

「**研究のバリア**」を打破する研究基盤デザインと研究データ利活用

# Design of Research Infrastructure and Utilization of Research Data for Breaking through '**Research Barriers**'

北本 朝展

Asanobu KITAMOTO

国立情報学研究所・総合研究大学院大学

National Institute of Informatics / SOKENDAI

<http://agora.ex.nii.ac.jp/~kitamoto/>

# Research Barriers



Research Data Sharing  
without barriers

<https://rd-alliance.org/>

- Barriers between you and others.
- Barriers between organizations.
- Barriers between communities.
- Barriers between scientists and citizens.

# Open Science

# What is Open Science?

Open Science is the **convergence of dreams** by people who are not satisfied with the current practice of science, and see a **possibility of revolution by leveraging the concept of openness.**

# Details in Another Slide

## Convergence to Open Science

オープンサイエンスへのコンバージェンス

Formation of a Community to Foster  
Shared Perception from Different Dreams  
同床異夢から共通認識を醸成するコミュニティの形成

National Institute of Informatics / SOKENDAI

Asanobu KITAMOTO

<http://agora.ex.nii.ac.jp/~kitamoto/>

2015/09/17

オープンサイエンスデータWS

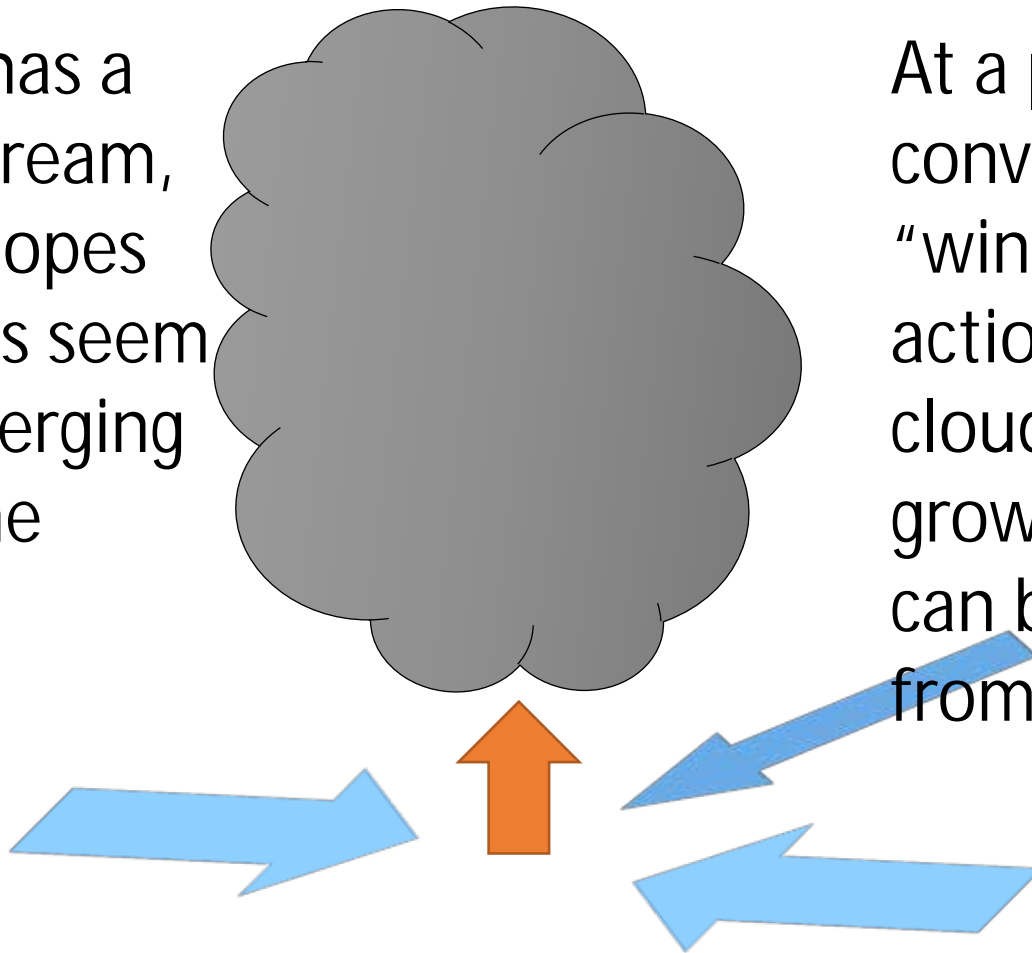
1

オープンサイエンス  
データ推進ワーク  
ショップ  
2015年9月17日～18日  
@京都大学

<http://agora.ex.nii.ac.jp/~kitamoto/research/publications/osd15.html.ja>

# Convergence to Open Science

Everyone has a different dream, but their hopes and actions seem to be converging to the same point.



At a point of convergence, “winds” of actions grow tall clouds. As it grows higher, it can be seen from a distance.

# Case Study: “Digital Typhoon” (Science)

# Digital Typhoon

<http://agora.ex.nii.ac.jp/digital-typhoon/>

The screenshot shows the 'Digital Typhoon: Typhoon Images and Information' website. The top navigation bar includes 'Home', 'Earth', and 'Digital Typhoon'. Below the navigation, there are links for 'Notice', 'Feedback', 'Wallpaper', 'Movie', 'Kids', 'RSS', 'Atom', 'Media RSS', and 'OpenSearch'. The main content area is divided into several sections:

- Real-time Typhoon Information:** This section displays three typhoon images for the years 201006, 201007, and 201008. Each image is accompanied by its name, number, and coordinates. For example, the 201006 typhoon is named 'WIPAC' with coordinates (163.8, 11.8).
- Number of Typhoons:** This section provides statistics for the current year and the average number of typhoons from 1951 to 2000.
- Latest Typhoon Information:** This section lists recent typhoon events, such as 'Typhoon 201004 (LUPWOCK) (2010-08-30)' and 'Typhoon 201007 (KOPRACU) (2010-09-01)'. It also includes links to 'Typhoon News Topics', 'Local Information Portal', and 'AMeRAS (AMeRAS Ranking)'.
- Typhoon Database:** This section offers various search options, including 'Search by Metadata', 'Search by Date / Season', 'Search by Name / Number', 'Search by Map', 'Search by Place (Lat / Lon)', 'Search by Pressure', 'Search by Wind', 'Search by Day', 'Activity Calendar', and 'Typhoon Timeline'.
- Meteorological Satellite Images:** This section displays a satellite image of a typhoon over the Pacific Ocean, dated '2010-09-01 11:00 (UTC)'.

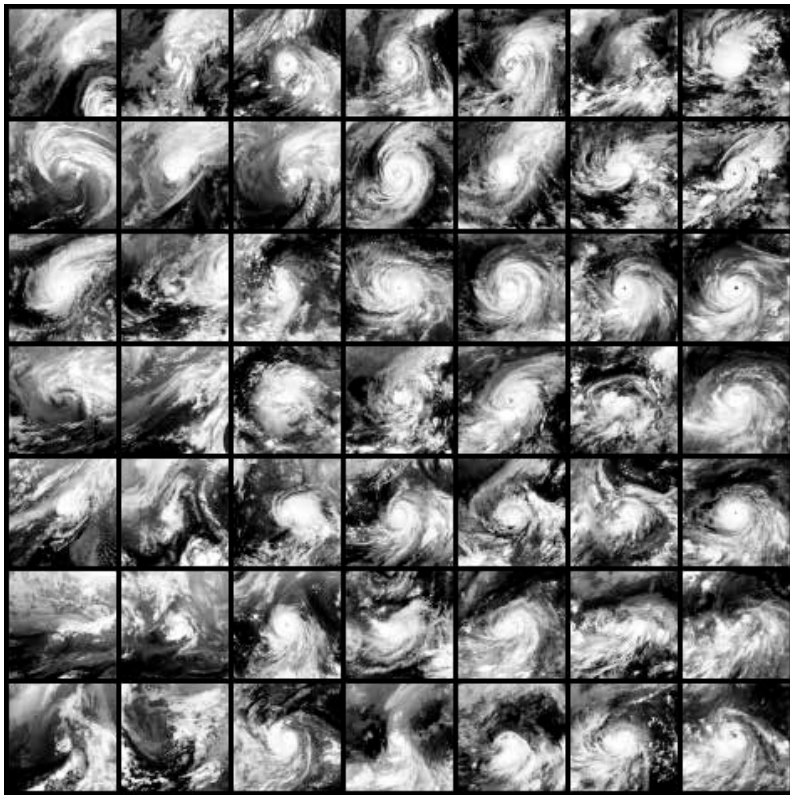
- Heterogeneous sources are integrated and indexed in real-time.
- Past data can be searched in the context of the current situation.
- Scientists and citizens use the websites for work, business, hobby.

One of the most famous typhoon information Website. About 200 million page views so far.

