in an academic journal in Japan

Examples of Information Transmission: Exhibition in Japan, International Lectures (Information Grand Voyage)

- Following last year, we participated in the Cutting-Edge IT & Electronics Comprehensive Exhibition (CEATEC) JAPAN 2008, which is drawing large attention from a wide range of users this year once again. The exhibition was highly successful and was frequently picked up by the media.
- In ICT 2008, the ICT-related major event in the EU, an invitational lecture (with Professor Kitsuregawa) was held. As seen in the fact that there were many requests for his lectures in addition to those scheduled in advance, interest in activity in Japan was very high.

Exhibition at CEATEC JAPAN 2008

- Date and place: September 30 (Tue.) to October 4 (Sat.) at Makuhari Messe
- Outline:

 - Exhibition booths (demonstration by 11 organizations including NTT docomo)
 - Collaboration corner (explanation of field-test businesses)
 - Main stage (explanation of overview)
 - Keynote speech (lectures by Professor Kitsuregawa, audience of 500)
 - Seminar (panel discussions concerning systems and intellectual properties: audience of 120)
- Number of visitors to the booth: A total of about 17,000 persons (over five days)
There were many inquiries concerning applications of the developed common technologies and collaboration from many institutes and companies, including that from a commercial officer from the Ambassador of Australia.
- Media coverage: *Nikkei BP, IT media, Toyokeizai, Nikkan Kogyo Shimbun, Denpa Shimbun, etc.*



Lecture at EU-IUT2008

- Date and place: November 25 (Tue.) to November 27 (Thu.) in Lyon, France
- Rostrum lectures:

 - (1) Lectures at the CHORUS meeting
 - (2) Lectures concerning "Prospects in Disruptive Innovation in Multimedia Search Engines"
 - (3) Lectures concerning "Networked Media & 3D Internet"
- Connection:

 - (1) The head of unit in the EU Committee expressed his thanks to the above lectures.
 - (2) Continuous information exchange was had with NESSI (Networked European Software and Services Initiative)
 - (3) Visited Exalead, a participant company in QUAERO



Examples of Information Transmission: International Conference, Academic Activity (Information Reliability, Sensing Web)

- International workshop (WICOW 2008: Second Workshop on Information Credibility on the Web, Napa Valley, California, October 30, 2008) was planned and held. In this workshop, information reliability technology was introduced and our leadership was shown.
- In the ICPR (International Conference on Pattern Recognition, which was held in 2008 in Tampa, Florida) which is the largest international conference in the field of pattern recognition, a workshop was founded, and our leadership in web-sensing technology was shown.
- In the Journal of the Japanese Society for Artificial Intelligence, which is an interdisciplinary academic journal covering the humanities and social science, a special issue featuring the “sensing wave” was published (March 2009 issue).



<http://www.dl.kuis.kyoto-u.ac.jp/wicow2>



(2) ICPR 2008 workshop

“Feature ‘Sensing Web’”: Noboru Babaguchi (Osaka University), Michihiko Mino (Kyoto University)
 “Sensing web – its concept and challenges”: Michihiko Mino (Kyoto University)
 “Importance of sensing web technology in the coordination program”: Shojiro Nishio (Osaka University)
 “Approach to privacy protection in a camera image” Michihiko Mino (Kyoto University) et al.
 “Protection of privacy information included in a voice”: Seiichi Nakagawa (Toyohashi University of Technology) et al.
 “Architecture for sensor information sharing and data format standards” Tsuneo Ajisaka (Wakayama University), et al.
 “Structuring of sensor information in sensing web” Rinichiro Taniguchi (Kyushu University), et al.
 “Control of see-through vision and privacy information”: Yuichi Ota (Tsukuba University), et al.
 “Observation-type real world contents: digital diorama” Noboru Babaguchi (Osaka University), et al.

(3) Special issue of the Journal of the Japanese Society for Artificial Intelligence

(1) Planning and holding WICOW2008

4. Cooperation during Each Project

Coordination Diagram between Each Technological Element of "Very Large Information Integration and Application Platform" Program

Research and Development on Information Credibility Evaluation Technology, etc. on Telecommunication Services (MIC)

- Technology for web content analysis
- Technology to analyze graphic, audio, and video information
- Technology to analyze text information
- Technology for the chronological analysis of meaning and content
- Technology for judging credibility of information transmitter
- Technology for judging credibility based on surface features of web page
- Technology for judging credibility based on information content
- Technology for judging credibility based on information from content/submitter

Information Grand Voyage Project (Development of technologies for search and analysis of needed information easily and appropriately) (METI)

- | | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> Collaboration platform Superfast pattern identification technology Recommendation technology • Resemblance evaluation technology on activity and preference • C-SVM recommendation technology • Integrated recommendation technology Graphic and video content understanding technology | <ul style="list-style-type: none"> Technology to analyze wide range of Japanese expressions Narrowing down search results via interaction Position data integration platform technology Personal information anonymization technology Image classification result visualization technology Sensor device gateway technology |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Collaboration

