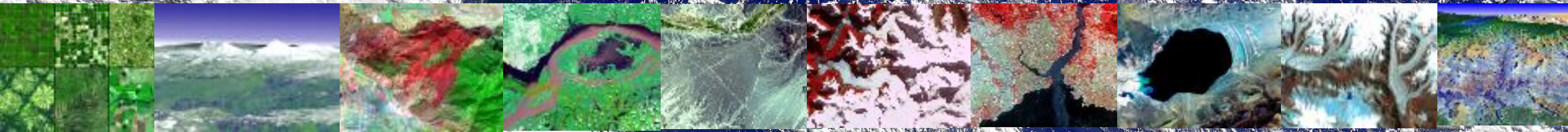
Data Intensive Research Project(s) at ITRI/AIST

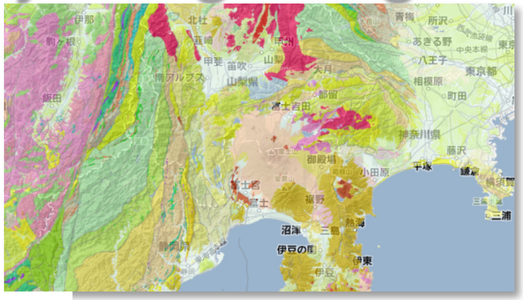
Jason H. Haga
Isao Kojima
National Institute of Advanced Science and Technology(AIST), JAPAN



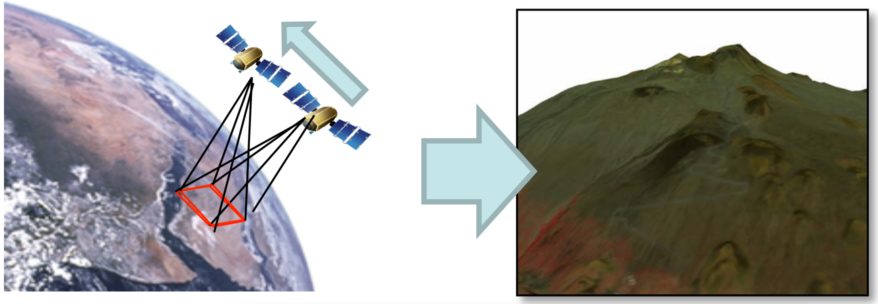
Data Integration Question

(same as last year)

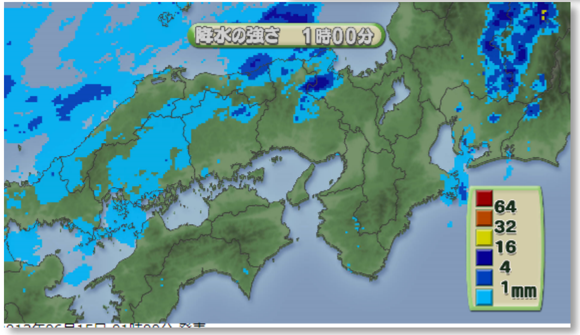
What knowledge can be obtained by integrating following data?



1. Geological Map
   - Geological Survey of Japan is a part of AIST
     - sedimentary rocks,
     - volcano rocks,
     - grantic rocks etc.



2. 3D Elevation Model
   - Created by our ASTER Satellite
     - Produce 3D-model by stereo-matching



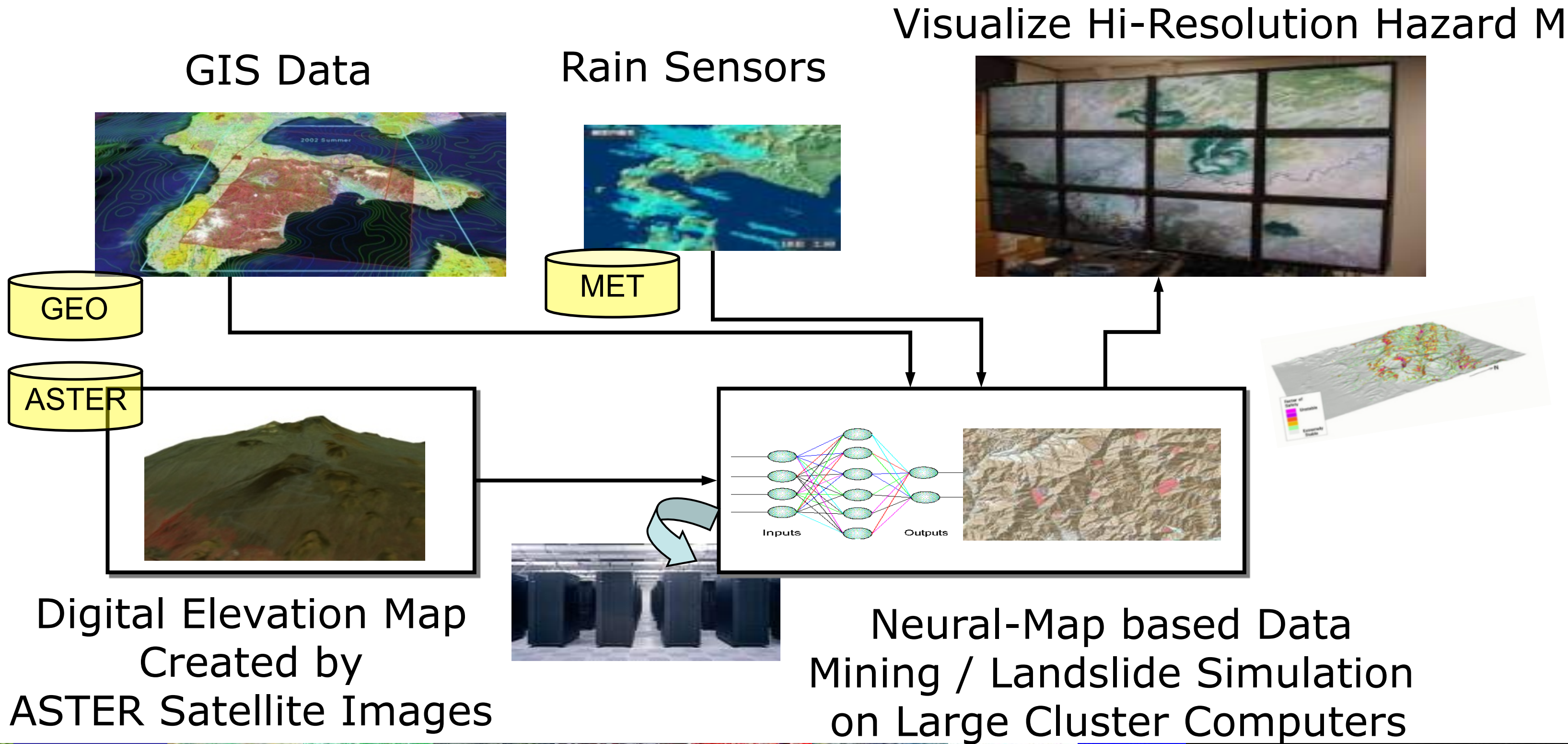
3. Real Time Rain Sensors
   - Provided by JMA(japan meteorological agency)

Answer: Hazard Map for Landslide

(One typical application of GEO Grid)

Key R&D Technologies

1. **Distributed Database Integration (Linked Data/Heterogeneous DB etc.)**
2. **Data Mining & Simulation on the Cloud (Neural-Net, Machine Learning)**
3. **Multi-Screen Visualization (Tiled Wall Software)**



Visualize Hi-Resolution Hazard Map

GIS Data

Rain Sensors

GEO

MET

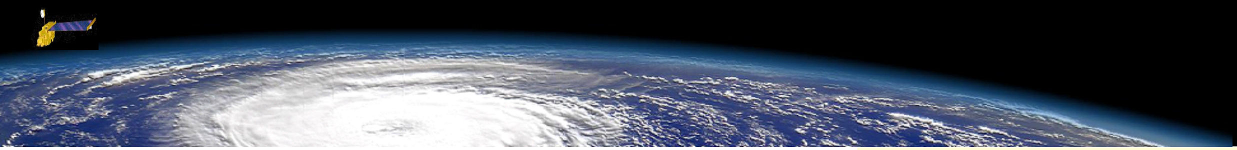
ASTER

Digital Elevation Map
Created by
ASTER Satellite Images

Neural-Map based Data
Mining / Landslide Simulation
on Large Cluster Computers

Background on AIST

- National Institute of **A**dvanced **I**ndustrial **S**cience and **T**echnology, Japan
  - Mission: Contribute to society through continuous advancement in technologies and support to Japanese industries
  - Supported by METI (Ministry of Economy, Trade and Industry)
- Established in 2001
  - Merging **15** different research institutes
    - Oldest is Geological Survey of Japan (est. 1882)
    - Set/maintain the kilogram calibration standard of Japan
- AIST ranked 7th in “Top 20 Japanese research institutions for all field”, Thomson Reuters, 2014



GEO Grid

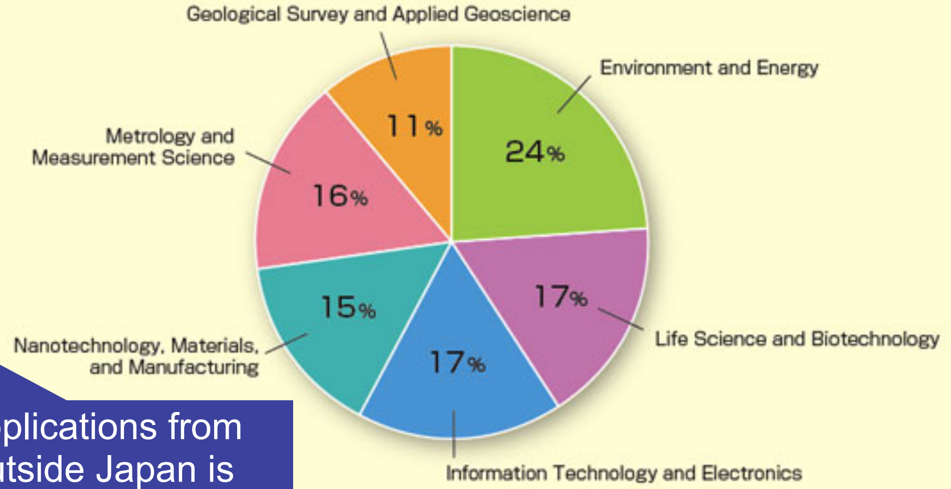
Grid-based e-infrastructure for geosciences

● Researchers (foreign nationals)	2,258 (96)
[Permanent]	[1,928]
[Fixed term]	[330]
● Administrative employees (foreign nationals)	675 (1)
Total number of employees:	2,933 (97)
● Executives (full time)	13
● Visiting researchers	159
● Postdoctoral researchers	200
● Technical staff	1,441

(As of April 1, 2015)

Applications from outside Japan is highly recommended

Composition of researchers by research field (As of April 1, 2012)