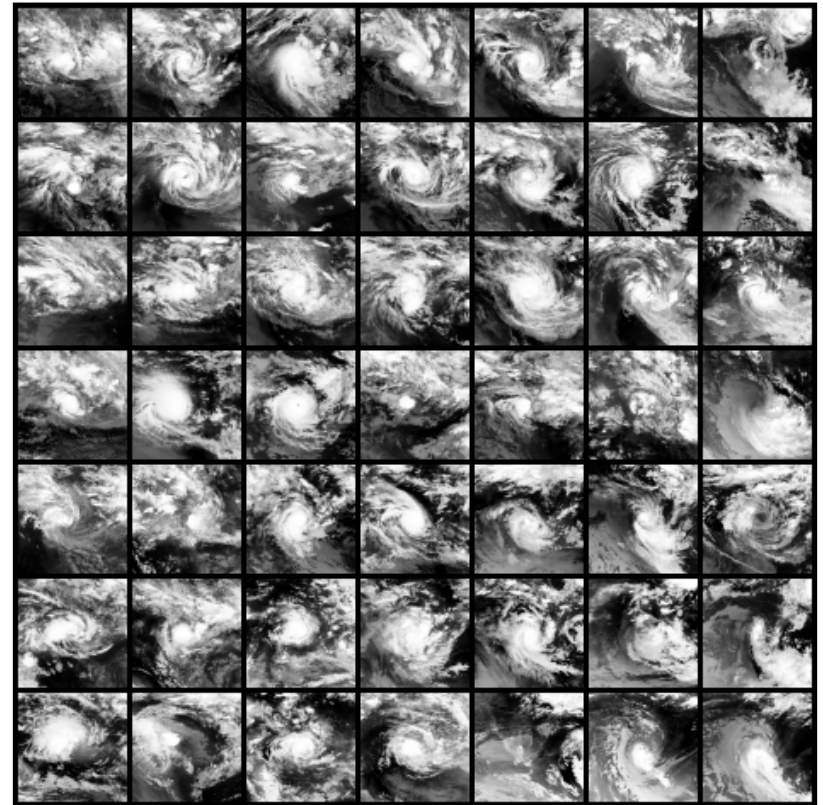


# Tropical Cyclone Image Collection

Since 1978, about 154,000 images for NH, and 35,100 images for SH.



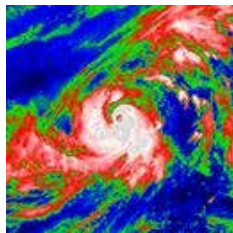
Northern Hemisphere



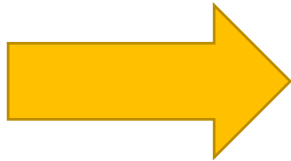
Southern Hemisphere

# Search by Situation

Decision making is often made in comparison to past events. Search related events to the current situation.



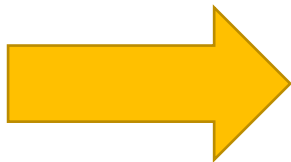
Satellite image



Find similar typhoon cloud patterns in the past.



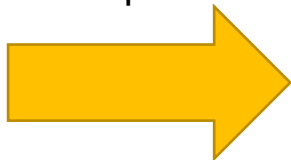
Online news



Find similar news articles or events in the past.

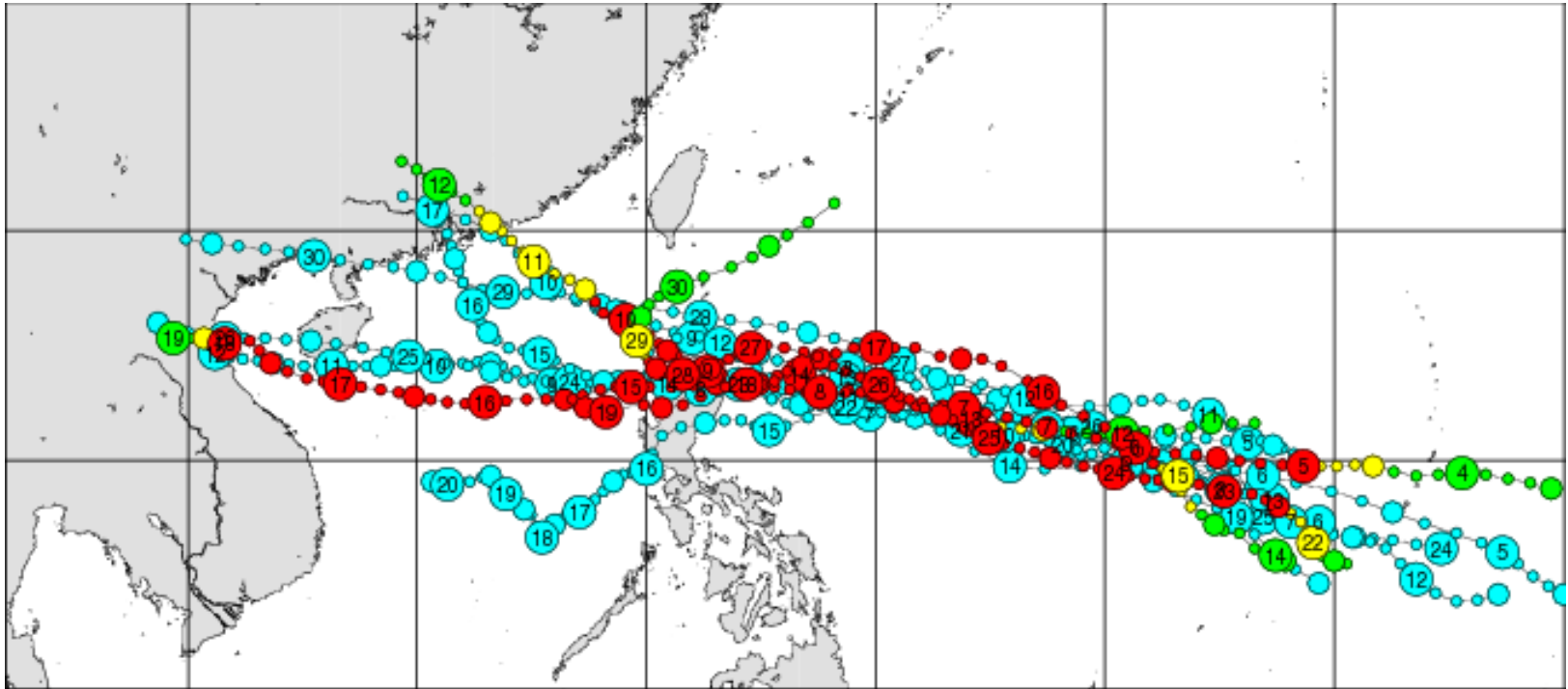


Rain pattern



Find similar precipitation patterns in the past.

# Search by Track Similarity



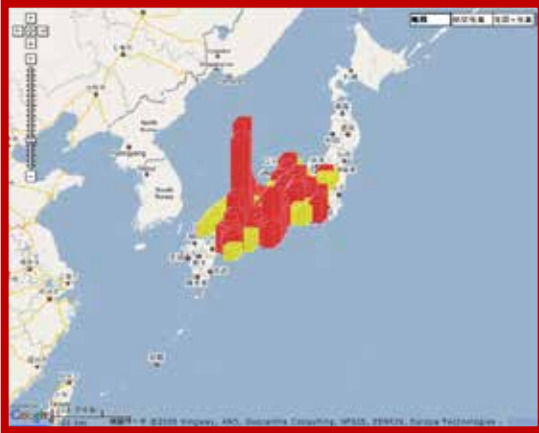
- Using dynamic time warping for evaluating similarity between tracks.

# Search by Image Similarity

Query 1	1	2	3	4
MTS110101803	GMS181031416	GMS492110508	GOE904101711	GMS595103114
201013 (WNP)	198101 (WNP)	199228 (WNP)	200423 (WNP)	199520 (WNP)
(N17.4, E122.6)	(N14.5, E161.4)	(N18.2, E134.8)	(N20.2, E130.2)	(N12.4, E130.9)
885 hPa / 125 kt	975 hPa / 60 kt	915 hPa / 100 kt	940 hPa / 85 kt	955 hPa / 80 kt
5	6	7	8	9
GMS491112714	GMS389042113	GMS597083013	MTS106102811	GMS502030313
199128 (WNP)	198902 (WNP)	199718 (WNP)	200619 (WNP)	200202 (WNP)
(N12.7, E143.5)	(N14.5, E148.5)	(N16.6, E138.6)	(N15.2, E126.5)	(N10.3, E135.1)
900 hPa / 115 kt	920 hPa / 100 kt	985 hPa / 50 kt	975 hPa / 65 kt	960 hPa / 75 kt
10	11	12	13	14
MTS109091718	GMS386051912	GMS179101518	GMS179051212	GMS501122111
200914 (WNP)	198603 (WNP)	197920 (WNP)	197904 (WNP)	200125 (WNP)
(N22.5, E139.3)	(N11.6, E156.2)	(N18.9, E129.4)	(N11.0, E120.2)	(N10.5, E157.0)
945 hPa / 85 kt	910 hPa / 120 kt	925 hPa / 100 kt	1000 hPa / 0 kt	965 hPa / 70 kt

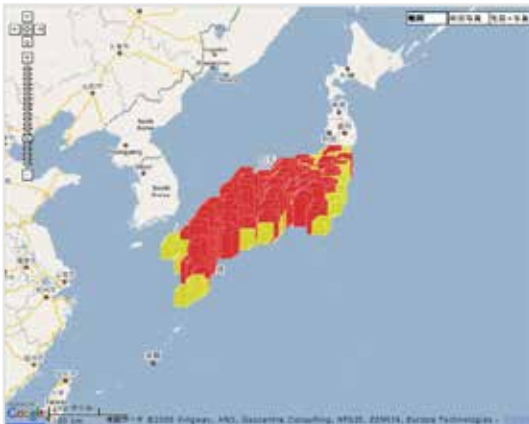
- **Content-based image retrieval**: search similar images to the query image.
- **Similarity is based on image features**; currently PCA, but ideally more complex.

