「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

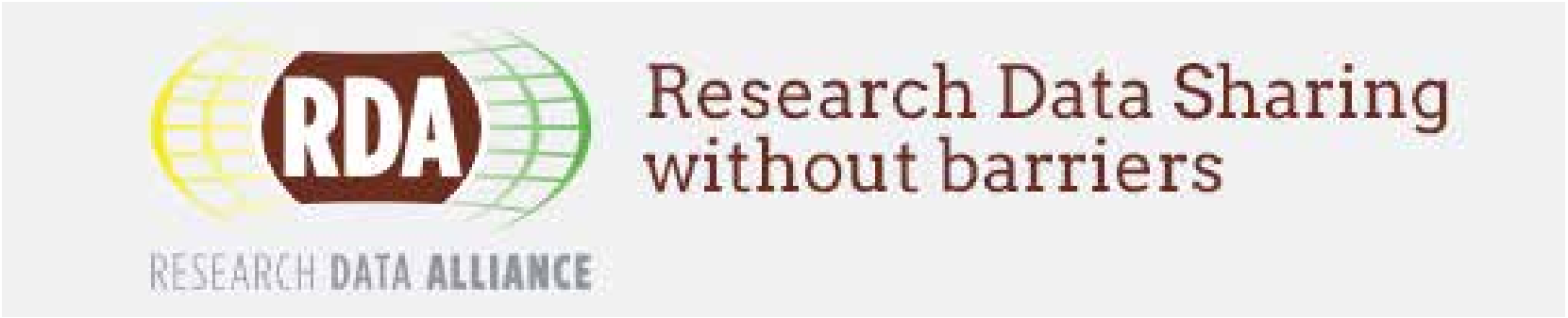
Asanobu KITAMOTO

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



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.