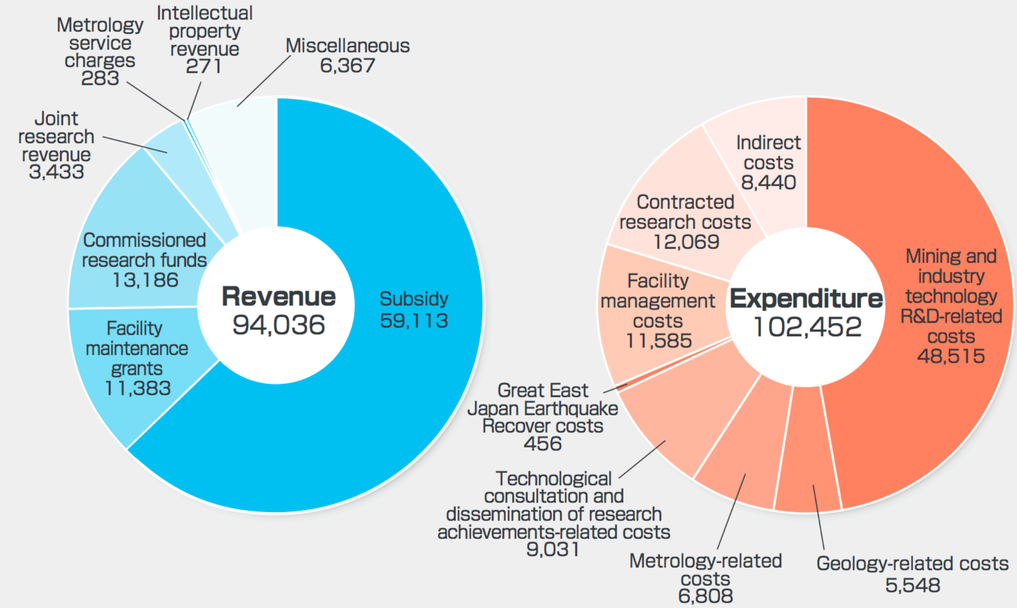


Number of researchers accepted through industry/academia/government partnerships

● Companies	1,774
● Universities	1,852
● Other organizations	972
	(foreign nationals :426)

(Total number of researchers accepted in FY 2013)

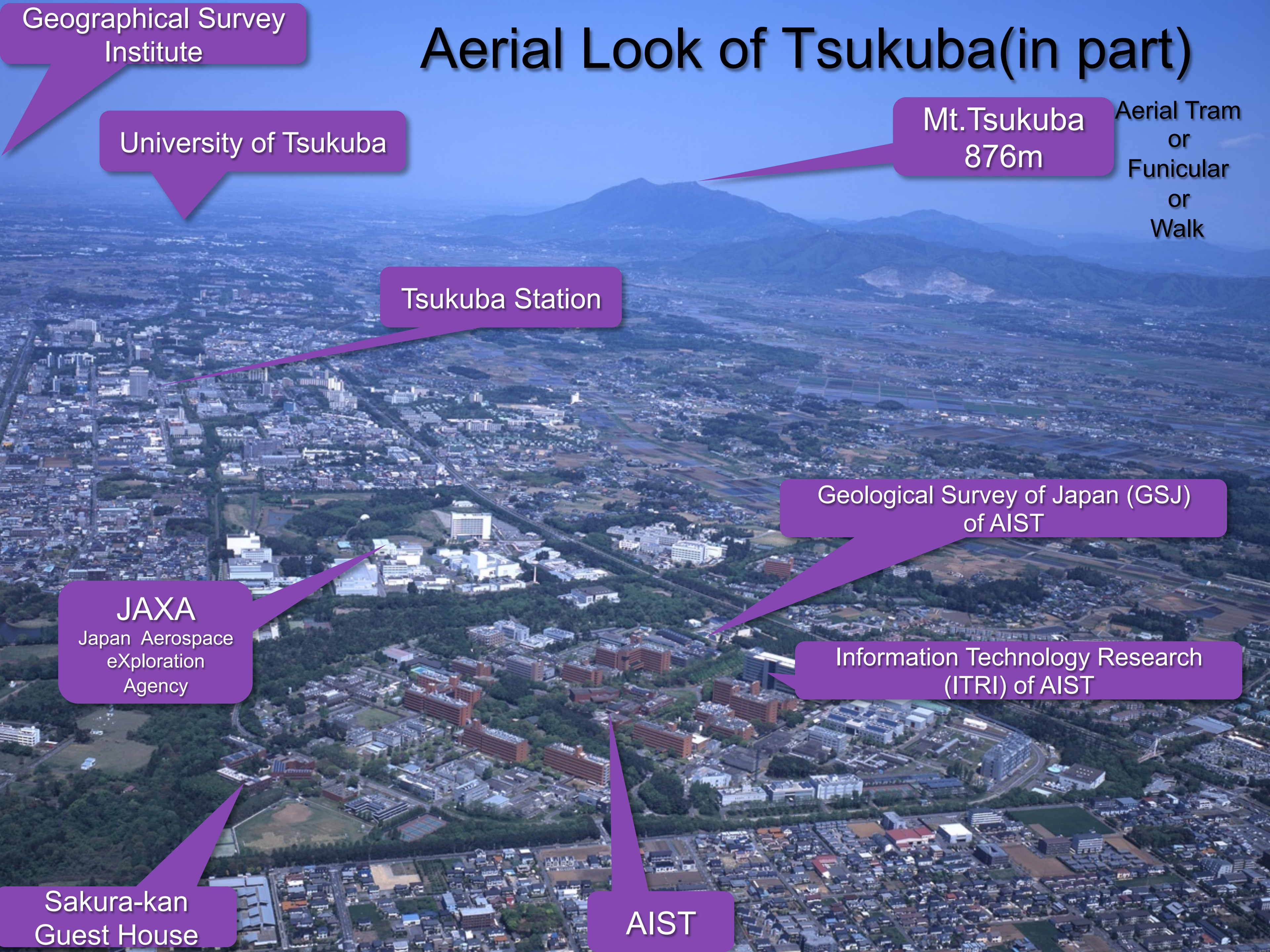
Financial Results for FY 2013 (unit : million yen)



Location of AIST

- Tsukuba (science) City
  - Government planned city
    - Est. in 1962
    - 1 university, 2 colleges
    - About 30 governmental research institutes including JAXA, KEK, NIMS
    - About 30~40 company labs
  - 60km Northeast from Central Tokyo
    - 45min with Tsukuba Express (TX)

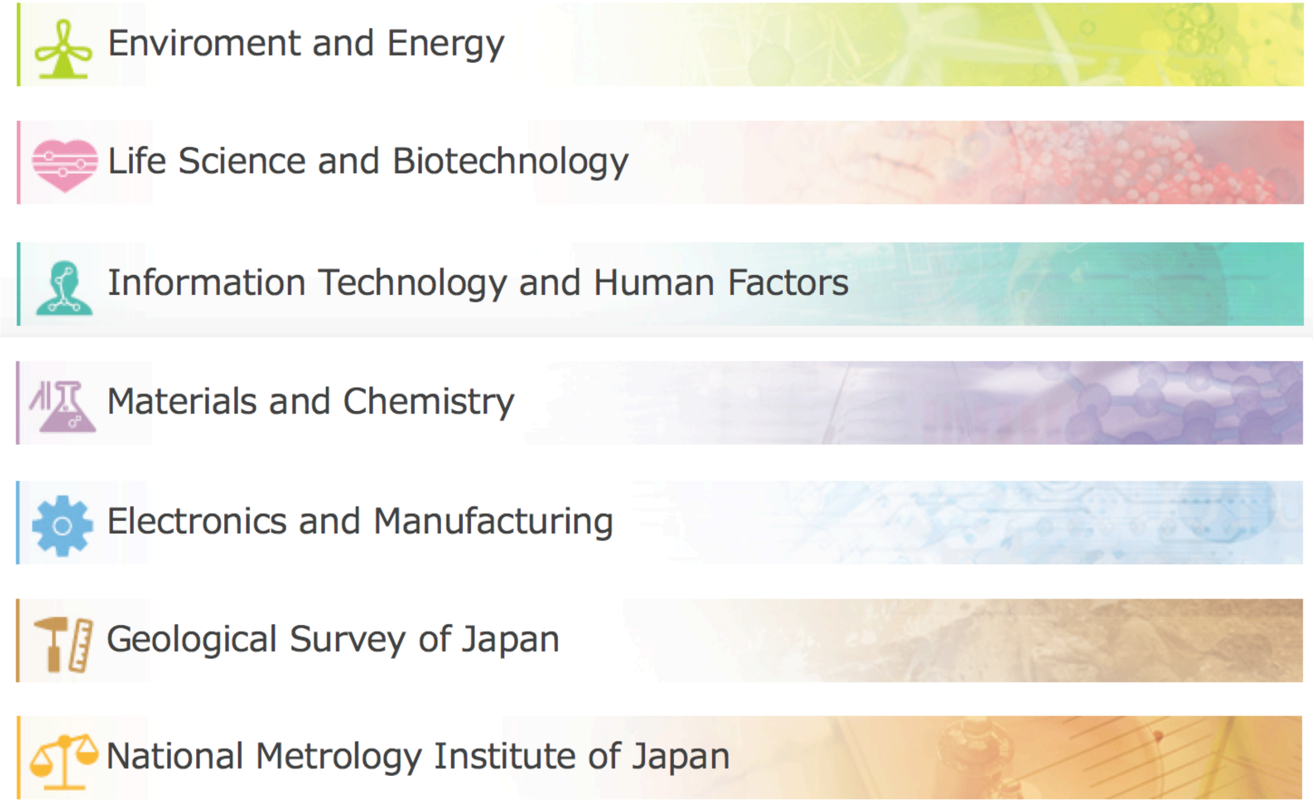
Aerial Look of Tsukuba(in part)



Geographical Survey Institute
University of Tsukuba
Mt.Tsukuba 876m
Aerial Tram or Funicular or Walk
Tsukuba Station
Geological Survey of Japan (GSJ) of AIST
JAXA Japan Aerospace eXploration Agency
Information Technology Research (ITRI) of AIST
Sakura-kan Guest House
AIST

Research at AIST

- 7 major research areas



- Enviroment and Energy
- Life Science and Biotechnology
- Information Technology and Human Factors
- Materials and Chemistry
- Electronics and Manufacturing
- Geological Survey of Japan
- National Metrology Institute of Japan

Good for Cross-Domain Research

Cross-Domain/ Interdisciplinary R&Ds

GeoScience + IT

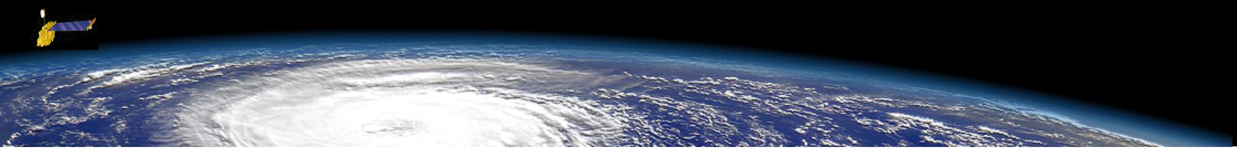
Bioscience + IT

Mechanics + IT

Etc.

Geoscience + IT

GeoGrid: An Example of Data Intensive Research Projects at ITRI



What is GEO Grid?