# Search by Rainfall Similarity

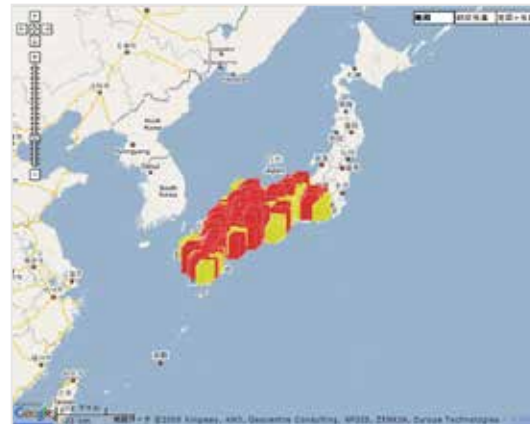


1998/09/20-1998/09/25

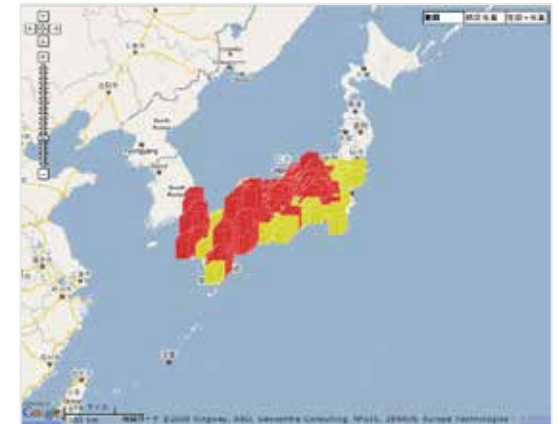
- Use a spatial pattern as a query, and retrieve other patterns.
- Similarity is based on the rainfall amount and distance.



2004/10/18-2004/10/21



1998/10/14-1998/10/18



1983/09/25-1983/09/29

2015/10/21

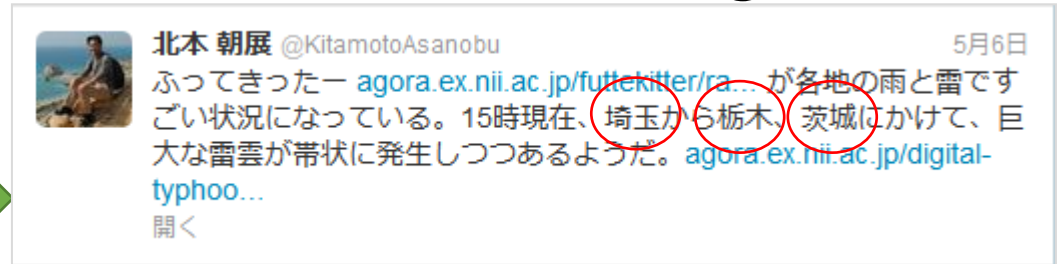
SPARC Japan Seminar 2015

13

# Social Weather Monitoring



Search by  
keyword



Social data

Toponym-based geotagging  
by GeoNLP

(time, latitude, longitude, situation)

Comparison of social and scientific  
streams in terms of precipitation



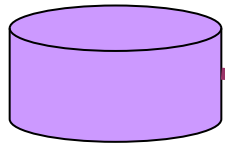
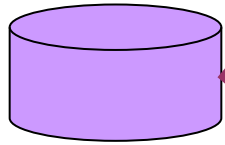
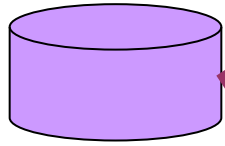
Scientific data



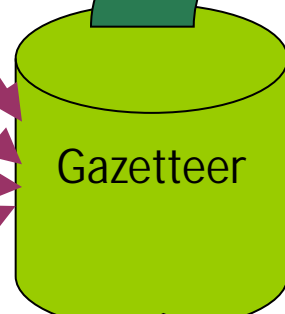
- Precipitation is easier to observe by the eye than weather in general.
- Weather radar can provide objective data or "ground truth."

# GeoNLP Framework

Public resources for placenames

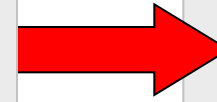


Collaboration  
↑  
Registration



Morph. Anal.  
↓  
Test Sample

Text or HTML for input into  
JSON-RPC API



Natural Language



Morphologic. Analysis



Pattern Matching



Placename Extraction

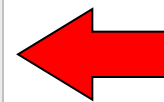


Placename Resolution

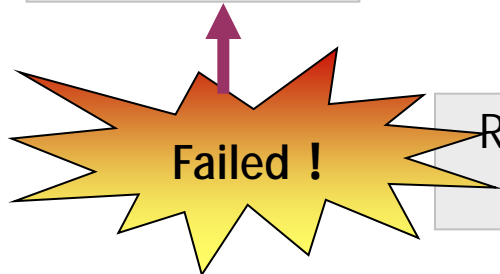


Metadata Annotation

GeoNLP Server

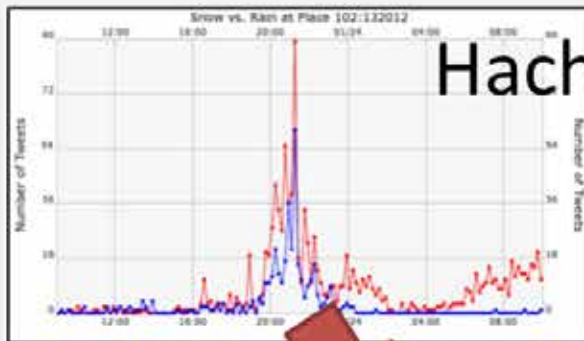


Response in JSON format or  
use it as CMS modules

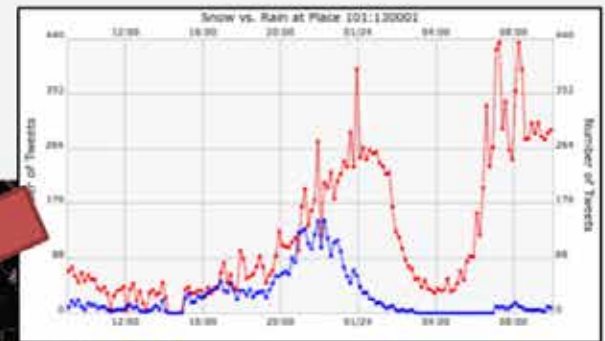


Update costs

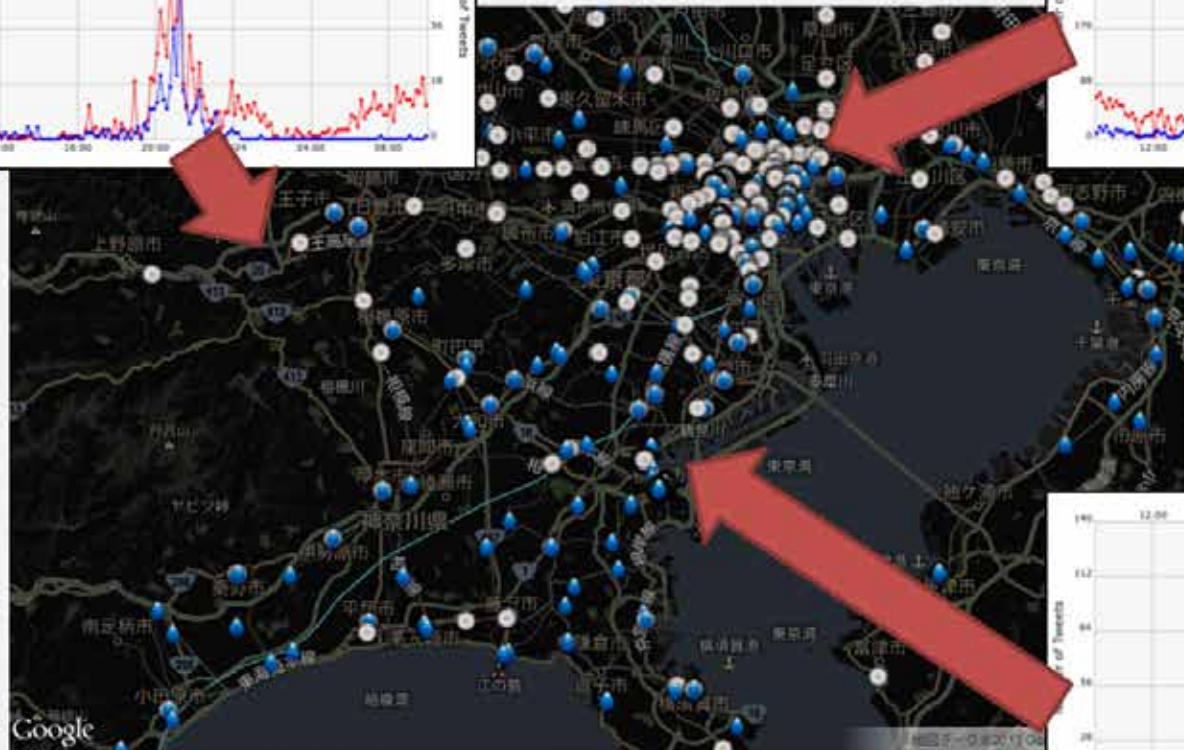
# Snow (white/red) and Rain (blue)