Development of Ultra-High Performance Database Engine Software Based on the Innovational Execution Principle (MEXT)

- Ultra-high performance database engine based on out of order principle
- Research on out of order database engine technology
- Research on resource regulation technology
- Research on monitoring technology
- Research on field-testing and evaluation

Content Extraction of Sensor Information for Societal Use (Complementary Subject)

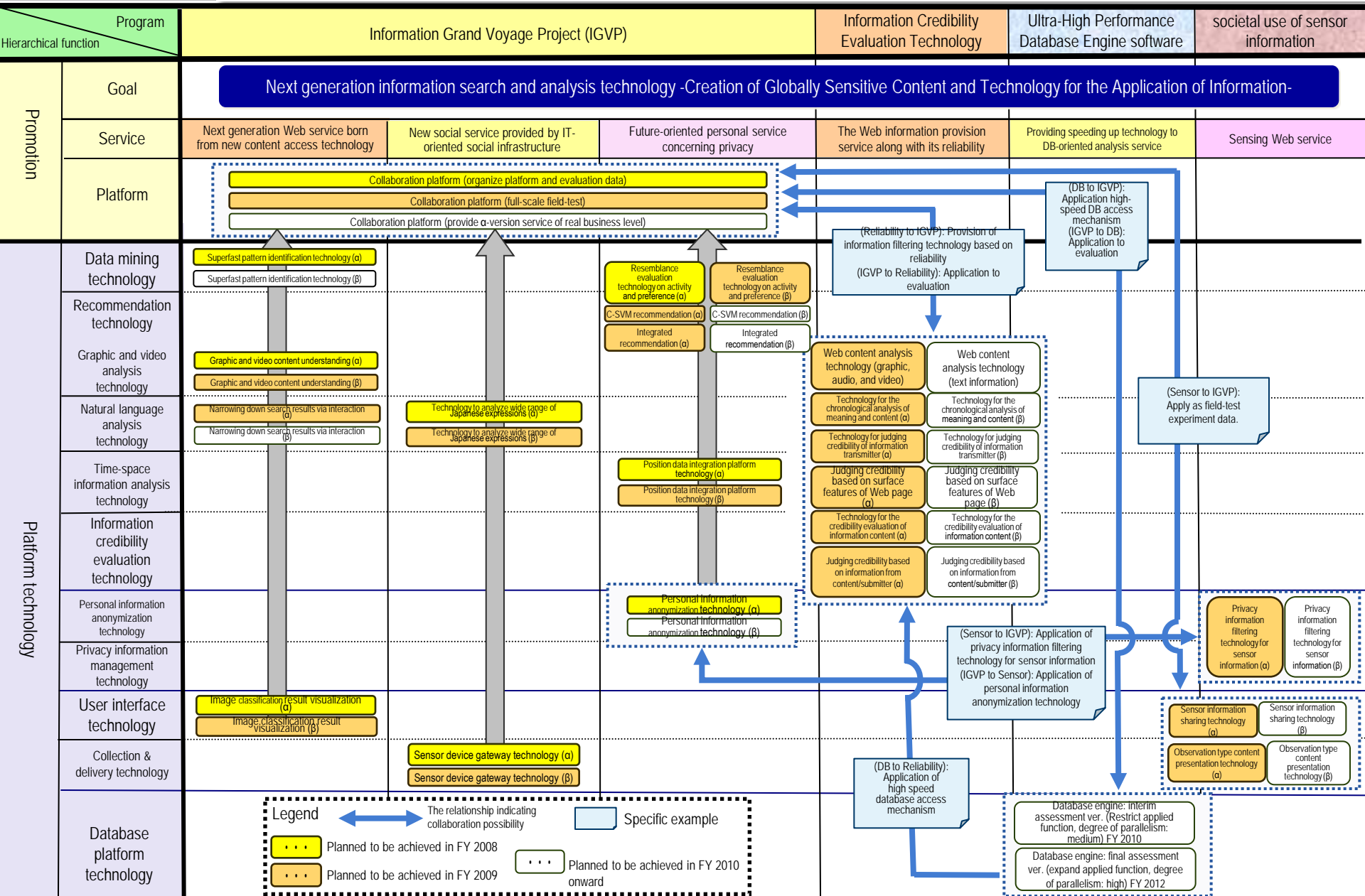
- Technology for privacy information management
- Technology for content extraction of sensor information
- Observation type real world content presentation technology (information application)



Goal

To **build the next generation information search and analysis technology platform** that would enable a person to collect and analyze needed information from diverse information inside and outside the web easily, accurately, and safely. Enable the **strategic application of super large volume data** in this age of information explosion by the year 2012.