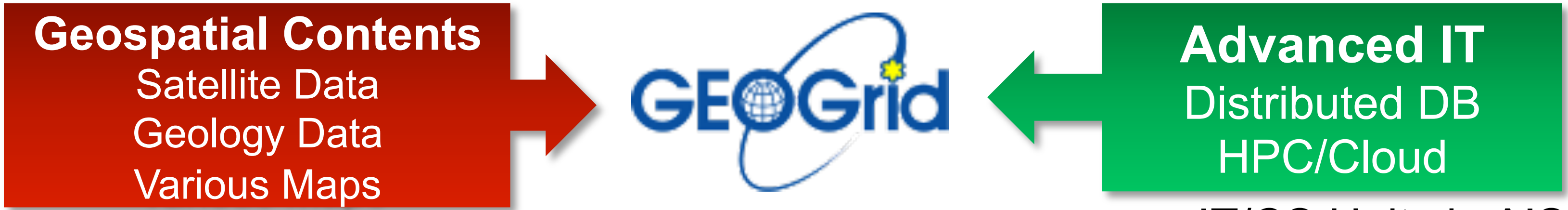
http://www.geogrid.org

GEO = Geospatial

Grid = Grid (cloud) Computing

e-Science infrastructure on heterogeneous data archives

– Cross-Domain (joint) project from 2004



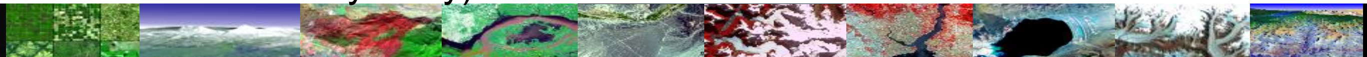
Geospatial Contents
Satellite Data
Geology Data
Various Maps

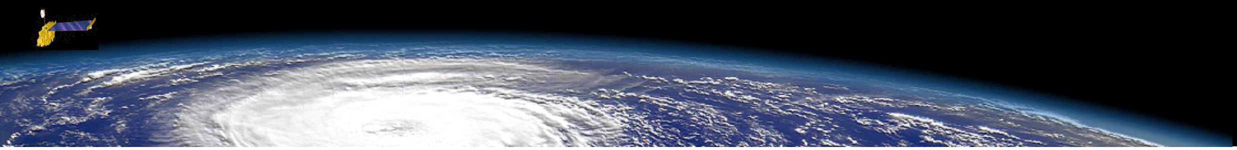
Advanced IT
Distributed DB
HPC/Cloud

Geology/Environment Units in AIST.

IT/CS Units in AIST

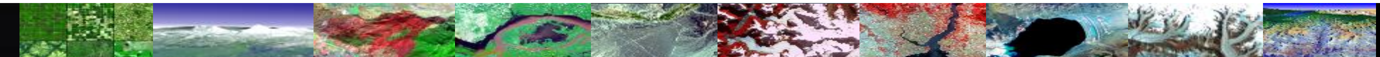
- **Core archive contents: Our Satellite Sensor Data**
  - ASTER satellite images >= 200TB(2000,000 scenes, y2000->)
  - Now extending to manage (Petabyte-Scale) PALSAR, PRISM, Landsat8 etc.
- **Core technologies: Grid Based => Parallel/Distributed R&D**
  - Distributed file system: Gfarm (started at AIST, Now at Tsukuba-U)
  - Database Integration: OGSA-DAI@Uk /Distributed SPARQL
  - Tsukuba-GAMA: Integrated Credential(Authentication) Management(some codes are included in MyProxy)





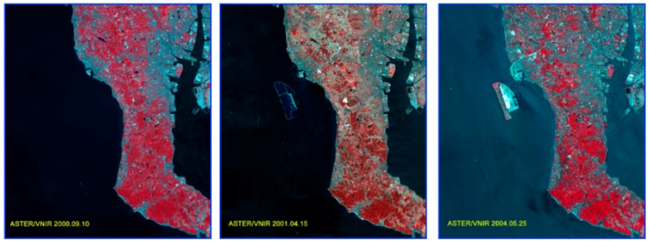
Major Technical Achievements of GEO Grid in 10 years

- **Petabyte-Class Large Scale Data Archive & Analysis**
  - Gfarm
- **Single sign-on system using Grid Security**
  - Tsukuba-GAMA
- **Heterogeneous Metadata Management based on OGC Standard**
  - AIST-CSW
- **Service-based Distributed Database Access**
  - OGSA-DAI(Web Services)

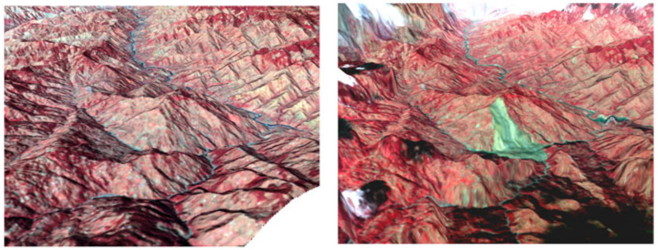


Data Archives

- ASTER sensor on NASA Terra satellite (2000~)
  - Resolution(Mid-range):15m(VNIR),30m(SWIR),90m(TIR)/px
  - 60km wide
    - 50~60GB daily Level 0 data transfer from NASA to JAPAN
  - 16 day observation cycle
    - Good for detecting long range change(= large computation)
  - 2 cameras with different angles
    - Can create DEM (Digital Elevation Model) by stereo matching



- Landsat-8 (by USGS)
  - Latest earth observation satellite launched 2013
    - 15m/Pan 30m/Color
    - 16 day observation cycle for the same area
  - Free and Open!



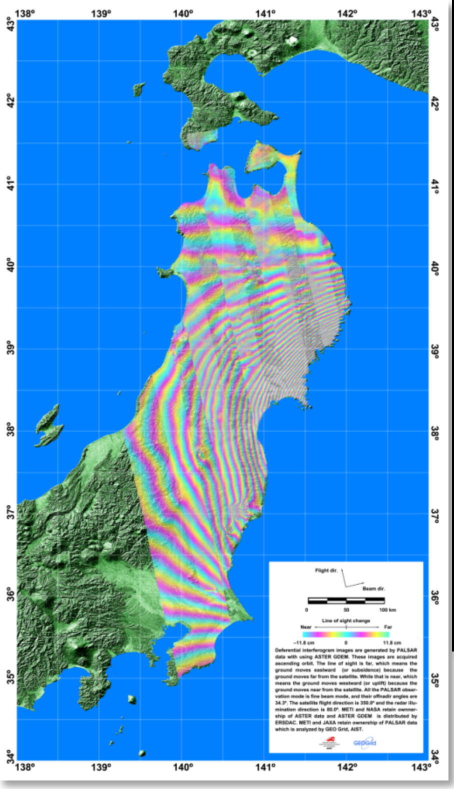
DEM of Pakistan Landslide 2005

- AIST set up the ground station for Landsat-8 (with Tokai-U)
  - Receives the daily data directly from the satellite
  - Can publish the data to the Internet in semi real-time
    - 2 hrs in AIST by our high performance computing (1 day in USGS)

3.11 science data examples produced by the GEO Grid

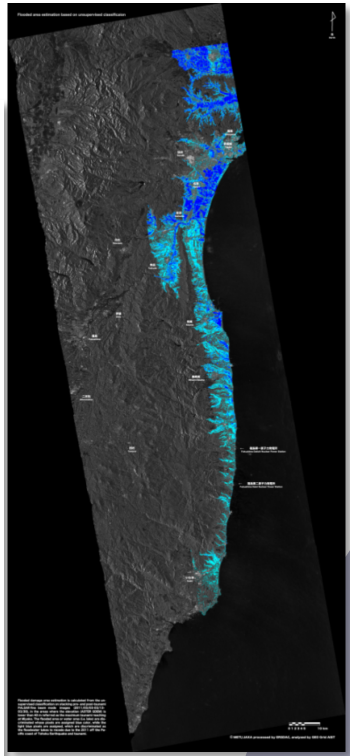
Grid-bas

Ground move with radar(SAR) satellite

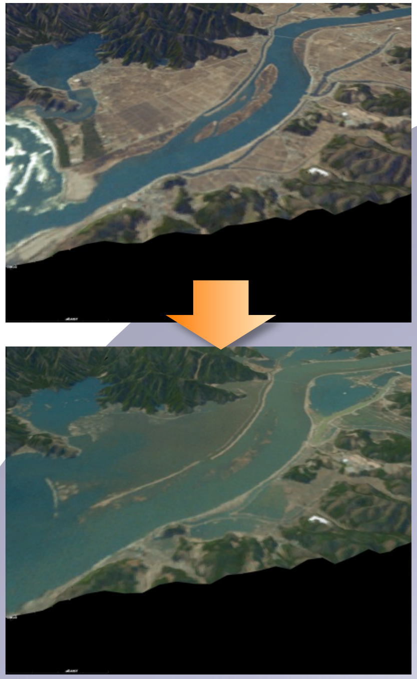


Deferential interferogram images are generated by PALSAR data with using ASTER GDEM. These images are acquired ascending orbit. The line of sight is far, which means the ground moves eastward (or subsidence) because the ground moves far from the satellite. While that is near, which means the ground moves westward (or uplift) because the ground moves near from the satellite. All the PALSAR observation mode is fine beam mode, and their offnadir angles are 34.3°. The satellite flight direction is 350.0° and the radar illumination direction is 80.0°. METI and NASA retain ownership of ASTER data and ASTER GDEM is distributed by ERSDAC. METI and JAXA retain ownership of PALSAR data which is analyzed by GEO Grid, AIST.

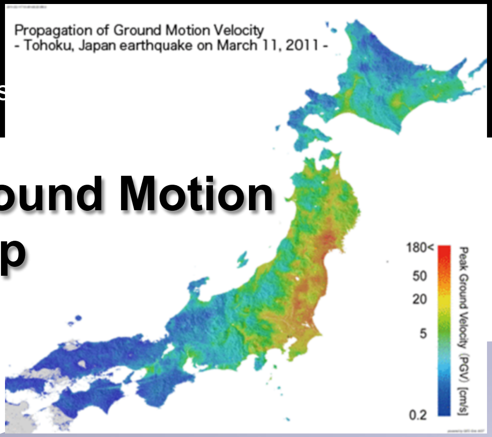
Flood simulation



ASTER images with 3D DEM

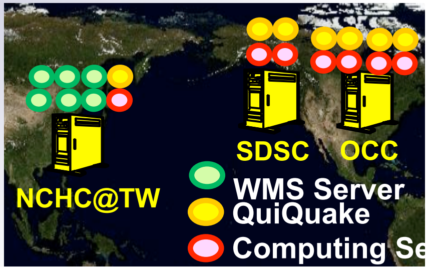


Ground Motion Map


Propagation of Ground Motion Velocity
- Tohoku, Japan earthquake on March 11, 2011 -
180<
50
20
5
0.2
Peak Ground Velocity (PGV) [cm/s]

GEO Grid archive/cluster is also damaged by 3.11 earthquake

We evacuated our environment using cloud technology and continued to process data in collaboration with OCC/SDSC/NCHC etc.