Hachioji



Tokyo



Yokohama

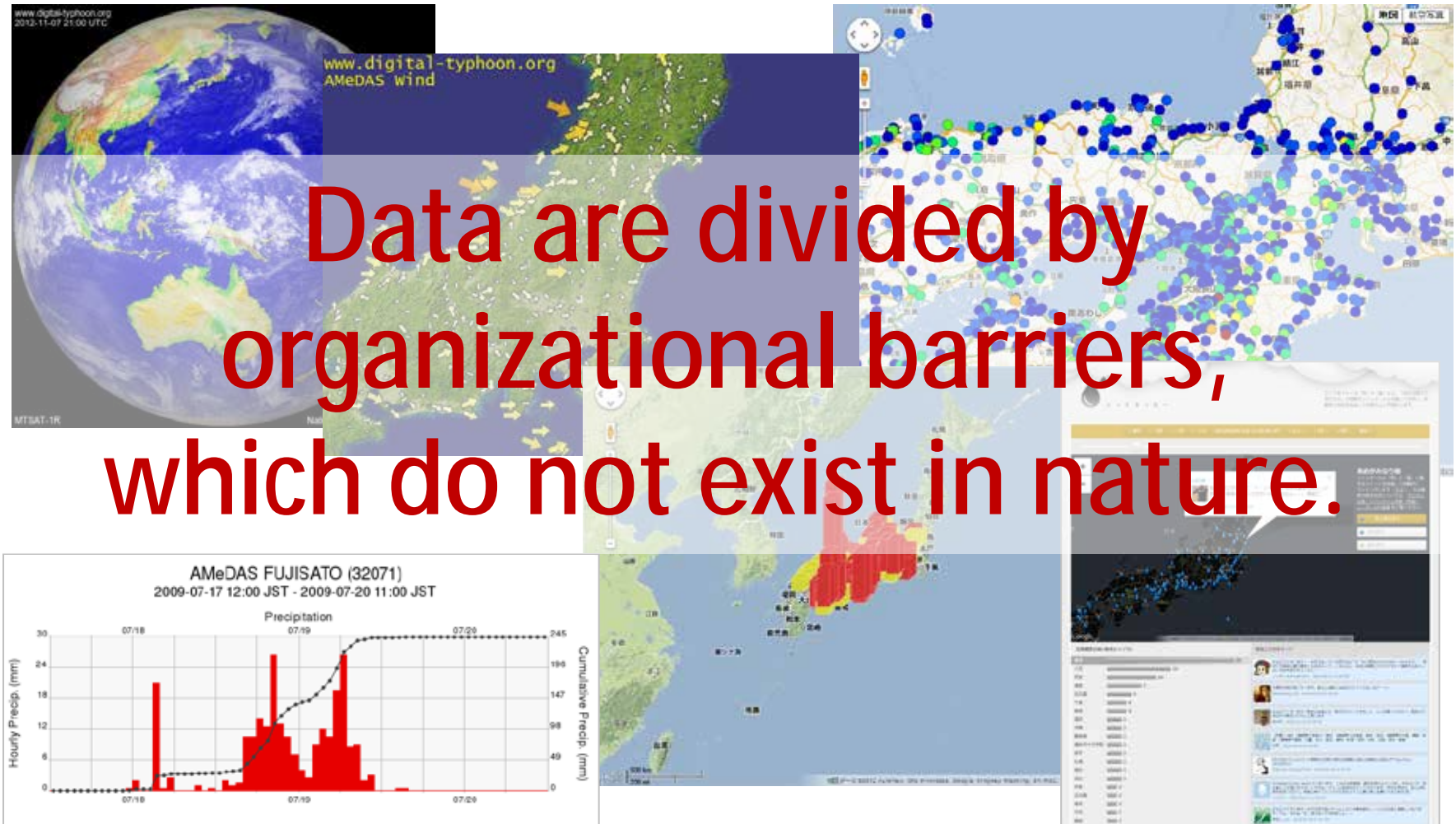


Jan. 23, 2012

Futtekitter: <http://agora.ex.nii.ac.jp/futtekitter/>



# Barriers between Organizations

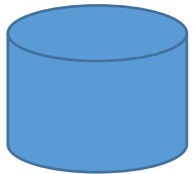


Data are divided by organizational barriers, which do not exist in nature.

# Barrier-Free Infrastructure

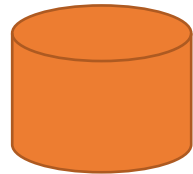
## Exclusive Design

Scientists



Professional  
Infrastructure

Citizens



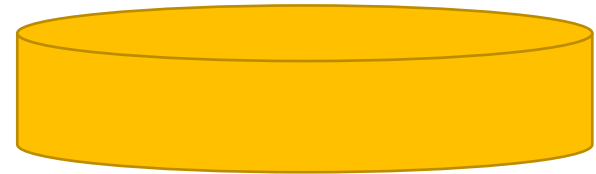
Outreach  
Infrastructure

## Inclusive Design

Scientists



Citizens



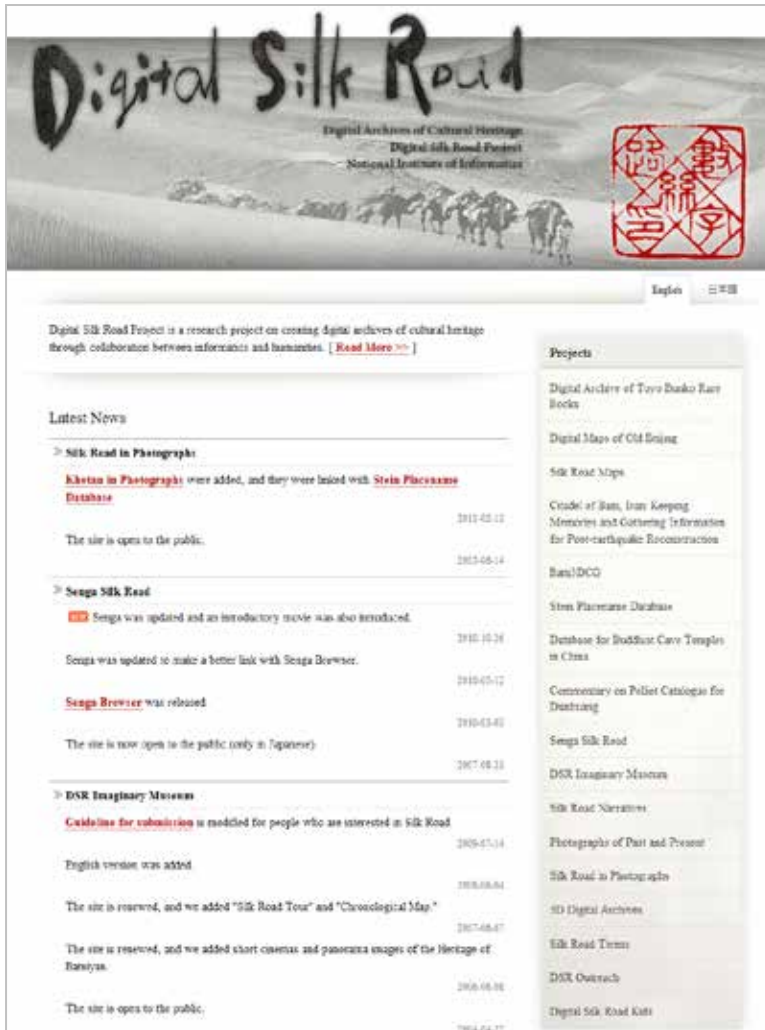
Universal Infrastructure with  
better experience for everyone

# Case Study: “Digital Silk Road” (Humanities)

# Digital Silk Road

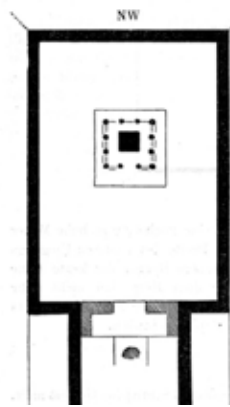
<http://dsr.nii.ac.jp/>

- Started in 2001.
- **Digital Humanities:** Collaborative work among informatics + humanities scholars.
- Databases and digital resources are **publicly accessible on the Web.**



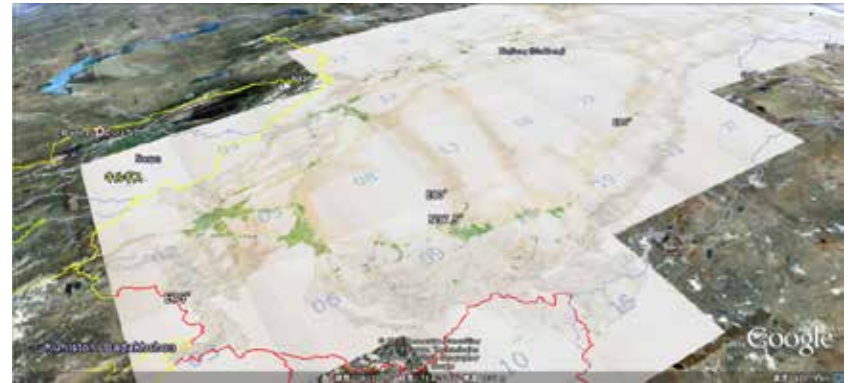
# Variety and Heterogeneity of Data

## Text



die obere sich wie eine in eine niedrigere 3,10 m tiefe Plattform eingepasste Bank darstellt (auf der Skizze schraffiert) und die Mitte offen liest. Vor dieser großen Unterstufe liegt der Rest eines mächtigen Sockels, in welchem ein tiefes Loch sich zeigt: hier hat also wohl eine große Statue oder eine Fahne gestanden. 12 m nach innen zu vom S-Rand der Plattform des Hauptbaues, 5,50 m von den Seitenmauern und 7 m vor der Rückmauer, erhebt sich eine niedrige, 8 m ins Geviert betragende Stufe, auf deren Mitte ein jetzt zerstörter, 2 m großer, viereckiger Sockel steht; um diesen Sockel geht ein Gang herum, vorne und an den Seiten je 1,50 m breit, hinten aber nur 90 cm breit. Dieser Umgang ist nach außen von einer Mauer umgeben, welche durch zwölf kleine Stäben in kleine Abteile geteilt ist, von denen der mittlere der Frontseite den Eingang bildet. Auf der Rückseite ist dies aus zwei Eck- und zwei Mittelstüben bestehende System sehr zerstört. Vor den sechs Interkolumnien der Seiten und den zwei Interkolumnien neben dem Eingang sind je noch Sockel für Statuen erhalten: nach mancherlei dekoratives Bei-