Coordination Diagram between Each Technological Element of "Very Large Information Integration and Application Platform" Program (Details)



5. Future Plans

Method for Future Promotion of the Coordination Program

To explore new fields in this age of information, the application of information is required

- ▶ Promote the construction of a next-generation information intellectual information access platform that would enable a person to collect and analyze necessary information from diverse information inside and outside the web easily, accurately, and safely.
- ▶ Investigate and consider various issues for the application of information together with the consideration of the legal system to promote activity that enables effective utilization of the achievements.

It is necessary to promote the active transmission of information, as well as to further strengthen the collaboration system

- ▶ Transmit information by holding a symposium and provide a place for discussion with experts.
- ▶ Further strengthen the collaboration system of related institutions for promoting research and development in the future and further improve international competitiveness.

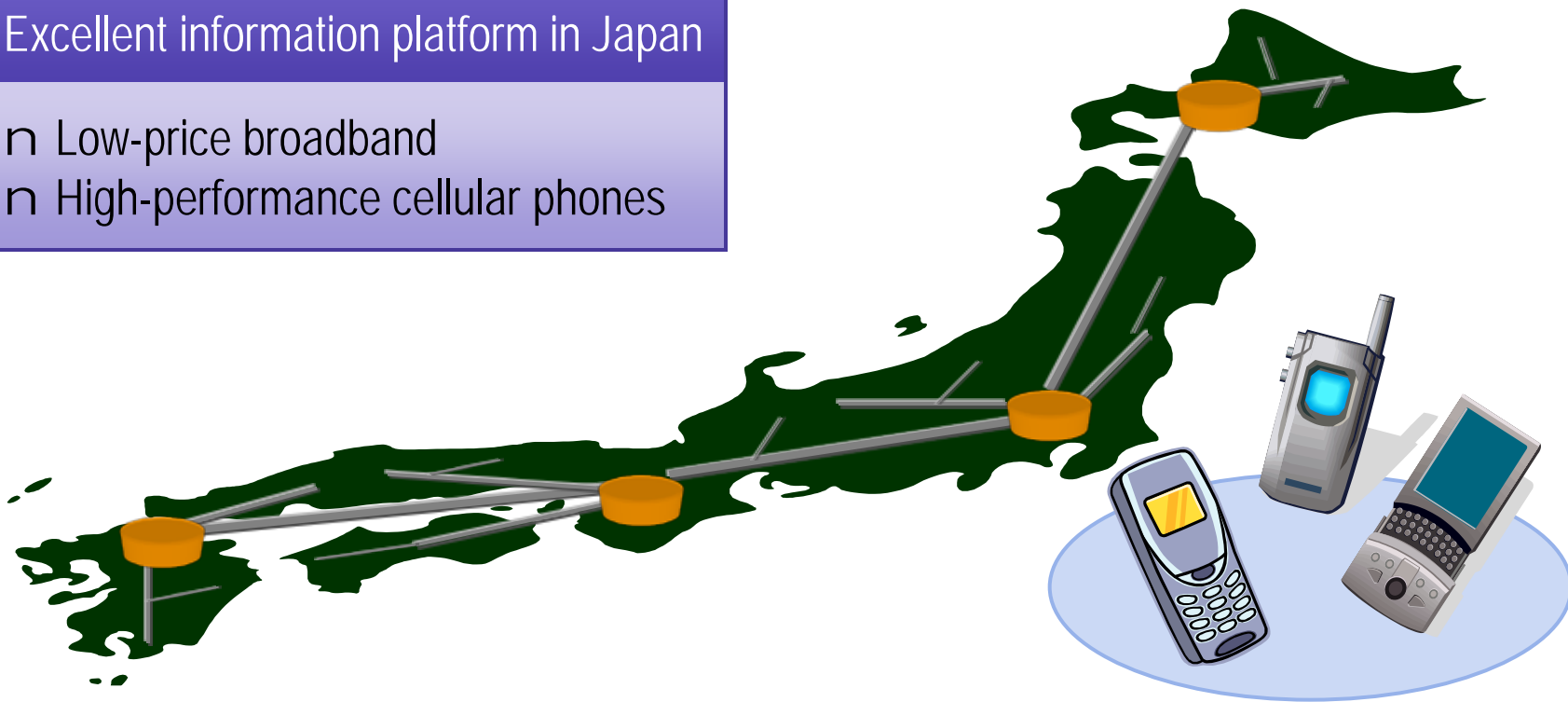
From the "Fountain of Knowledge" to "Weaving Threads of Knowledge"



The Japanese Information Platform: Leading the World

Excellent information platform in Japan

- n Low-price broadband
- n High-performance cellular phones



Japan is fated to exercise leadership towards the construction of a platform for weaving threads of knowledge.