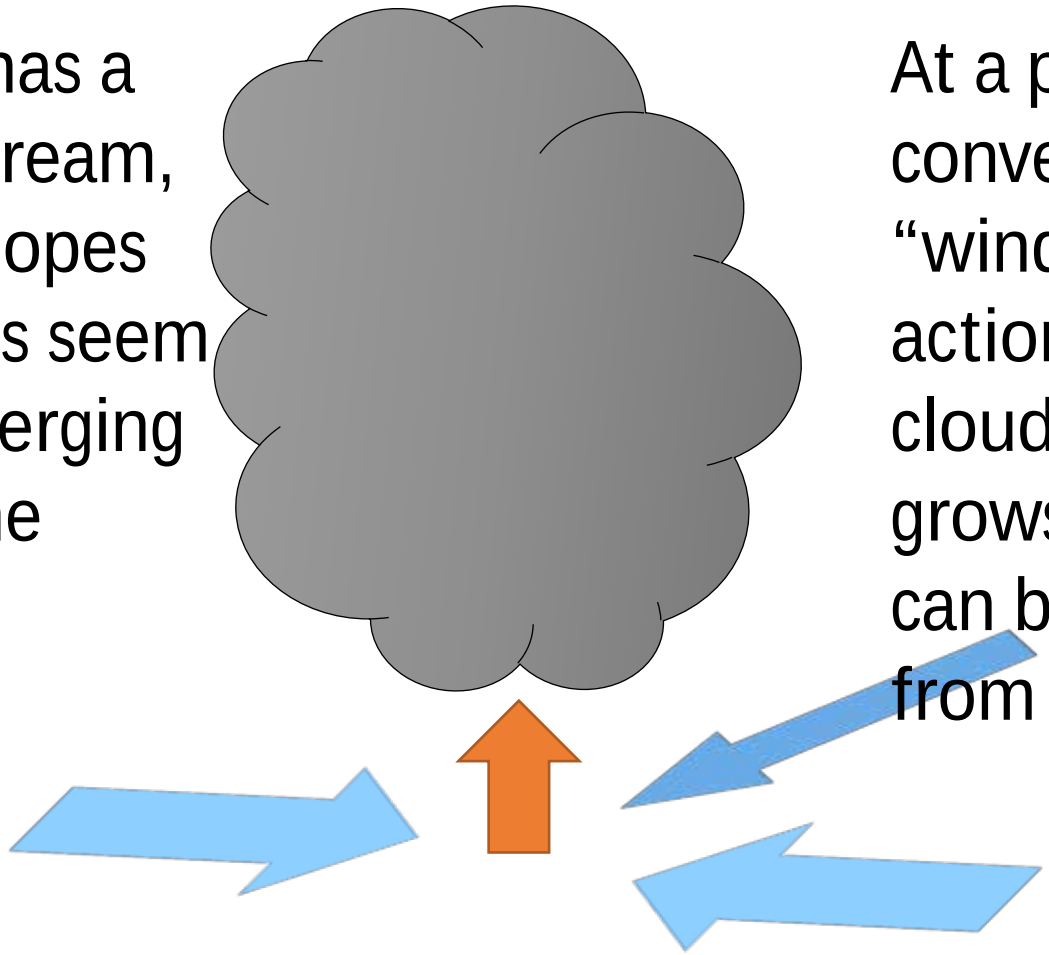
Details in Another Slide

**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.

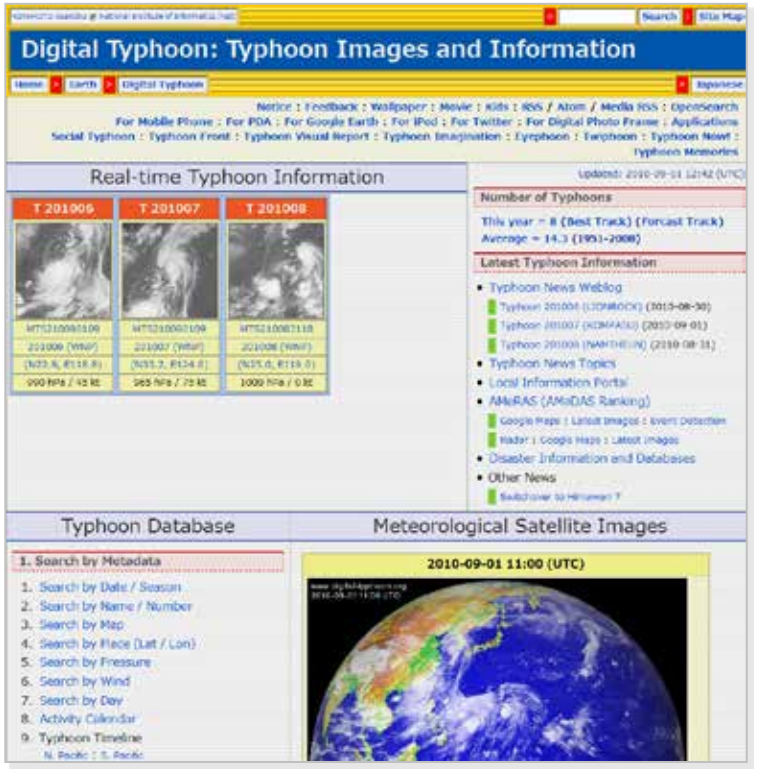


At a point of convergence, “winds” of actions grow tall clouds. As it grows higher, it can be seen from a distance.

Case Study: “Digital Typhoon” (Science)

Digital Typhoon

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

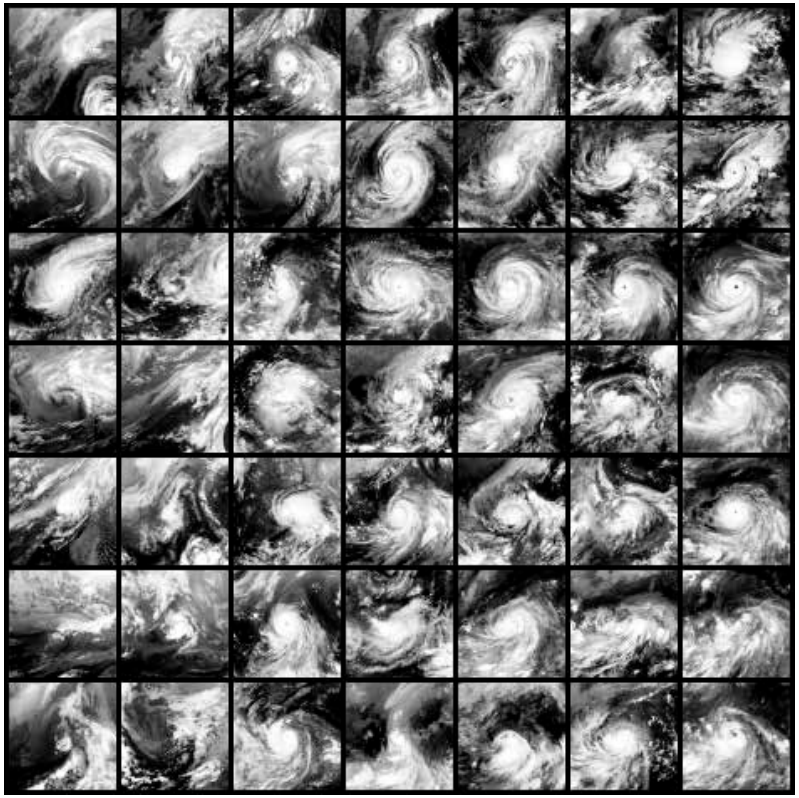


One of the most famous typhoon information Website. About 200 million page views so far.