## Map



## Photograph

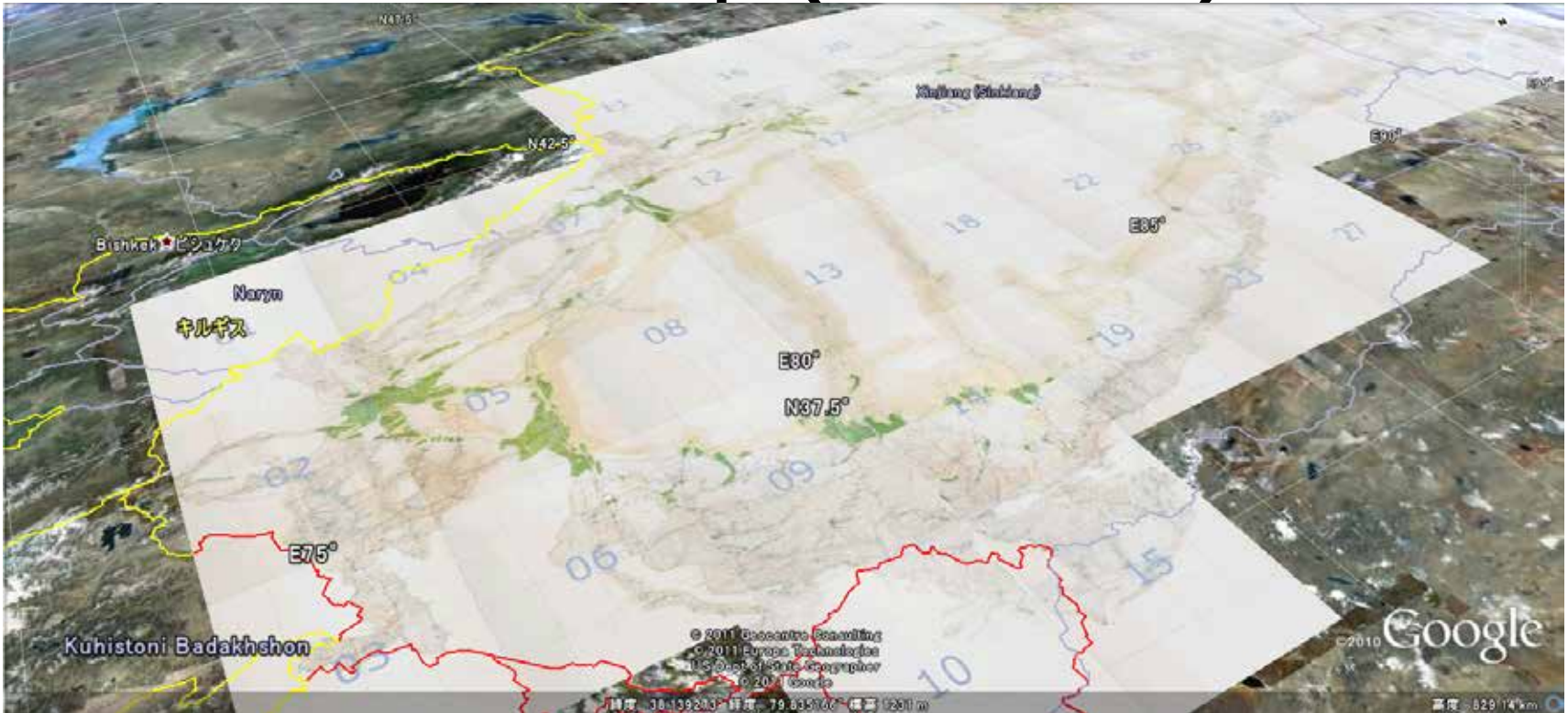


Y  
1904(Le Coq, 1913, Tafel. 70, I)

## Gazetteer

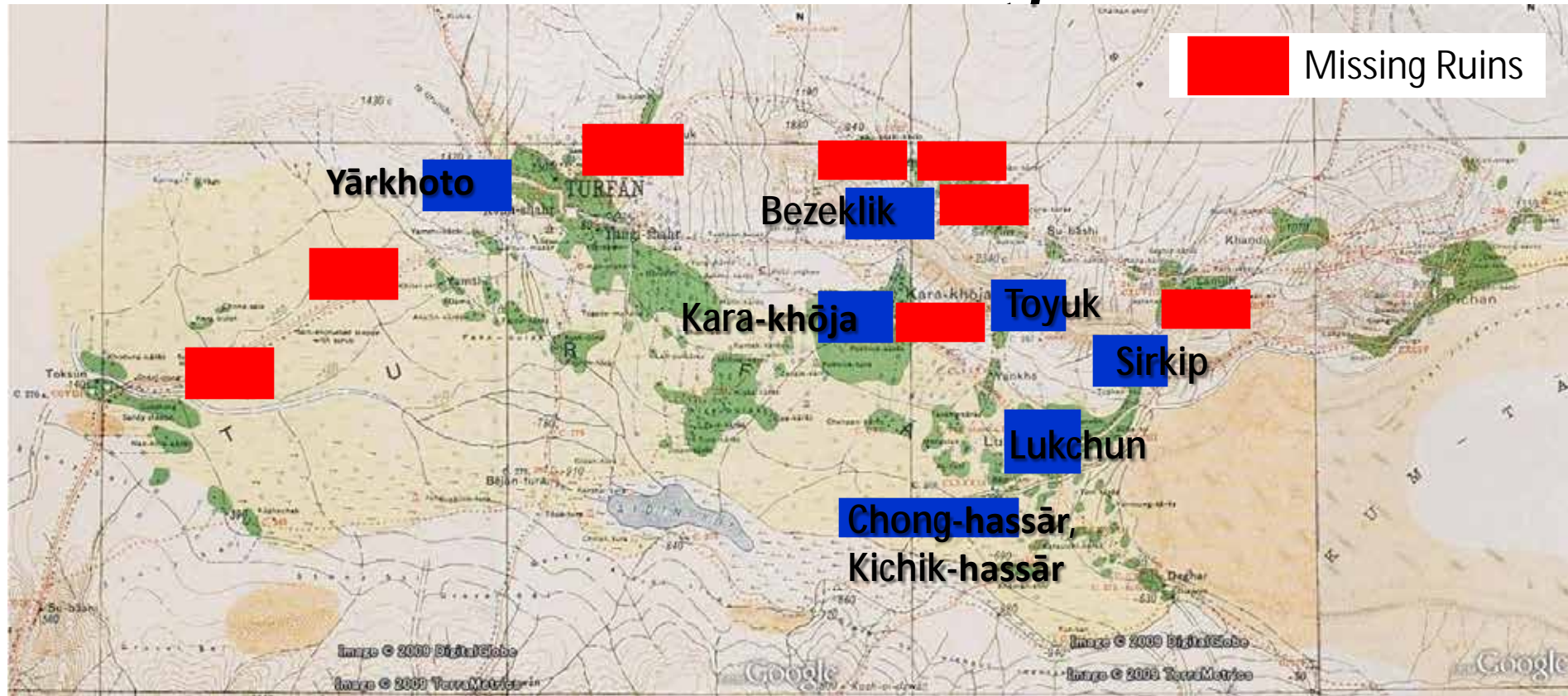
Abāb-langar, habit., 14. B. 3.	Aechhik-bulak (of Turfān), spring, 28. B. 4.
Abād (of Ak-su), market-town, 12. A. 3.	Aechhik-bulak (of Yai-döbe), spring, 4. C. 4.
Abād (of Kara-yulghun), vill., 12. B. 1.	Aechhik-daryā, river, 21. A. 2.
Abād (of Karghalik), vill., 5. C. 4.	Aechhik-dawān, pass, 9. B. 3.
Abād (of Kāshgar), vill., 5. A. 2.	Aechhik-jilga (of Duwa), valley, 9. B. 3.
Abād (of Turfān), vill., 28. C. 3.	Aechhik-jilga (of Kara-tāsh), valley, 2. D. 3.
Abād (of Yārkand), vill., 5. C. 2.	Aechhik-jilga (of Khotan), valley, 9. C. 3.
Abād-jilga, valley, 12. B. 2.	Aechhik-jilga (of Sampula), valley, 14. A. 3.
Abdal, vill., 30. B. 2.	Aechhik-jilga (of Tawak-kél), loc., 14. A. 1.
Abdalkash-mazār, shrine, 14. C. 3.	Aechhik-köl, lake, 15. D. 1.
Abdul-ghafūr-langar, loc., 10. C. 1.	Aechhik-kuduk (of Kapa), well, 23. A. 1.
Abdul-rahmān-jilga, valley, 9. A. 4.	Aechhik-kuduk (of Kuruk-tāgh), well, 28. C. 4.
Abshak-bēl, Pass, 2. B. 1.	Aechhik-kuduk (of Marāl-bāshi), well, 5. D. 2.
Ach-tāgh, hill and vill., 7. C. 2.	Aechhik-otan, loc., 7. C. 2.
Acha-dong (of Chizghān), hill, 19. C. 3.	Aechhik-su, loc., 31. A. 4.
Acha-dong (of Yārkand R.), loc., 7. D. 4.	Aechhik-tügemen, loc., 5. D. 2.
Acha-kuduk, loc., 7. D. 4.	Achi-tāgh, hill, 32. B. 1.
Acha-shipang, loc., 22. D. 4.	Achik-aghzi, loc., 9. D. 3.
Achak-aghzi, loc., 5. A. 4.	Achma (of Hanguva), vill., 14. A. 2.
Achal (of Ak-su), vill., 12. A. 3.	
Achal (on Charchak R.), loc., 21. C. 2.	

# Stein Map (Silk Road)



- Stein's map "Innermost Asia" was registered and displayed on Google Earth satellite images.