National Applied Research Laboratories National Center for High-Performance Computing Better HPC Better Living

UC San Diego

Open Cloud Consortium

CTC

Université Lille1

Orkney

NSPO

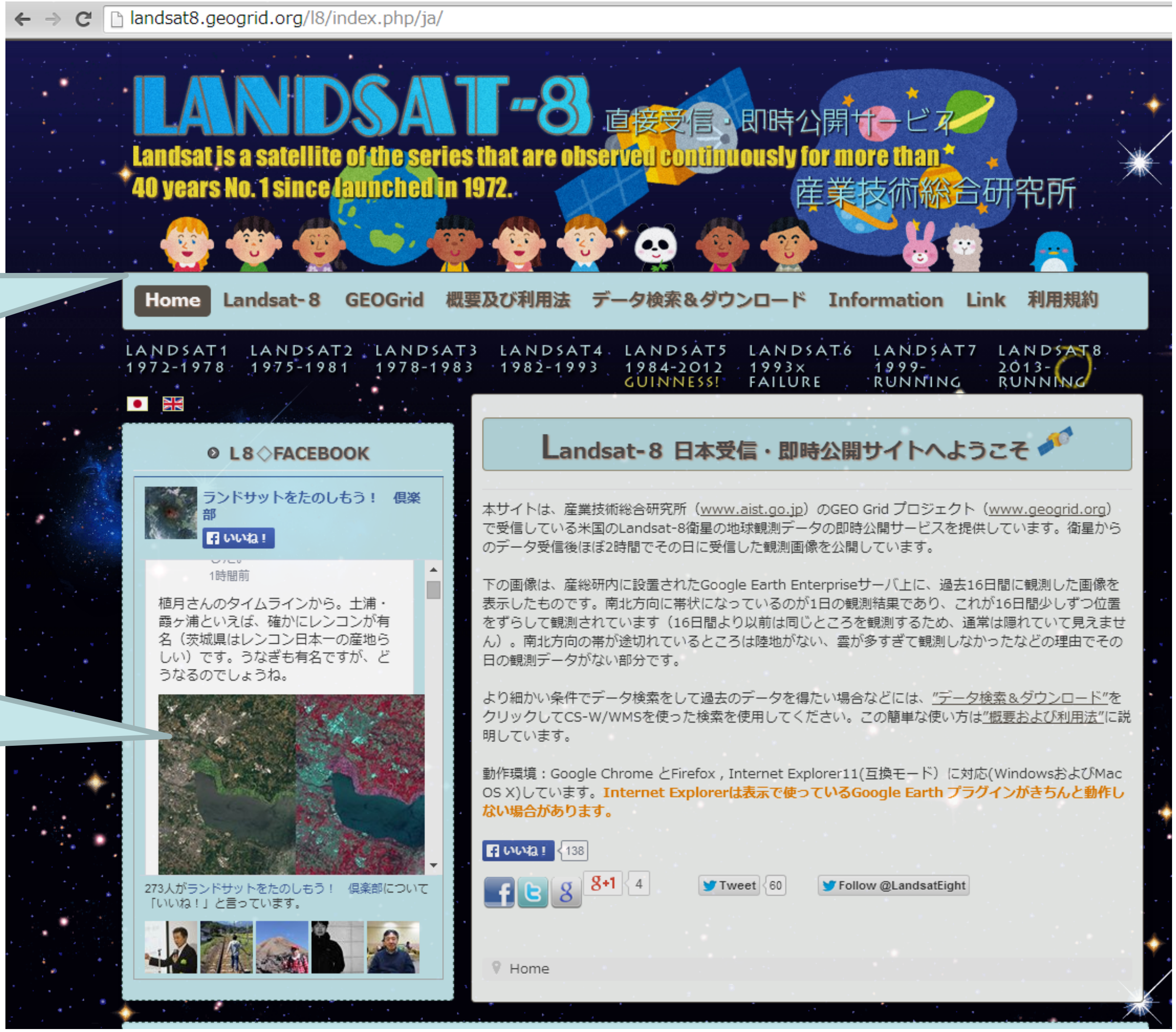
ITT

株式会社 NTTデータCCS

Public Service http://landsat8.geogrid.org

Latest/Historical Data can be Downloaded and Viewed

User Contributions Like
“I found interesting things!”
by Facebook



Landsat is a satellite of the series that are observed continuously for more than 40 years No.1 since launched in 1972.

植月さんのタイムラインから。土浦・霞ヶ浦といえば、確かにレンコンが有名（茨城県はレンコン日本一の産地らしい）です。うなぎも有名ですが、どうなるのでしょうね。

273人がランドサットをたのしもう！　倶楽部について「いいね！」と言っています。

Landsat-8 日本受信・即時公開サイトへようこそ

本サイトは、産業技術総合研究所（www.aist.go.jp）のGEO Grid プロジェクト（www.geogrid.org）で受信している米国のLandsat-8衛星の地球観測データの即時公開サービスを提供しています。衛星からのデータ受信後ほぼ2時間でその日に受信した観測画像を公開しています。

下の画像は、産総研内に設置されたGoogle Earth Enterpriseサーバ上に、過去16日間に観測した画像を表示したものです。南北方向に帯状になっているのが1日の観測結果であり、これが16日間少しずつ位置をずらして観測されています（16日間より以前は同じところを観測するため、通常は隠れていて見えません）。南北方向の帯が途切れているところは陸地がない、雲が多すぎて観測しなかったなどの理由でその日の観測データがない部分です。

より細かい条件でデータ検索をして過去のデータを得たい場合などには、”データ検索&ダウンロード”をクリックしてCS-W/WMSを使った検索を使用してください。この簡単な使い方は”概要および利用法”に説明しています。

動作環境：Google Chrome とFirefox , Internet Explorer11(互換モード）に対応(WindowsおよびMac OS X)しています。**Internet Explorerは表示で使っているGoogle Earth プラグインがきちんと動作しない場合があります。**

Constellation

- ASTER = 16 days cycle
- Landsat-8 = 16 days cycle
- ASTER+Landsat-8 = 8 days cycle (same orbits)

Target: Daily change detection

- – Example: Skybox (which is acquired by Google) has a plan to launch 20 satellites

We are investigating to do the same thing with existing (and new) satellites

Analysis

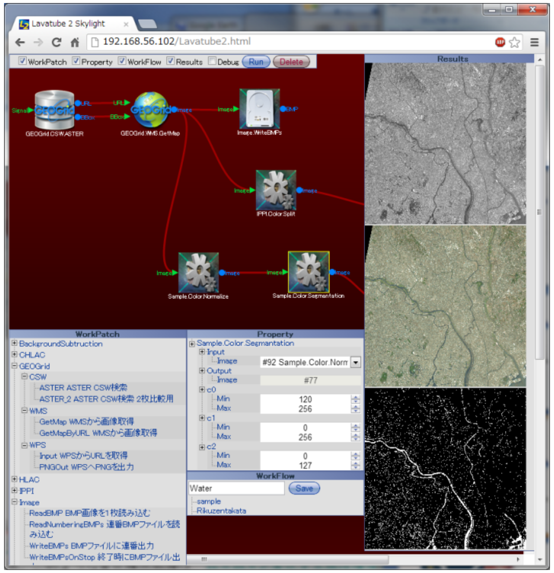
Workflow engine: Lavatube

Machine Learning System: Hivemall

Our yet another Workflow Engine: Lavatube

for spacio-temporal data and image/movie processing

1. Support rest-based OGC (OpenGeospatial Consortium) services
2. Support various image/movie processing modules
3. Provide High-Level interaction



WorkPatch

- ⊞ BackgroundSubtruction
- ⊞ CHLAC
- ⊟ GEOGrid
  - ⊟ CSW
    - ASTER ASTER CSW検索
    - ASTER_2 ASTER CSW検索 2枚比較用
  - ⊟ WMS
    - GetMap WMSから画像取得
    - GetMapByURL WMSから画像取得
  - ⊟ WPS
    - Input WPSからURLを取得
    - PNGOut WPSへPNGを出力
- ⊞ HLAC
- ⊞ IPPI
- ⊟ Image
  - ReadBMP BMP画像を1枚読み込む
  - ReadNumberingBMPs 連番BMPファイルを読み込む
  - WriteBMPs BMPファイルに連番出力
  - WriteBMPsOnStop 終了時にBMPファイル出

Property

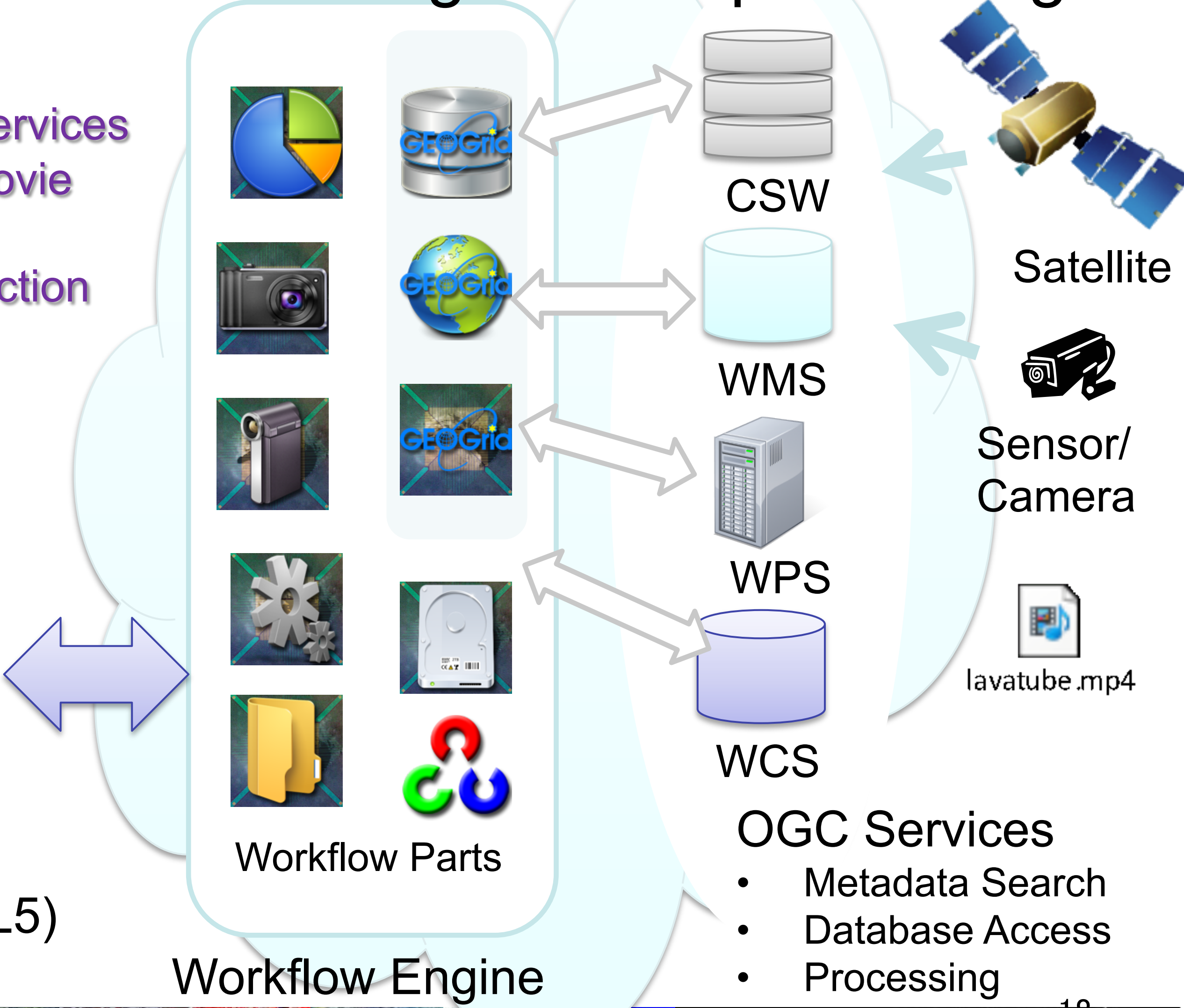
- ⊞ Sample.Color.Segmantation
  - ⊞ Input
    - Image: #92 Sample.Color.Norm
  - ⊞ Output
    - Image: #77
  - ⊞ c0
    - Min: 120
    - Max: 256
  - ⊞ c1
    - Min: 0
    - Max: 256
  - ⊞ c2
    - Min: 0
    - Max: 127