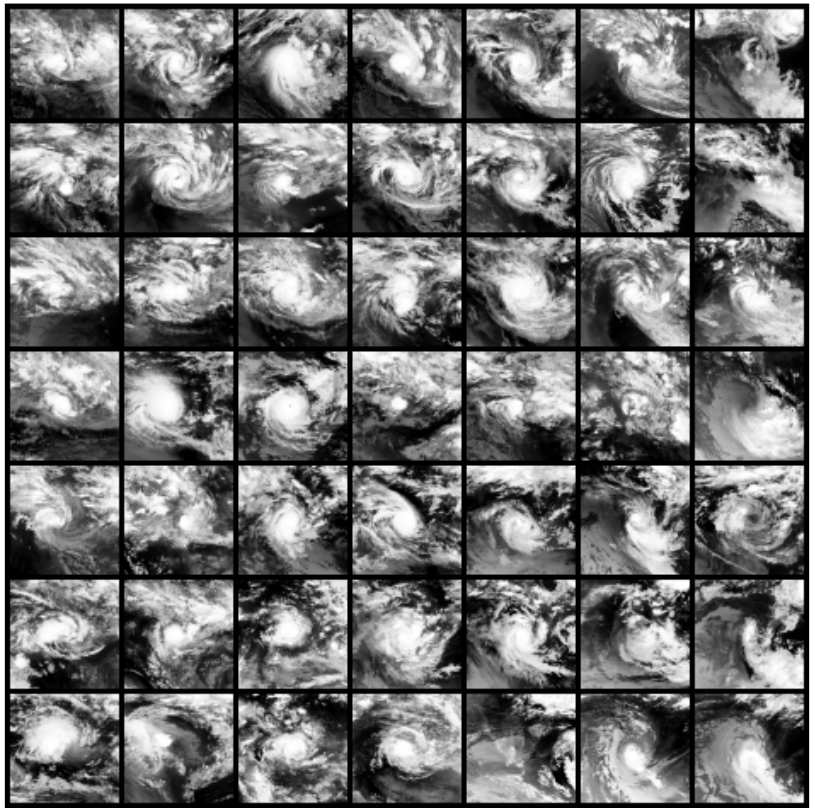
- Heterogeneous sources are integrated and indexed in real-time.
- Past data can be searched in the context of the current situation.
- Scientists and citizens use the websites for work, business, hobby.

Tropical Cyclone Image Collection

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



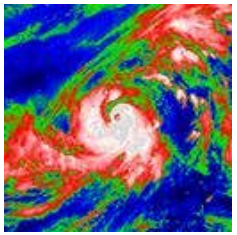
Northern Hemisphere



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.