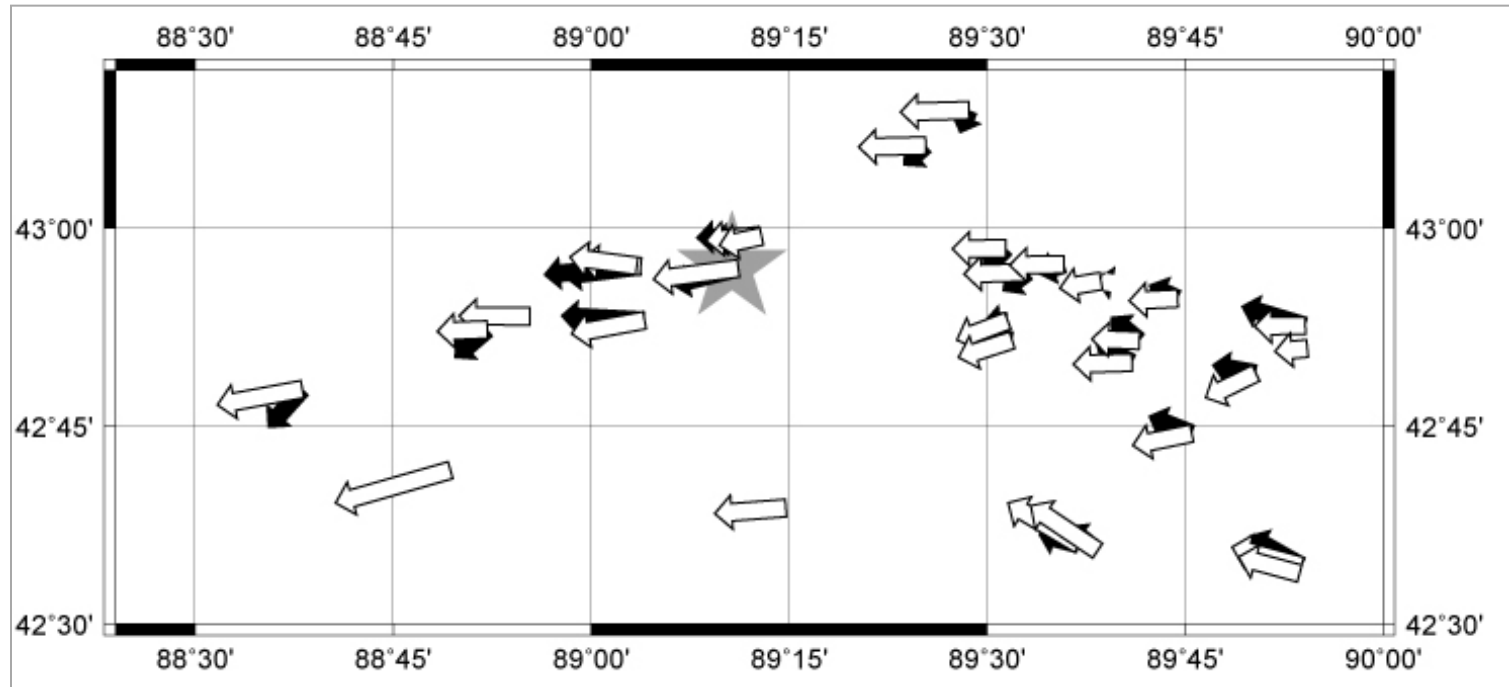
# Problem of "Missing" Ruins



Oi-tam, ruined fort  
 Bögan-tura  
 Buluyuk ( Shipang, Sassik-bulak, Kazma )  
 Murtuk-ruins

Yoghan-tura  
 Chikkan-köl  
 Bedaulat's town, Bēsh-kāwuk, Kosh-gumbaz  
 Yutōgh

# Error Distribution in Turfan

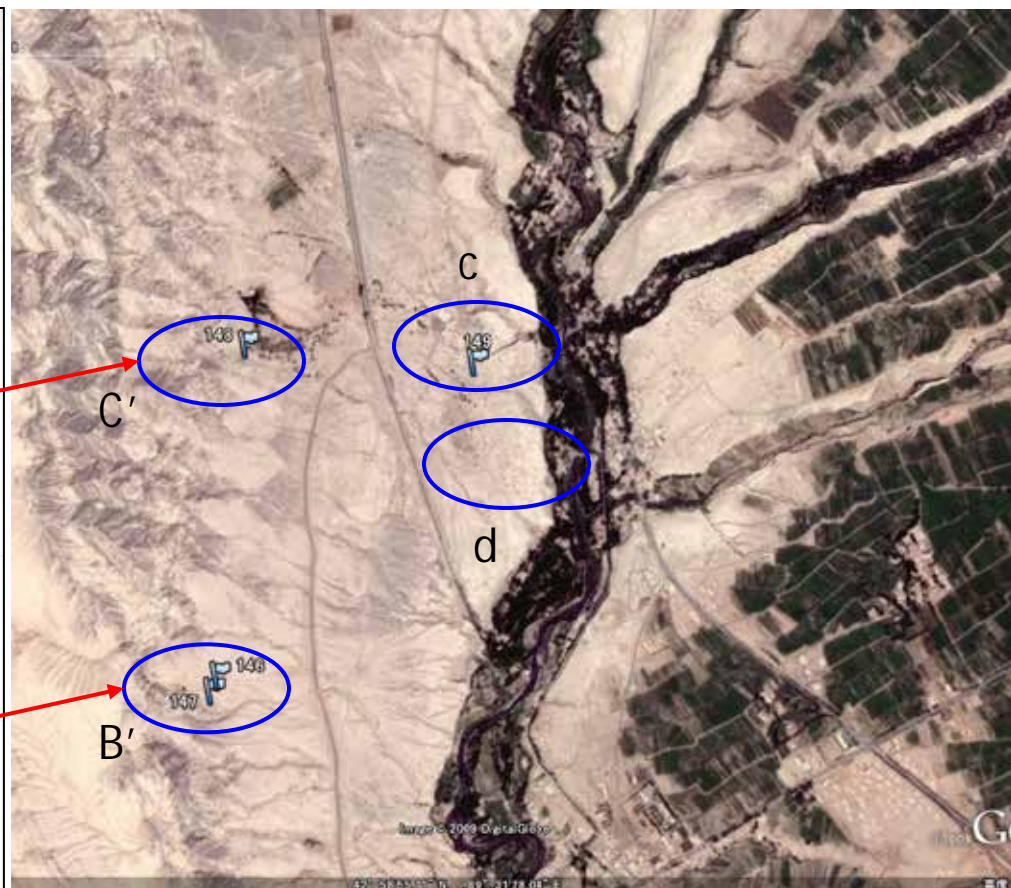
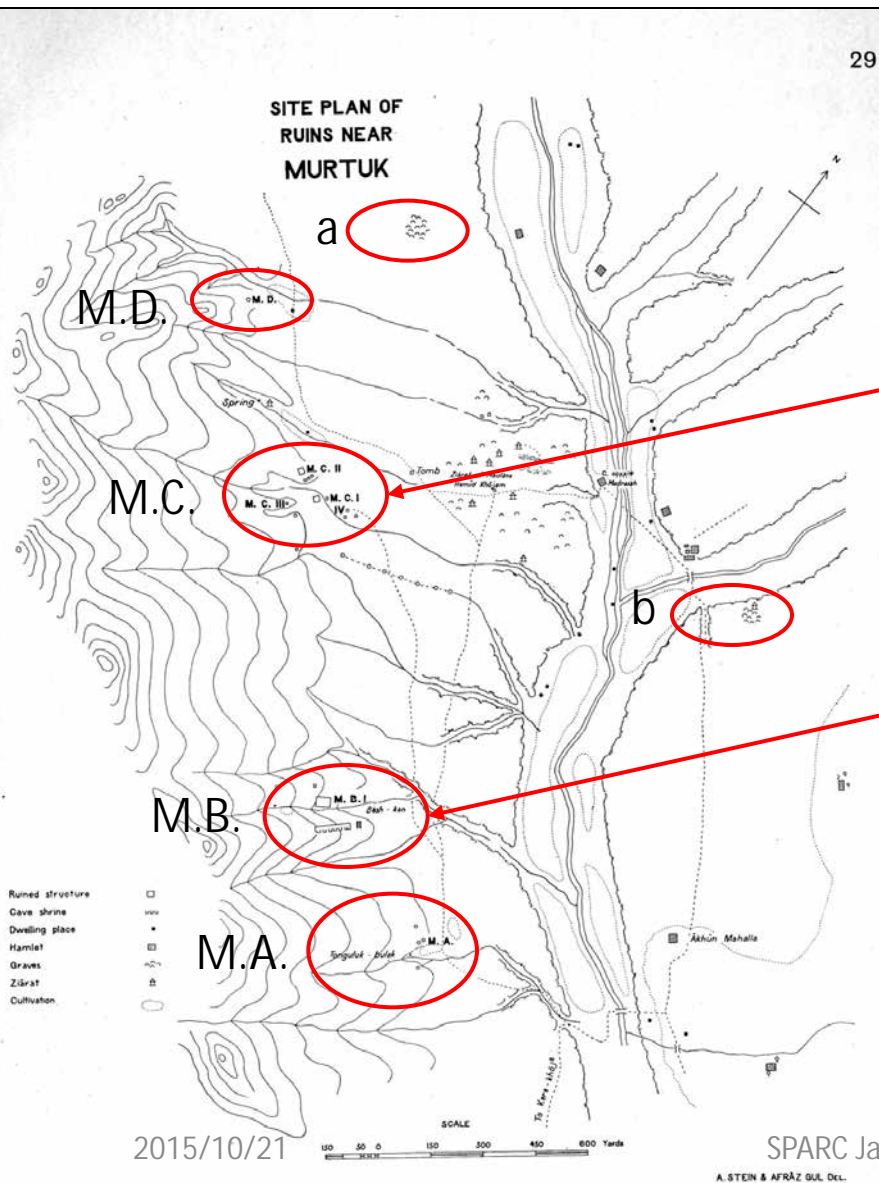


Error Distribution in Turfan Basin / White: Innermost Asia / Black: Serindia

- Some ruins were reported by 20<sup>th</sup> expeditions, but are missing in recent survey reports.



# Matching Entities



Stein's map and satellite images for the same area. Each source reports different ruins due to different conceptualization.

# Barriers between Communities

- Our research was criticized by humanities scholars as “not understandable” or “too different from our traditional approach.”
- Concept of the approach is changed from technical to humanistic viewpoints for better communication across communities.
- Some “early-adopters” quickly understood our concept, although some “laggards” are very slow to accept our approach...