WorkFlow

Water Save

- sample
- Rikuzentakata

Browser Interface(HTML5)
or Windows engine



Workflow Parts

Workflow Engine

OGC Services

- Metadata Search
- Database Access
- Processing

- Hivemall: Scalable Machine Learning Library for Apache Hive
- A collection of machine learning algorithms as Hive UDFs/UDTFs
  - Classification & Regression
  - Recommendation
  - k-Nearest Neighbor Search
- An open-source project on Github
  - Licensed under LGPL
  - github.com/myui/hivemall (bit.ly/hivemall)

Application

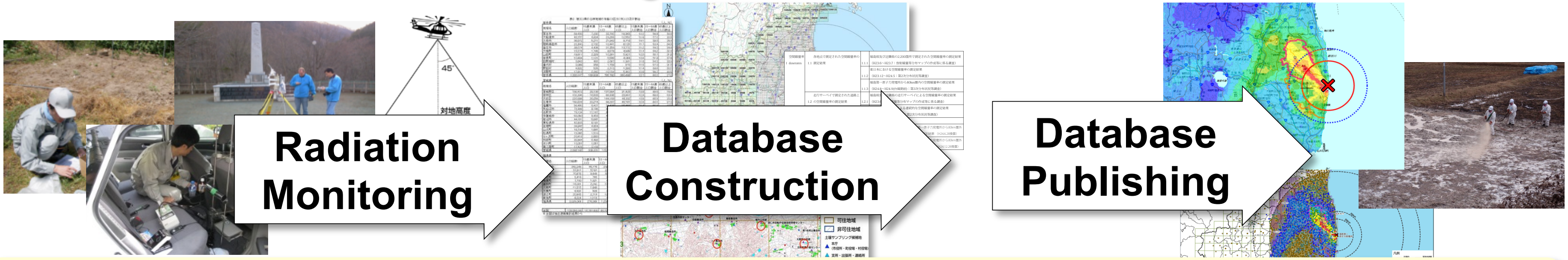
Radiation Monitoring Database for Fukushima

Project Overview

Radiation Monitoring Data is important to:

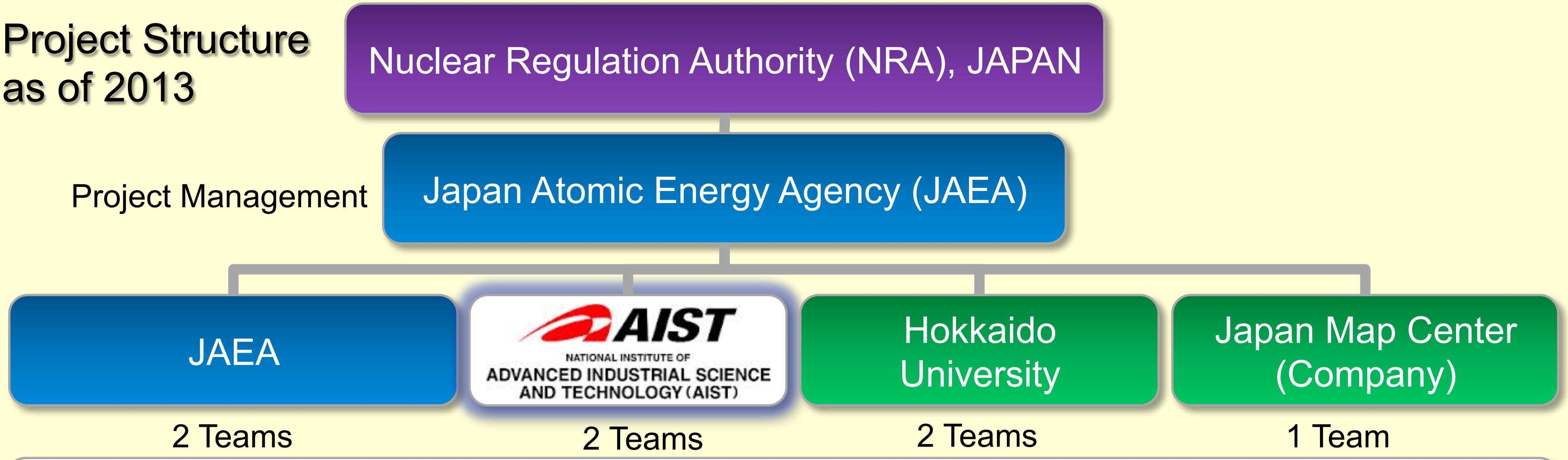
- Understand what happened at the accident in the past
- Help the decision making for the future

National Project to continuously Monitor/Construct/Publish Radiation Monitoring Database of Fukushima Area



Radiation Monitoring → **Database Construction** → **Database Publishing**

Project Structure as of 2013



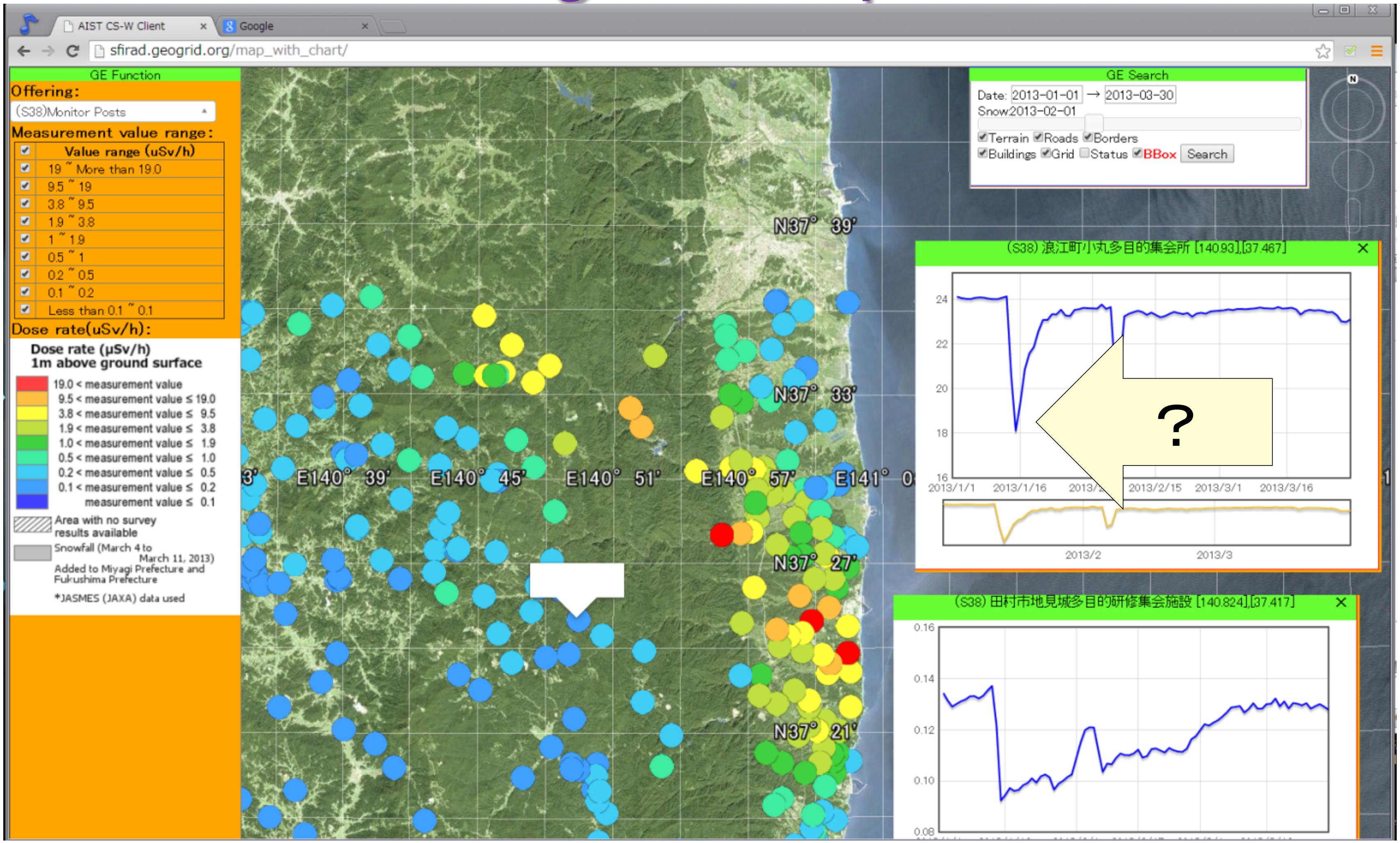
Nuclear Regulation Authority (NRA), JAPAN

Project Management — Japan Atomic Energy Agency (JAEA)

JAEA	AIST NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)	Hokkaido University	Japan Map Center (Company)
2 Teams	2 Teams	2 Teams	1 Team

Partners ・・・・・ GIS.FCU Okayama-U,
The Institute of Statistical Mathematics,,

Example Data Integration Application using OGC specs



AIST CS-W Client | Google

sfirad.geogrid.org/map_with_chart/

GE Function

Offering:

(S38)Monitor Posts

Measurement value range:

☑	Value range (uSv/h)
☑	19 ˜ More than 19.0
☑	9.5 ˜ 19
☑	3.8 ˜ 9.5
☑	1.9 ˜ 3.8
☑	1 ˜ 1.9
☑	0.5 ˜ 1
☑	0.2 ˜ 0.5
☑	0.1 ˜ 0.2
☑	Less than 0.1 ˜ 0.1

Dose rate(uSv/h):

Dose rate (μSv/h)
1m above ground surface

- 🟥 19.0 < measurement value
- 🟧 9.5 < measurement value ≤ 19.0
- 🟨 3.8 < measurement value ≤ 9.5
- 🟩 1.9 < measurement value ≤ 3.8
- 🟩 1.0 < measurement value ≤ 1.9
- 🟩 0.5 < measurement value ≤ 1.0
- 🟦 0.2 < measurement value ≤ 0.5
- 🟦 0.1 < measurement value ≤ 0.2
- 🟦 measurement value ≤ 0.1
- ▨ Area with no survey results available
- ⬜ Snowfall (March 4 to March 11, 2013) Added to Miyagi Prefecture and Fukushima Prefecture