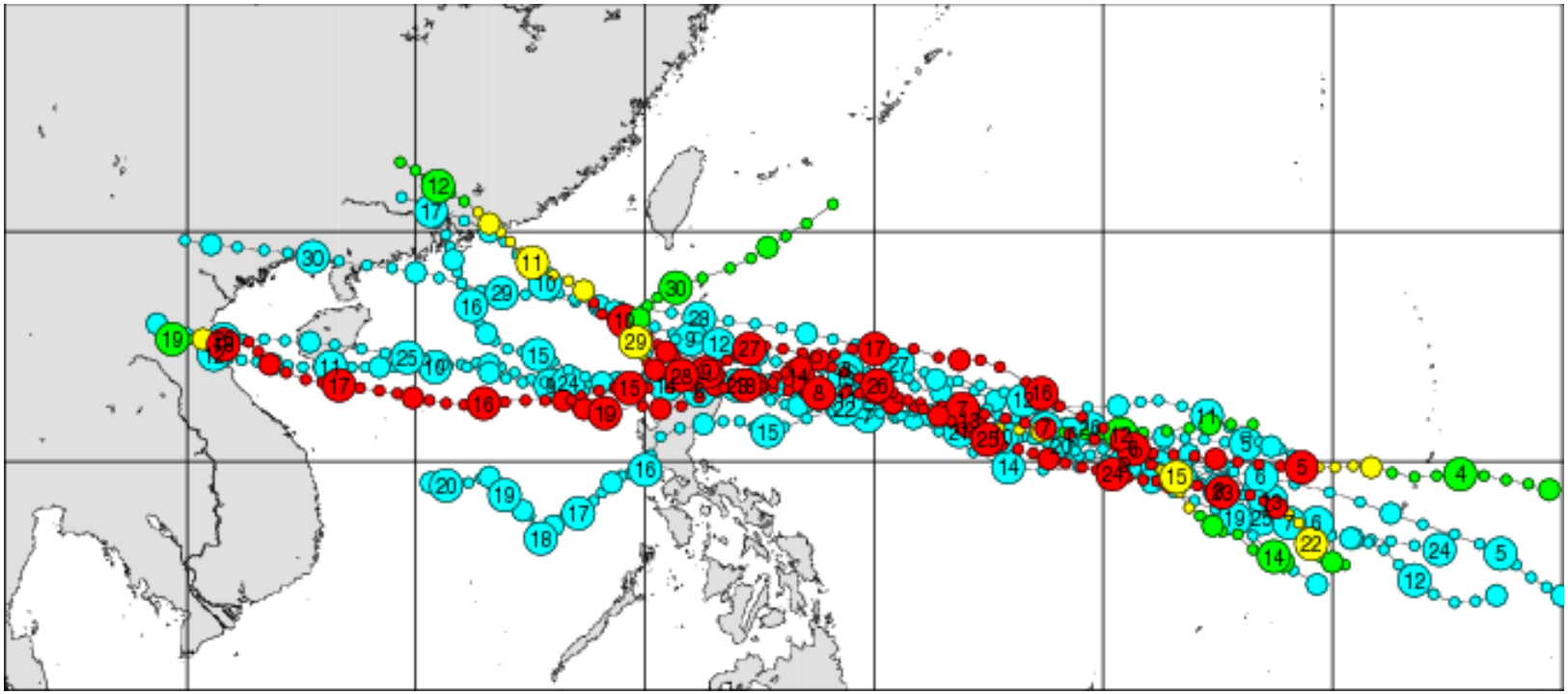
Online news

Find similar news articles or events in the past.

Rain pattern

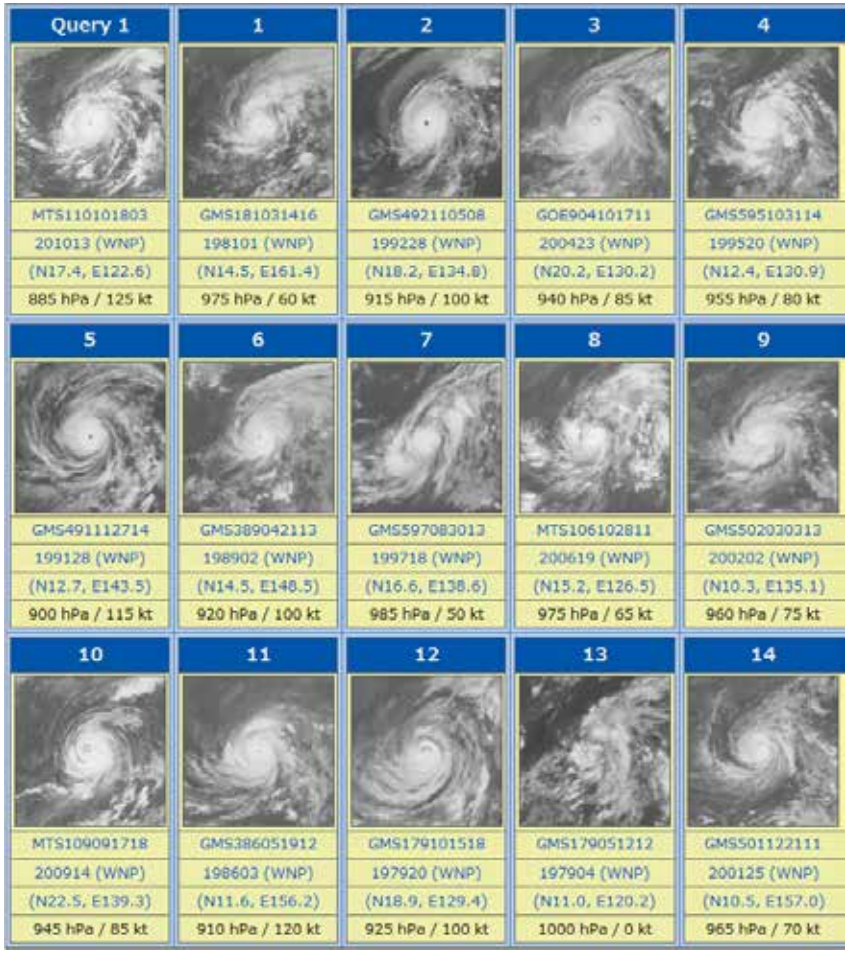
Find similar precipitation patterns in the past.

Search by Track Similarity



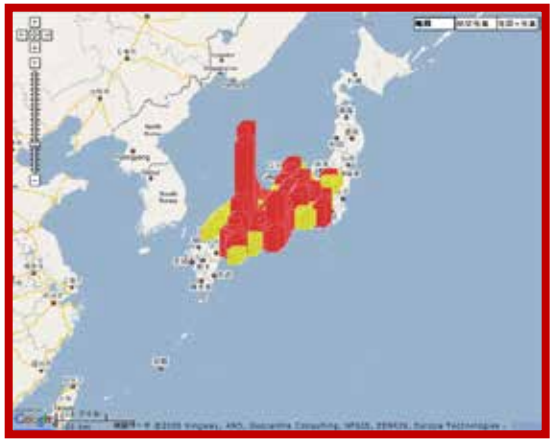
- Using dynamic time warping for evaluating similarity between tracks.