# Barrier-Breaking Research

## Ordinary Results

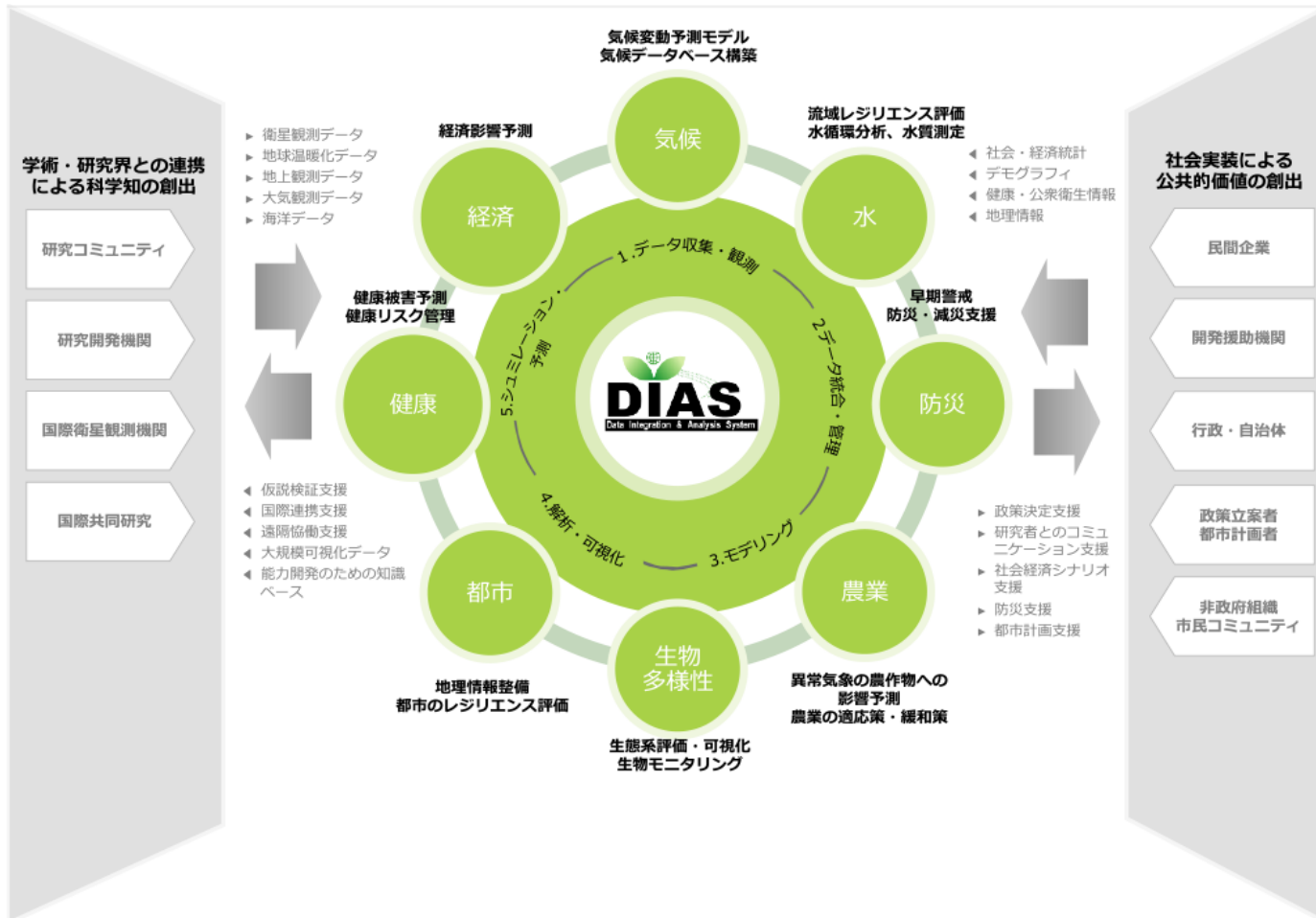
- Take data prepared by another community.
- Apply existing methods that you are already familiar with.
- Obtain results which are already known, or trivially correct.

## Innovative Results

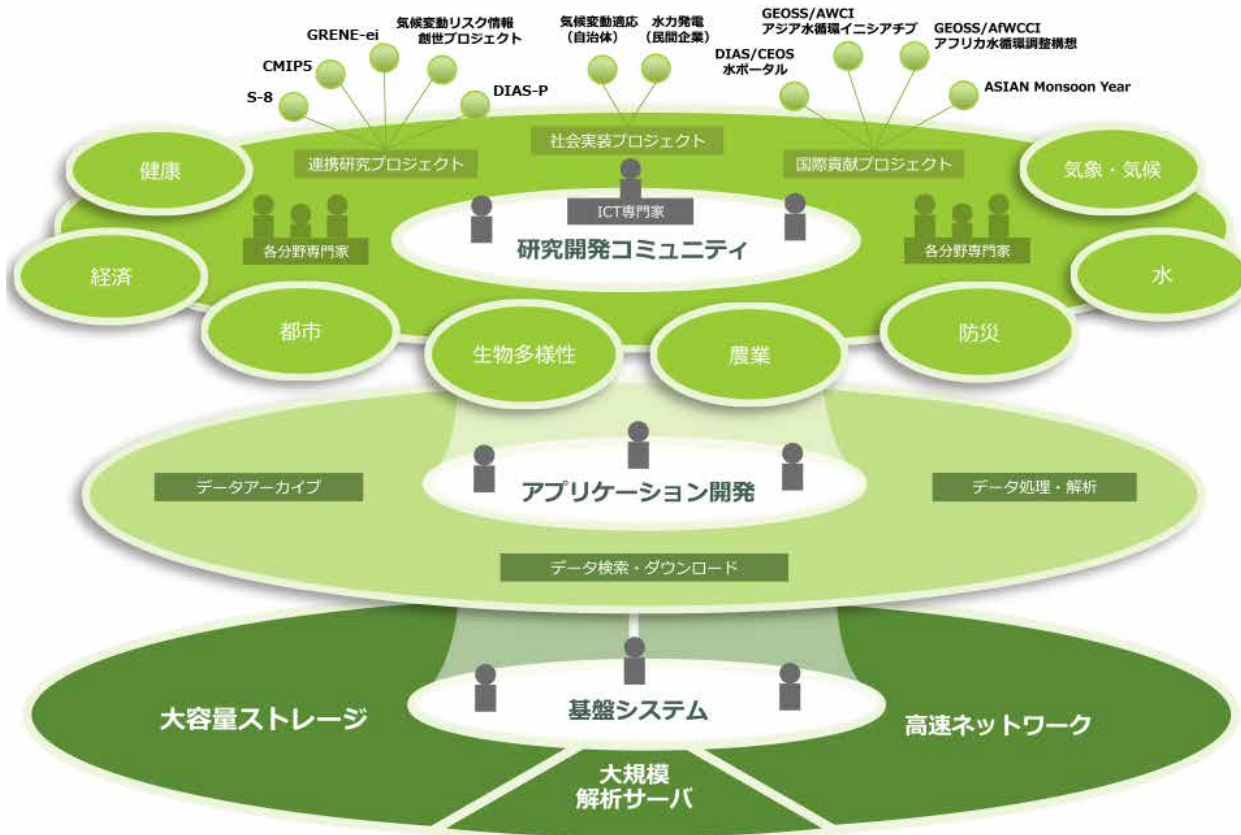
- Deepen knowledge about the problem.
- Propose a new concept for innovating the viewpoint of research.
- Obtain results which cannot be realized without collaboration.

# Forces to Break through Barriers

# DIAS : Data Integration and Analysis System



# DIAS = Data + Community

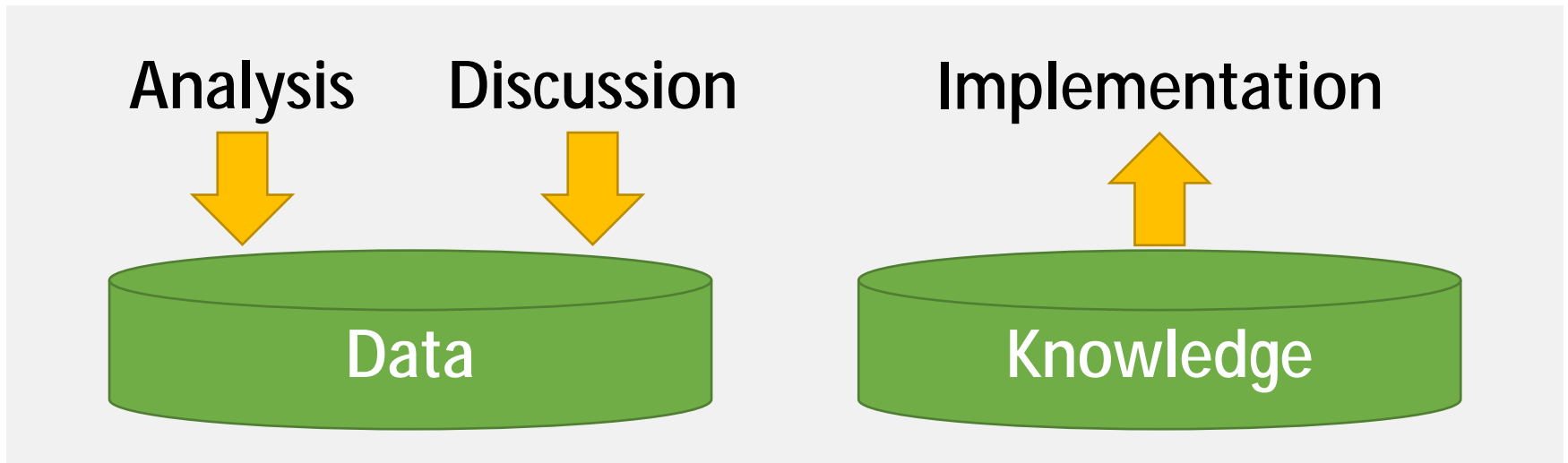


Community  
Infrastructure

Data  
Infrastructure

<http://www.diasjp.net/about/system/>

# Data Sharing or Knowledge Sharing



- **Data = inflow**. Analysis and discussion on the same data is shared across communities.
- **Knowledge = outflow**. Implementation does not require interaction across communities.