*JASMES (JAXA) data used

GE Search

Date: 2013-01-01 → 2013-03-30

Snow:2013-02-01

☑Terrain ☑Roads ☑Borders
☑Buildings ☑Grid □Status ☑BBox Search

(S38) 浪江町小丸多目的集会所 [140.93],[37.467] ✕

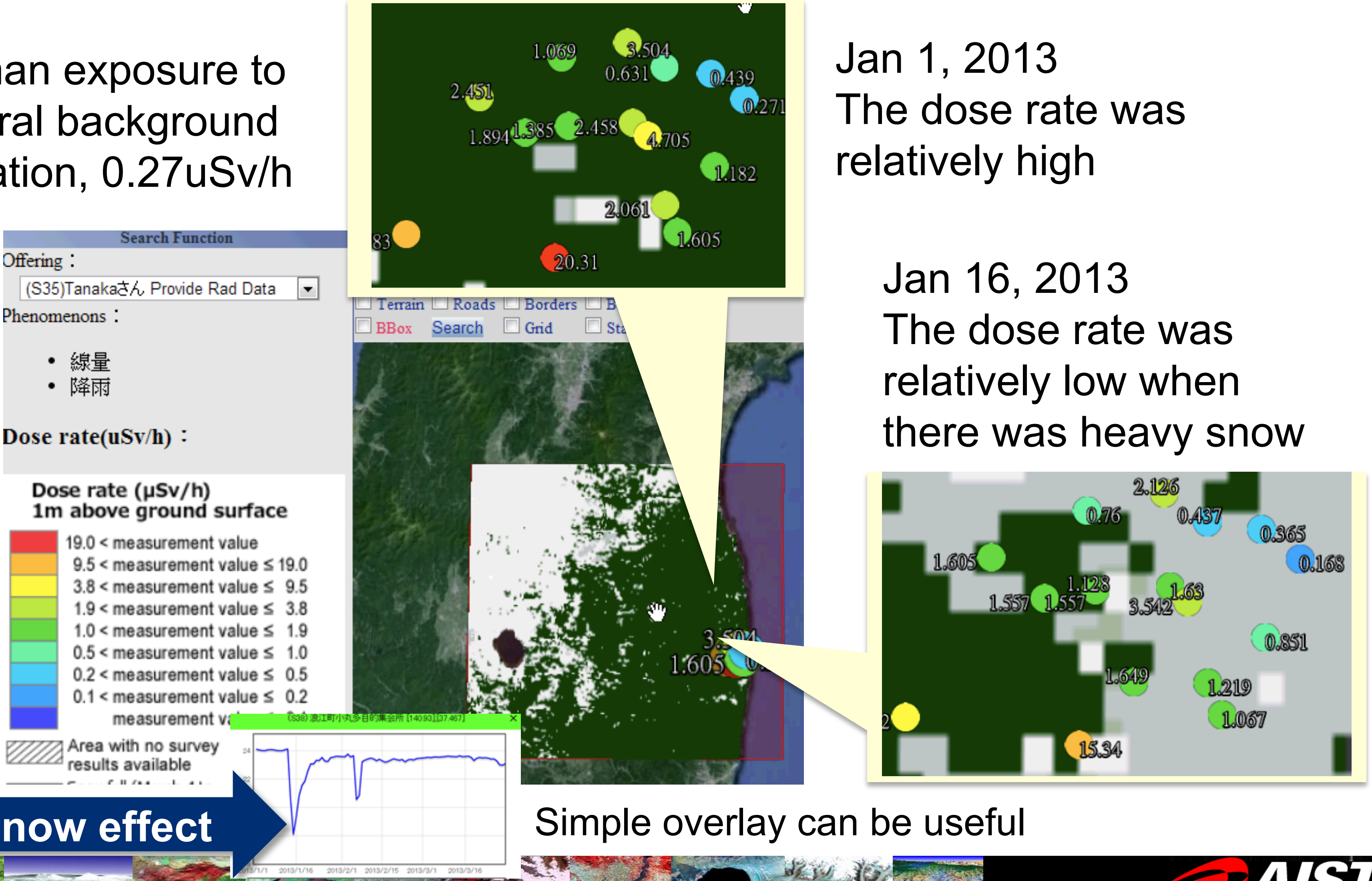
(S38) 田村市地見城多目的研修集会施設 [140.824],[37.417] ✕

?

Combine SOS (Sensor Observation Service) with other WMS (Web Map Service) data source (Weather)

Human exposure to natural background radiation, 0.27uSv/h

Jan 1, 2013
The dose rate was relatively high



Search Function

Offering：
(S35)Tanakaさん Provide Rad Data

Phenomenons：

- 線量
- 降雨

Dose rate(uSv/h)：

Dose rate (μSv/h)
1m above ground surface

🟥	19.0 < measurement value
🟧	9.5 < measurement value ≤ 19.0
🟨	3.8 < measurement value ≤ 9.5
🟩	1.9 < measurement value ≤ 3.8
🟩	1.0 < measurement value ≤ 1.9
🟩	0.5 < measurement value ≤ 1.0
🟦	0.2 < measurement value ≤ 0.5
🟦	0.1 < measurement value ≤ 0.2
🟦	measurement v
▨	Area with no survey results available

Jan 16, 2013
The dose rate was relatively low when there was heavy snow

Snow effect

Simple overlay can be useful

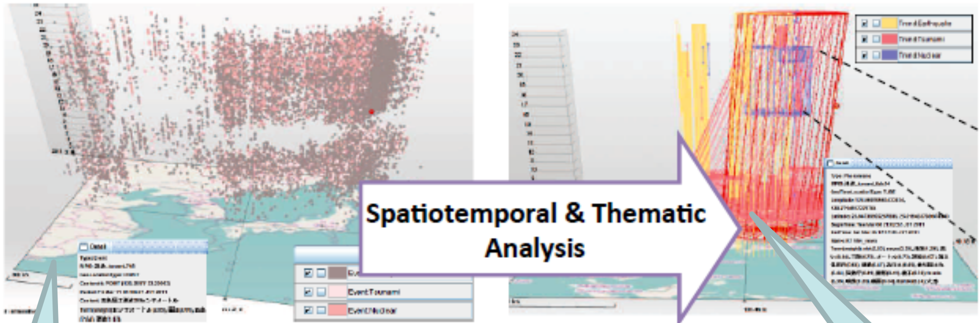
Directions

Social, Mobile & Crowdsourcing

Data Integration Issue: Administrative & Non-Administrative Data

- Administrative Data (Current GEO Grid Data)
  - Governmental & official data
  - Limited amount with controlled quality
- Non-Administrative Data
  - NPO, Social media, crowdsourcing (Twitter, etc.)
  - Large amount, variable quality

Application Examples



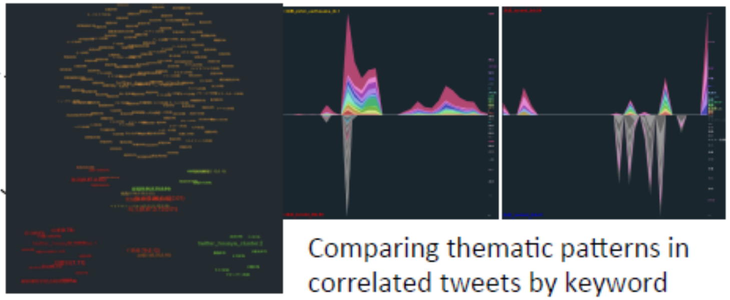
Spatiotemporal & Thematic Analysis

Mapping, clustering, and regression: (2011/03/02-2011/03/24)

1) earthquake-related tweets
2) tsunami-related tweets
3) nuclear-related tweets

地理情報システムの プカンファレンス
ACM GIS 2011で優秀デ を受賞

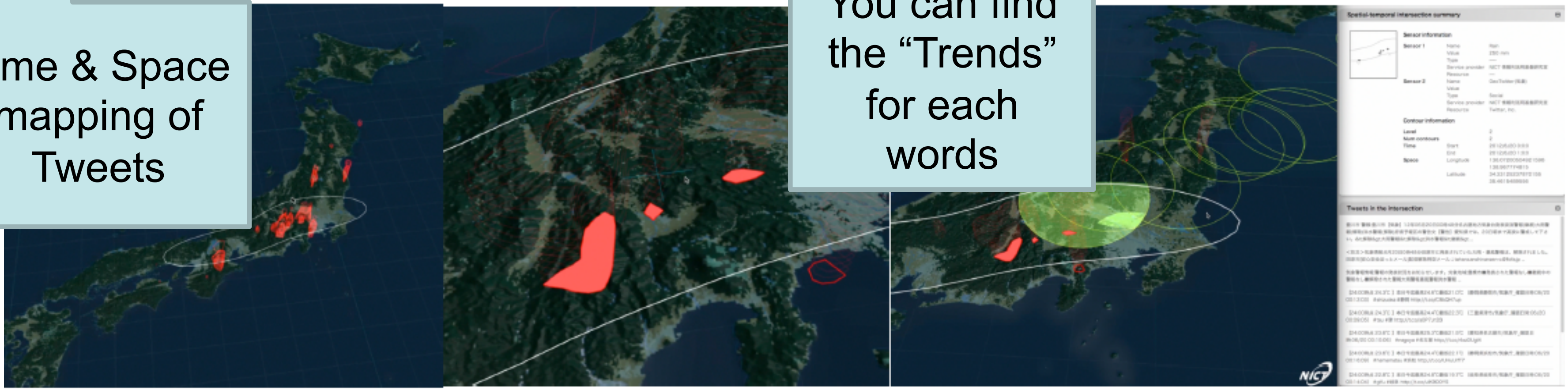
- mTrend (ACMGIS2011 Demonstration)



Comparing thematic patterns in correlated tweets by keyword streams

Time & Space mapping of Tweets

You can find the “Trends” for each words


Spatial-temporal intersection summary
Tweets in the Intersection
NICT

- Cyber-Physical Data Cloud: An Infrastructure for Interconnecting Heterogeneous Sensor Data (WTP2012 Demonstration)

Situation creation on the basis of intersection area of outbreaks between tweets and natural phenomena

Crowdsourcing and Notifications Joint Research with Taiwan ITRI

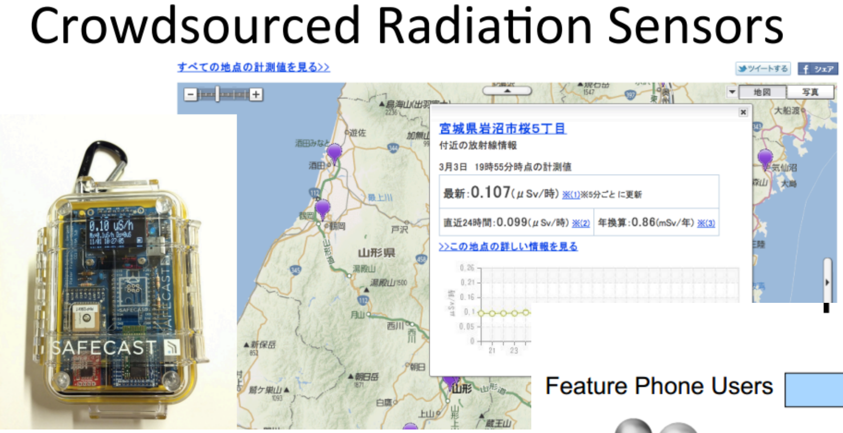
工業技術研究院
Industrial Technology
Research Institute

ITRI & AIST Joint Research Project
Data Integration and Messaging Fram
Crowdsourced and Administrative Sen