Search by Image Similarity



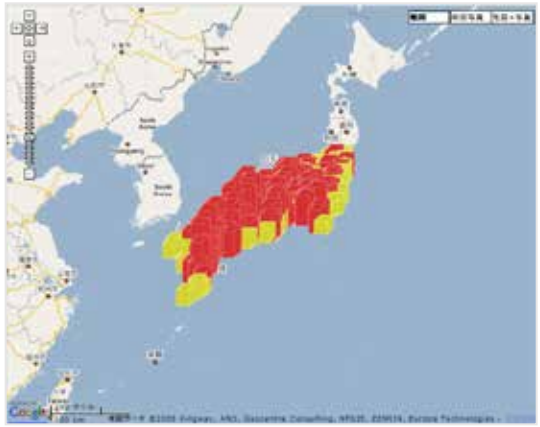
- Content-based image retrieval: search similar images to the query image.
- Similarity is based on image features; currently PCA, but ideally more complex.

Search by Rainfall Similarity

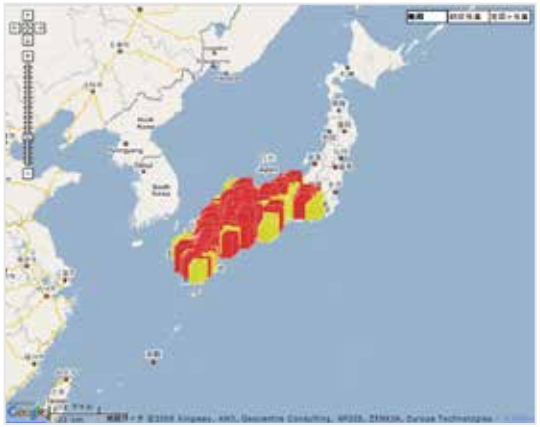


1998/09/20-1998/09/25

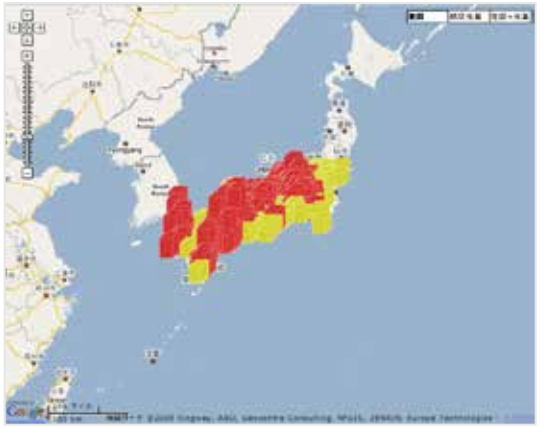
- Use a spatial pattern as a query, and retrieve other patterns.
- Similarity is based on the rainfall amount and distance.