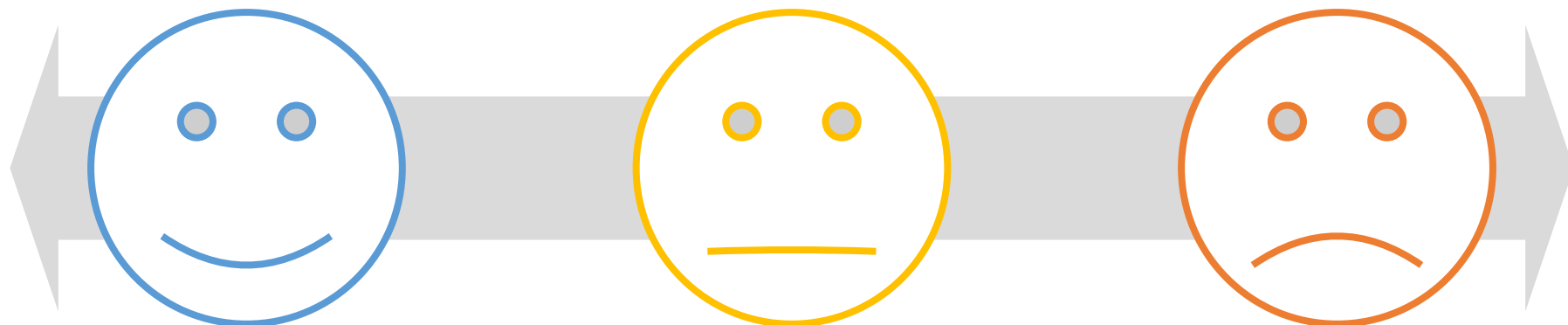
# Infrastructure's Gravitational Field

- What is the attractive force of research infrastructure for data and community?
- **Long-term sustainable operation** accumulates the mass of dependability.
- **Large data mass** indicates the value of adding new data and joining the community.
- **Increased gravity leads to positive feedback** to attract more data and community.



# Barriers In You

# Mission and Openness



Government

Research

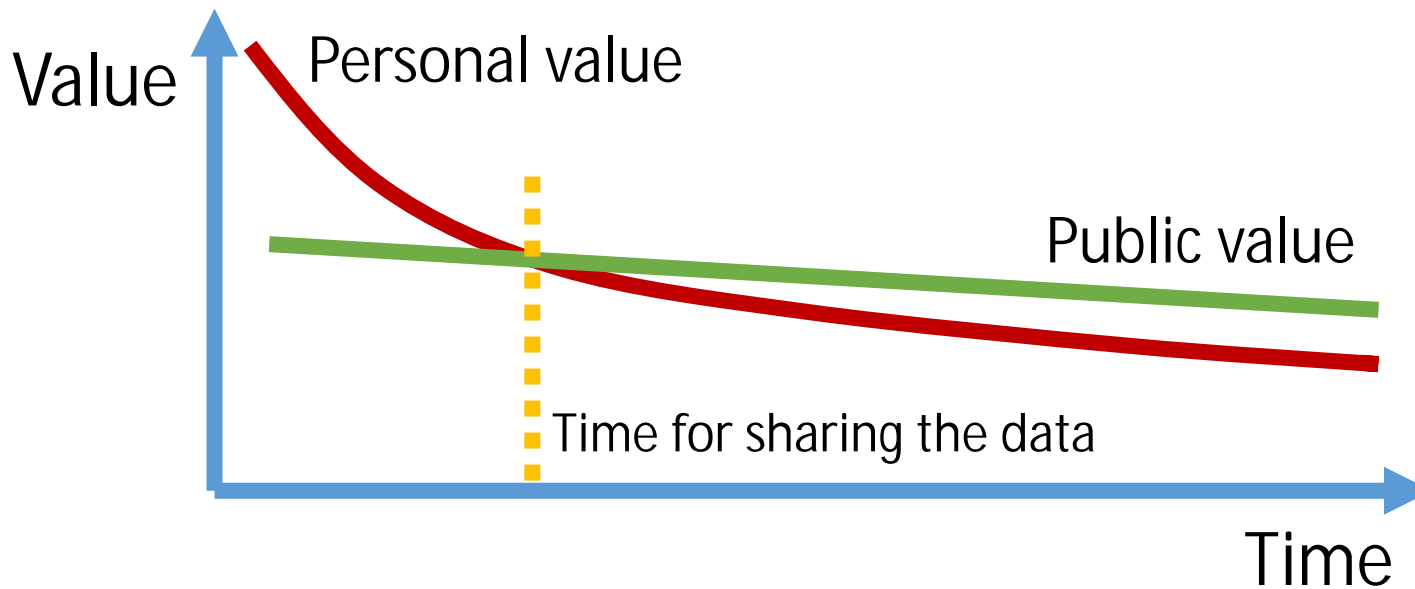
Business

Mission:  
public benefit  
= open by  
default

Mission:  
innovation =  
eventually  
open

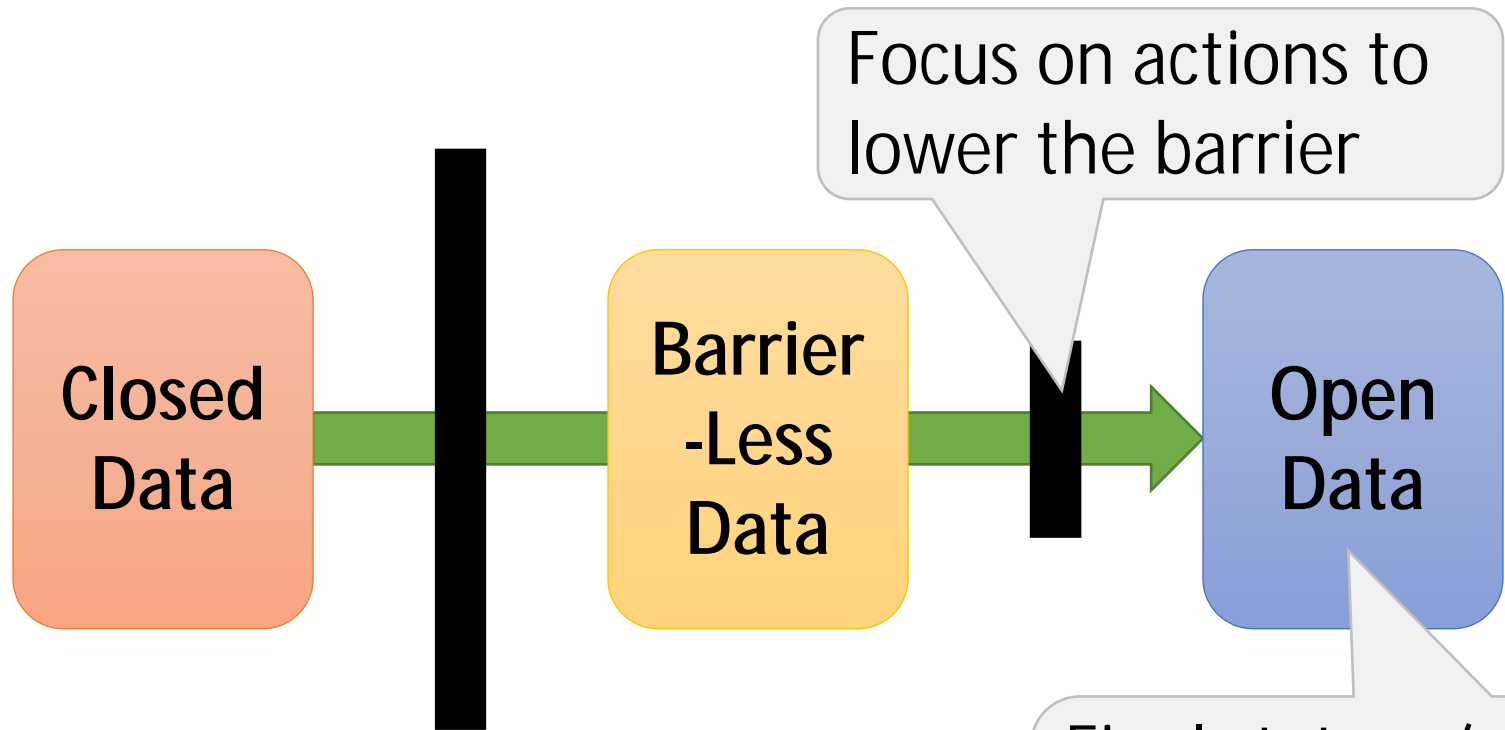
Mission:  
(internal) profit  
= open by  
strategy

# Eventual Openness of Data



- Raise public value by **increasing citation, evaluating open innovation, reducing data management cost.**
- At the end of your life, you want to share the data!

# Barrier-Less Data



**We welcome ANY actions to lower the barriers!**

# Are You Ready for Actions?



Yes, I am ready!



No, I am satisfied with the current system.

- Living within barriers is comfortable. **Is there any advantage in breaking the barriers?**

# Research across Barriers

- Barriers between you and others.
  - Barriers between organizations.
  - Barriers between research communities.
  - Barriers between scientists and citizens.
- 
- Share and discuss over research data is a good way to promote collaboration.
  - **Stay hungry and foolish to break the barriers!**

# Related Resources

- Digital Typhoon
  - <http://agora.ex.nii.ac.jp/digital-typhoon/>
- Digital Silk Road
  - <http://dsr.nii.ac.jp/>
- Open Science
  - <http://agora.ex.nii.ac.jp/~kitamoto/research/open-science/>
- Researchmap
  - <http://researchmap.jp/kitamoto/>