莊國煜
Kuo-Yu slayer Chuang

slayer@itri.org.tw

Crowdsourced Radiation Sensors

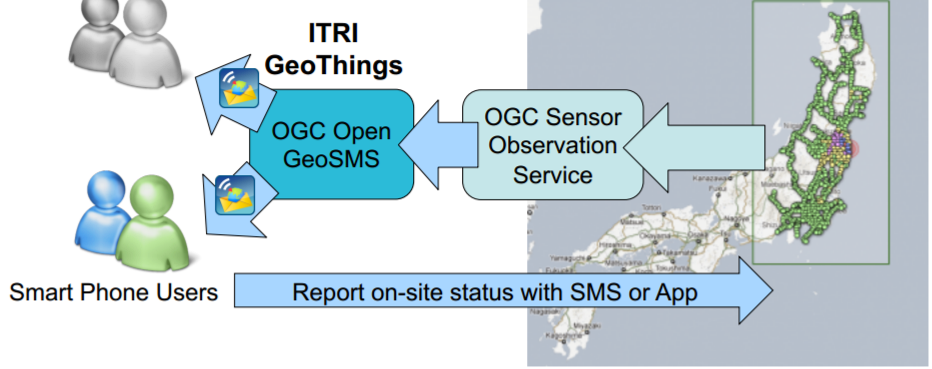


すべての地点の計測値を見る>>

宮城県岩沼市桜5丁目
付近の放射線情報
3月3日　19時55分時点の計測値

最新:0.107(μSv/時) ※(1)※5分ごと に更新
直近24時間:0.099(μSv/時) ※(2)　年換算:0.86(mSv/年) ※(3)

>>この地点の詳しい情報を見る



Linked Open Data

Federated SPARQL with “Best-Effort” Query Processing

Linked Open Data (LOD)

Try to create a huge linked knowledge cloud

- The data is written with RDF (Resource Description Framework)
  - The Standard for the Semantic web community
- Highly distributed and rapidly increasing
  - More than 300 sites, billion~trillion triples
- Cross-Domain
  - GEO, BIO, Government, Media, etc
  - Many governmental data is going to be published as LOD
- Issues
  - Distributed SPARQL processing may be slow, but a centralized data service lacks freshness
  - Heterogeneity with SPARQL Endpoints, plain RDF Texts

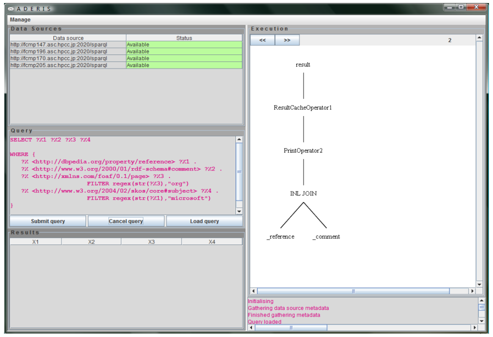
Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences
As of September 2011

Our Approach
Hybrid Adaptive Query Processing

Based on the **freshness, coverage** and the **response time**

Adaptive Query optimization

Pre-defined Query Processing Schedule



Data source	Status
http://fcmp147.asc.hpcc.jp:2020/sparql	Available
http://fcmp196.asc.hpcc.jp:2020/sparql	Available
http://fcmp170.asc.hpcc.jp:2020/sparql	Available
http://fcmp205.asc.hpcc.jp:2020/sparql	Available

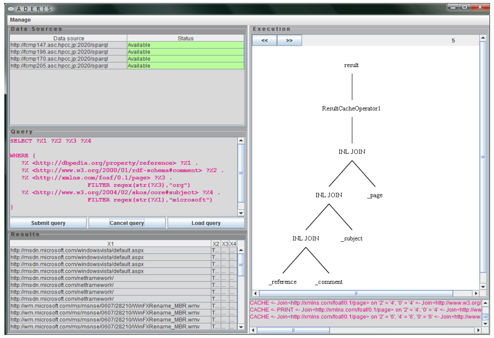
```
SELECT ?X1 ?X2 ?X3 ?X4

WHERE {
   ?X <http://dbpedia.org/property/reference> ?X1 .
   ?X <http://www.w3.org/2000/01/rdf-schema#comment> ?X2 .
   ?X <http://xmlns.com/foaf/0.1/page> ?X3 .
                    FILTER regex(str(?X3),"org")
   ?X <http://www.w3.org/2004/02/skos/core#subject> ?X4 .
                    FILTER regex(str(?X1),"microsoft")
}
```

X1	X2	X3	X4

```
Initialising
Gathering data source metadata
Finished gathering metadata
Query loaded
```

Network delays,
Too many results
Site troubles, etc...



Data source	Status
http://fcmp147.asc.hpcc.jp:2020/sparql	Available
http://fcmp196.asc.hpcc.jp:2020/sparql	Available
http://fcmp170.asc.hpcc.jp:2020/sparql	Available
http://fcmp205.asc.hpcc.jp:2020/sparql	Available

```
SELECT ?X1 ?X2 ?X3 ?X4

WHERE {
   ?X <http://dbpedia.org/property/reference> ?X1 .
   ?X <http://www.w3.org/2000/01/rdf-schema#comment> ?X2 .
   ?X <http://xmlns.com/foaf/0.1/page> ?X3 .
                    FILTER regex(str(?X3),"org")
   ?X <http://www.w3.org/2004/02/skos/core#subject> ?X4 .
                    FILTER regex(str(?X1),"microsoft")
}
```

X1	X2	X3	X4
http://msdn.microsoft.com/windowsvista/default.aspx	T...	...	...
http://msdn.microsoft.com/windowsvista/default.aspx	T...	...	...
http://msdn.microsoft.com/windowsvista/default.aspx	T...	...	...
http://msdn.microsoft.com/windowsvista/default.aspx	T...	...	...
http://msdn.microsoft.com/netframework/	T...	...	...
http://msdn.microsoft.com/netframework/	T...	...	...
http://msdn.microsoft.com/netframework/	T...	...	...
http://msdn.microsoft.com/netframework/	T...	...	...
http://wm.microsoft.com/ms/msnse/0607/28210/WinFXRename_MBR.wmv	T...	...	...
http://wm.microsoft.com/ms/msnse/0607/28210/WinFXRename_MBR.wmv	T...	...	...
http://wm.microsoft.com/ms/msnse/0607/28210/WinFXRename_MBR.wmv	T...	...	...
http://wm.microsoft.com/ms/msnse/0607/28210/WinFXRename_MBR.wmv	T...	...	...

```
CACHE <- Join<http://xmlns.com/foaf/0.1/page> on '2' = '4', '0' = '4' <- Join<http://www.w3.org/
CACHE <- PRINT <- Join<http://xmlns.com/foaf/0.1/page> on '2' = '4', '0' = '4' <- Join<http://ww
CACHE <- Join<http://xmlns.com/foaf/0.1/page> on '2' = '6', '4' = '6', '0' = '6' <- Join<http://www.
```

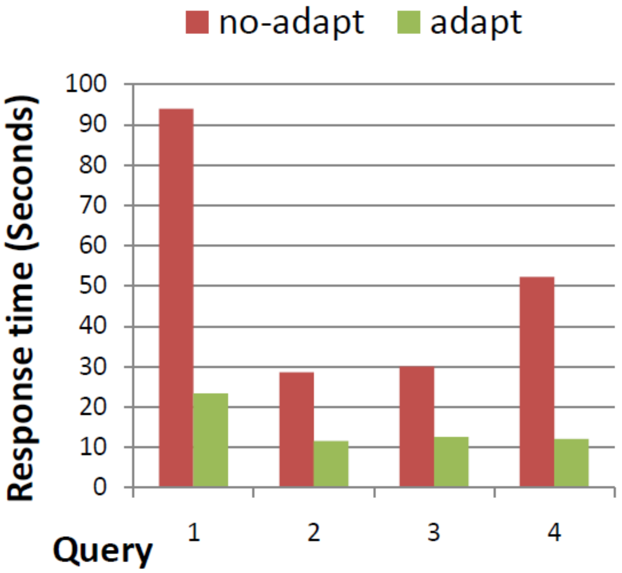
Modified Processing Adaptively

```
dbo: <http://dbpedia.org/ontology/>
dbp: <http://dbpedia.org/property/>
owl: <http://www.w3.org/2002/07/owl#>

Query 1 (Result size = 150):
select * where {
 ?x dbp:reference ?ref .       777,679
 ?x rdf:comment ?comment .     10,000
 ?x skos:subject ?subj .       9971
 ?x foaf:page ?page .          10,000
 ?x rdf:type ?type .           800,000
 FILTER ( regex(str(?subj),"building") )
}
Query 2 (Result size = 8):
select * where {
 ?x dbp:reference ?ref .      777,679
 ?x rdf:comment ?comment .    10,000
 ?x skos:subject ?subj .      9971
 ?x foaf:page ?page .         10,000
 ?x rdf:type dbo:book         3105
}
Query 3 (Result size = 8):
select * where {
 ?x dbp:reference ?ref .      777,679
 ?x rdf:comment ?comment .    10,000
 ?x skos:subject ?subj .      9971
 ?x foaf:page ?page .         10,000
 ?x rdf:type dbo:book         3105
 ?x dbo:releaseDate ?date     (DBP) 126,737
}
```

```
rdf:  <http://www.w3.org/2000/01/rdf-schema#>
skos: <http://www.w3.org/2004/02/skos/core#>
foaf: <http://xmlns.com/foaf/0.1/>

Query 4 (Result size = 13):
select * {
 ?book rdf:type dbo:Book .   3105
 ?book foaf:page ?p .        10,000
 ?book owl:sameAs ?link      (DBP) 10,121,699
}
```

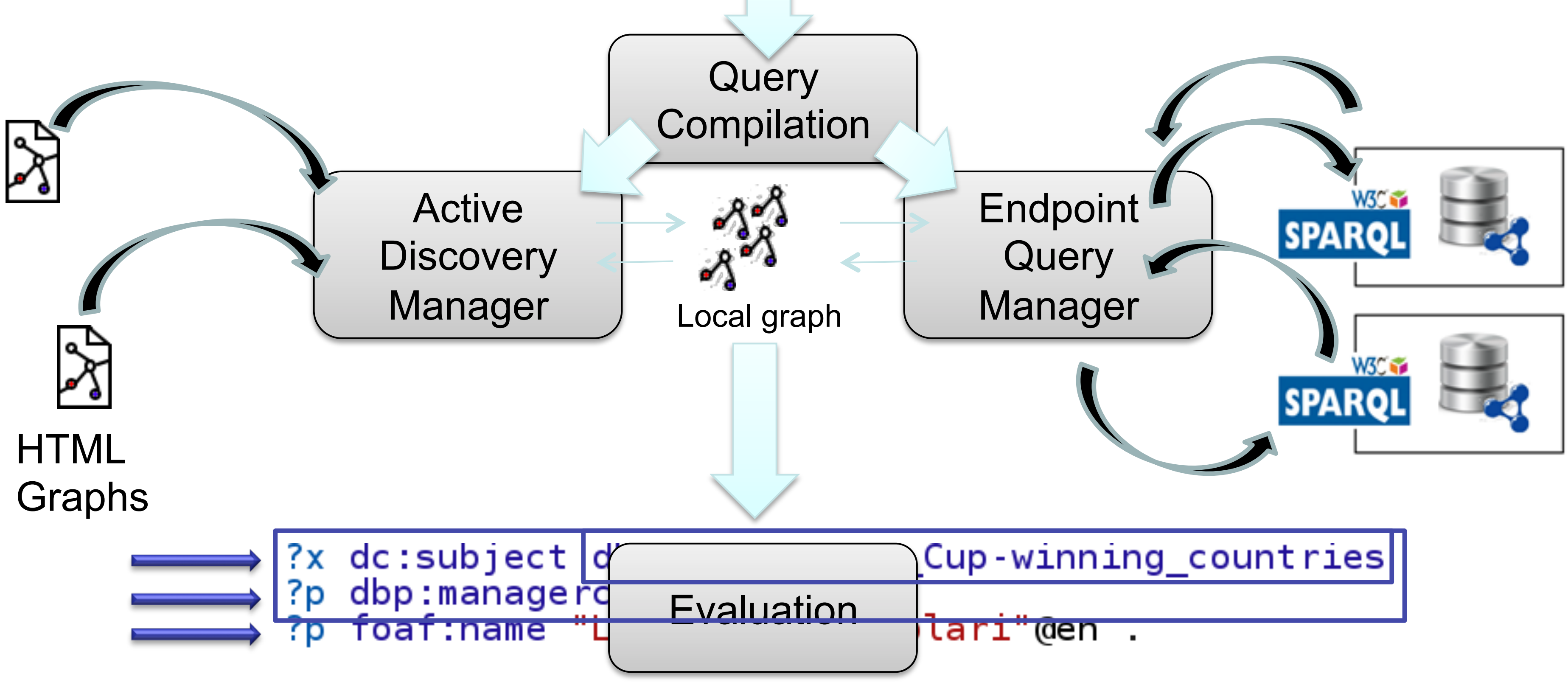


Achieve good performance around 10 distributed SPARQL endpoints (still small for 300 ;-<)