2004/10/18-2004/10/21



1998/10/14-1998/10/18



1983/09/25-1983/09/29

Social Weather Monitoring

Search by keyword

Social data

Toponym-based geotagging by GeoNLP

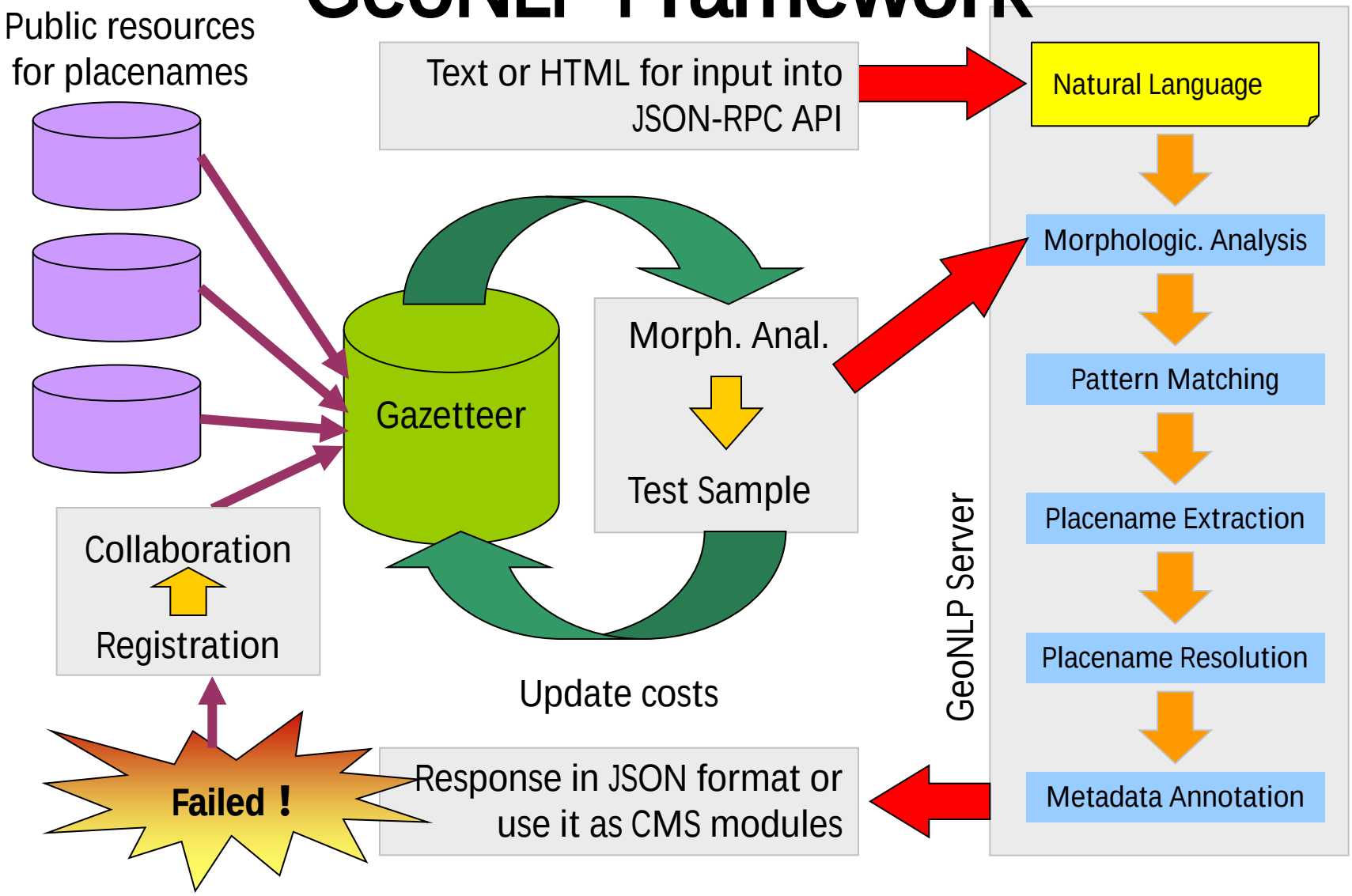
(time, latitude, longitude, situation)

Scientific data

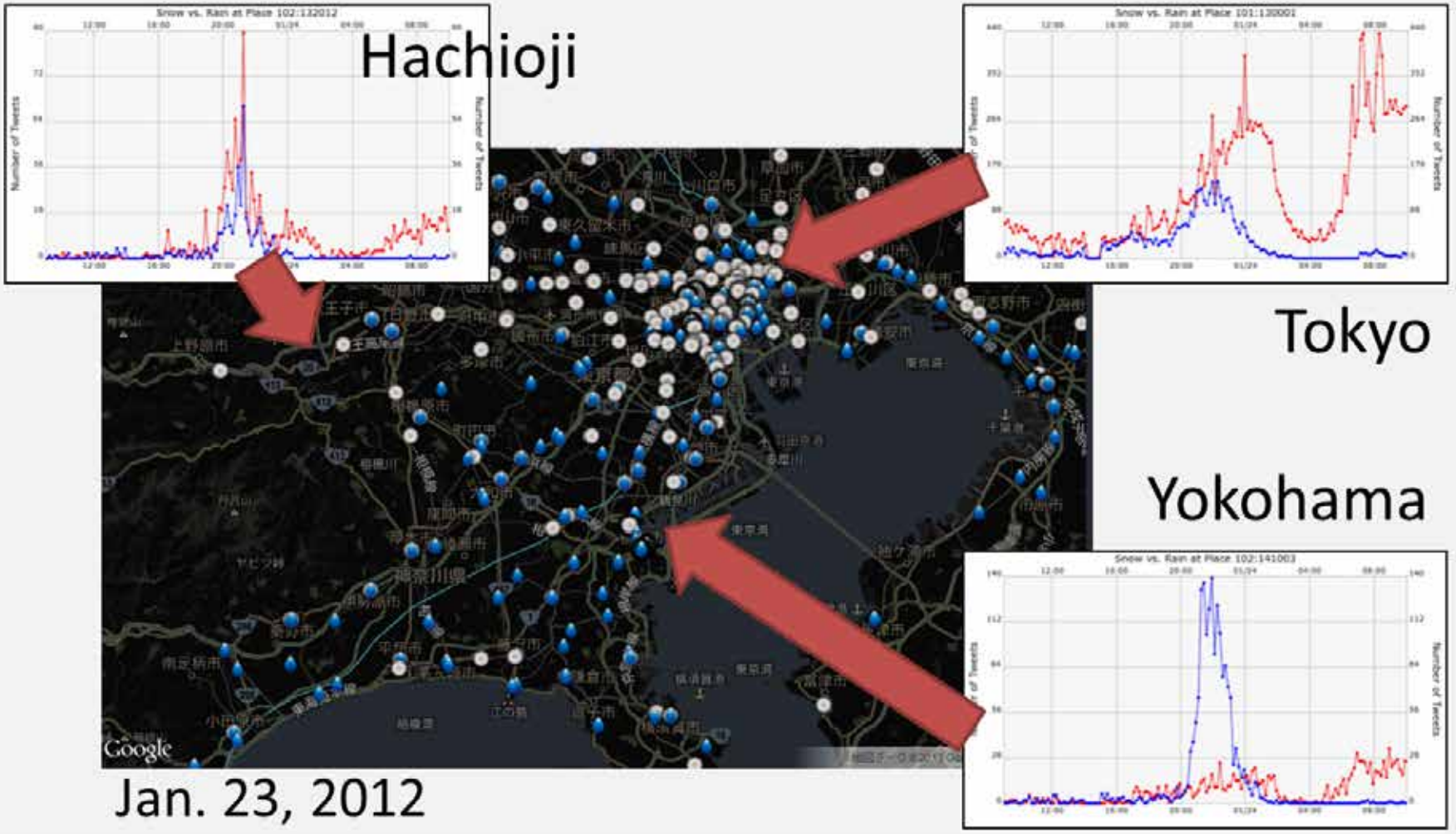
Comparison of social and scientific streams in terms of precipitation

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

GeoNLP Framework

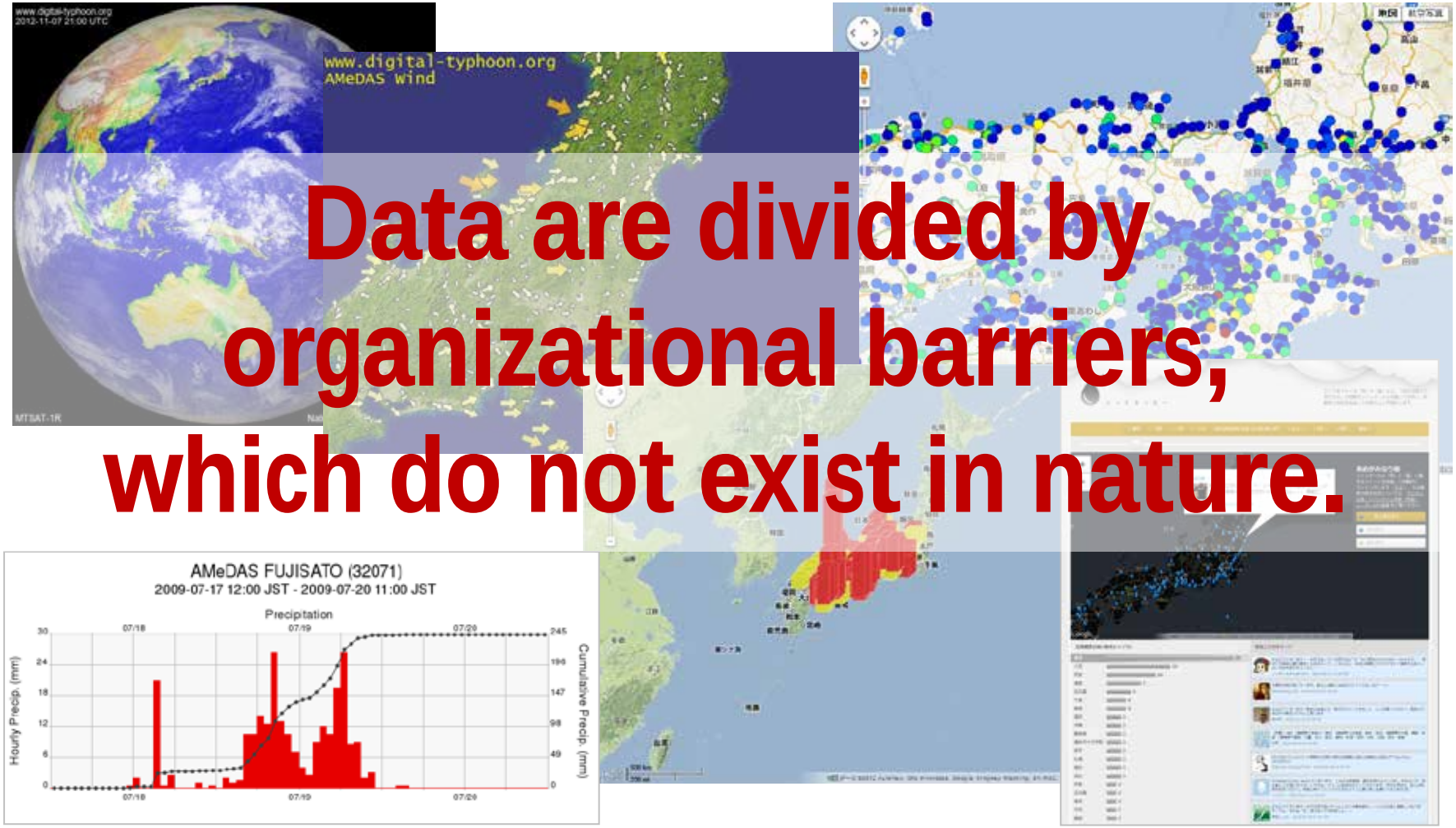


Snow (white/red) and Rain (blue)