Hybrid & Adaptive Query Processing

User's SPARQL Query

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dbp: <http://dbpedia.org/resource/property>
SELECT * WHERE {
    ?x dc:subject dbp:FIFA_World_Cup-winning_countries .
    ?p dbp:managerclubs ?x .
    ?p foaf:name "Luiz Felipe Scolari"@en .
}
```



Query Result

You can get rough answer in 10 seconds,,
or
More accurate answer with a long time

BioScience + IT

BIO-CAD/LEAD

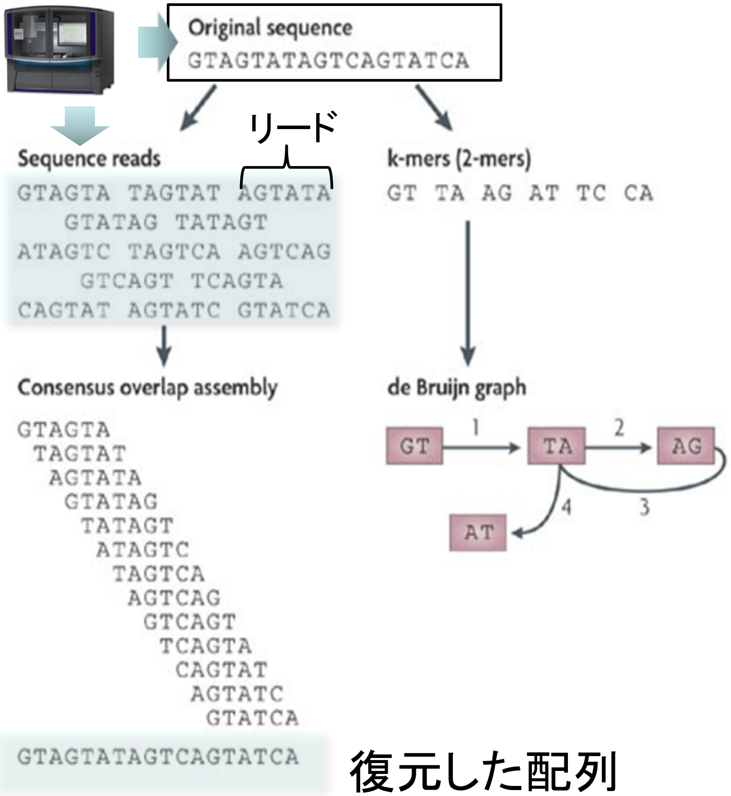
Hydra: Molecular Visualization

High Performance Genomics Assembly

- **Next Generation Sequencers**

  Huge set of short reads are obtained

  - **1 read： ATGC（base）100(50basex2)**
  - **Total ： 100million reads just for 1 run**



Original sequence
GTAGTATAGTCAGTATCA

リード

Sequence reads
GTAGTA TAGTAT AGTATA
GTATAG TATAGT
ATAGTC TAGTCA AGTCAG
GTCAGT TCAGTA
CAGTAT AGTATC GTATCA

Consensus overlap assembly
GTAGTA
TAGTAT
AGTATA
GTATAG
TATAGT
ATAGTC
TAGTCA
AGTCAG
GTCAGT
TCAGTA
CAGTAT
AGTATC
GTATCA

GTAGTATAGTCAGTATCA

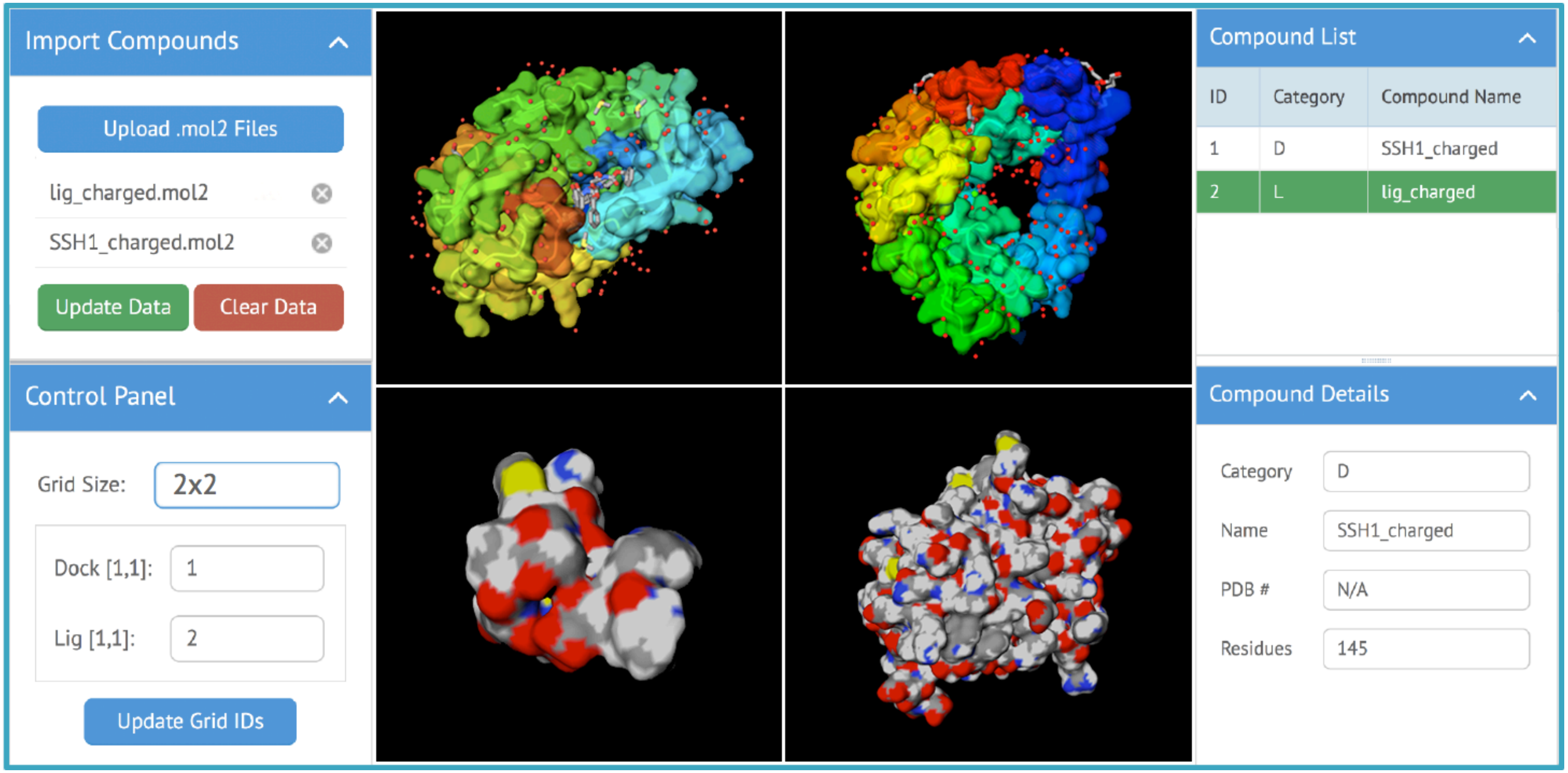
復元した配列

k-mers (2-mers)
GT TA AG AT TC CA

de Bruijn graph

- **Hybrid Assembly Workflow**
  - MPI parallelized (SAET, ASiD)
  - Improve the algorithms (Velvet)
  - To achieve scalability and performance enhancements

Hydra Molecular Visualization



Import Compounds

lig_charged.mol2

SSH1_charged.mol2

Control Panel

Grid Size: 2x2

Dock [1,1]: 1

Lig [1,1]: 2

Compound List

ID	Category	Compound Name
1	D	SSH1_charged
2	L	lig_charged

Compound Details

Category	D
Name	SSH1_charged
PDB #	N/A
Residues	145

- Create a more device agnostic tool

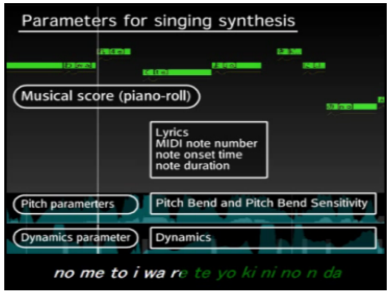
  Visualization of multiple protein-ligand interactions

Mechanics + IT

Media-related R&D

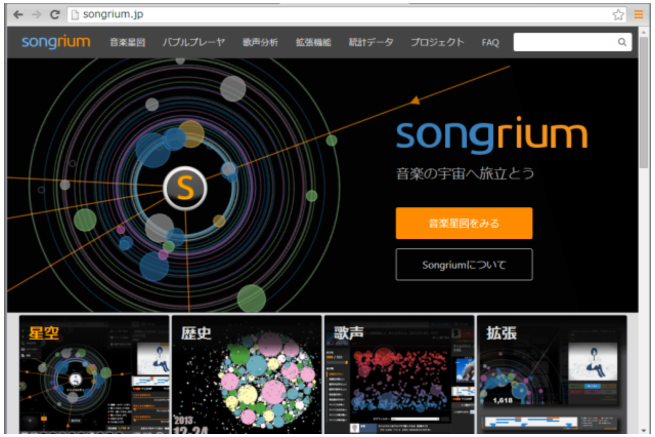
Media-Related R&D

- IT behind the robot
  - Computer Singing Systems
    - **VocaListener**
    - **VocaWatcher**


Parameters for singing synthesis
Musical score (piano-roll)
Lyrics
MIDI note number
note onset time
note duration
Pitch parameters
Pitch Bend and Pitch Bend Sensitivity
Dynamics parameter
Dynamics
no me to i wa re te yo ki ni no n da

Sagashituzuketa FutarinohoNtono Prologue
探し続けた 二人のほんとの プロローグ

- Active Music-Listening Web Service
  - **Songrium**



Summary

- AIST has many interdisciplinary data-oriented R&D projects
  - Geospatial
  - Linked Data
  - Bioinformatics
  - Multimedia (Music/Songs)
- Looking forward to the OSDC students contribution

- AIST YouTube: https://www.youtube.com/user/aistchannel

Acknowledgements

Isao Kojima

Kyoungsook Kim

Steven Lynden

Hirotaka Ogawa

Tsutomu Ikegami

Yuan Zhao