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

Barriers between Organizations



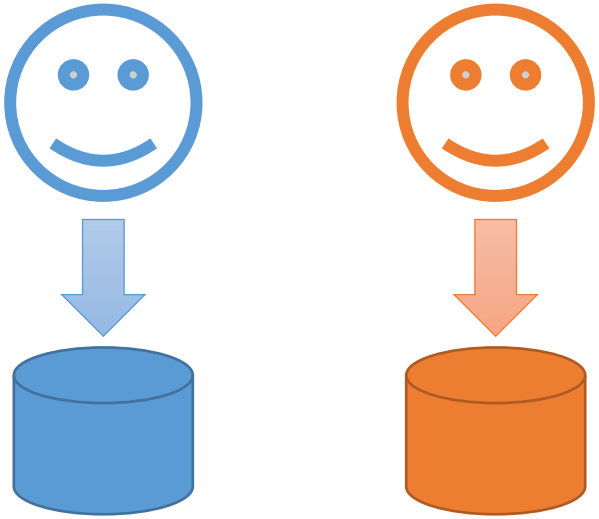
Data are divided by organizational barriers, which do not exist in nature.

Barrier-Free Infrastructure

Exclusive Design

Scientists

Citizens



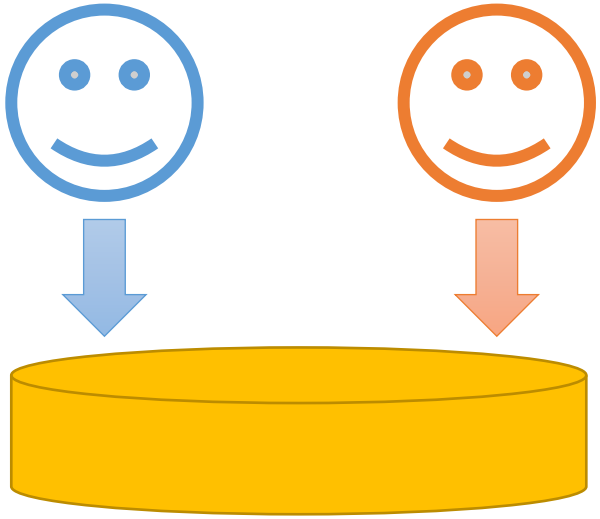
Professional Infrastructure

Outreach Infrastructure

Inclusive Design

Scientists

Citizens



Universal Infrastructure with better experience for everyone

Case Study: “Digital Silk Road” (Humanities)

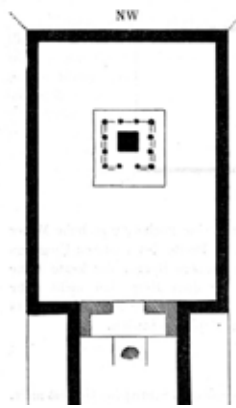
Digital Silk Road

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.

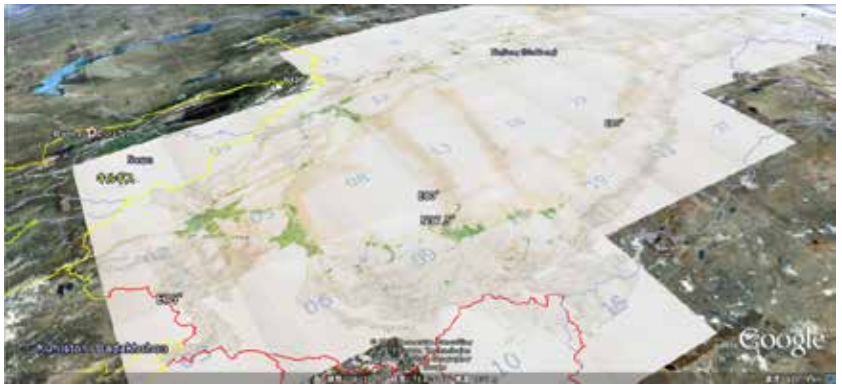
Variety and Heterogeneity of Data

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 läßt. 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 Säulen in kleine Abteile geteilt ist, von denen der mittlere der Frontseite den Eingang bildet. Auf der Rückseite ist dies aus zwei Eck- und zwei Mittelsäulchen bestehende System sehr zerstört. Vor den sechs Interkolumnien der Seiten und den zwei Interkolumnien neben dem Eingang sind je noch Sockel für Statuen erhalten: auch mancherlei dekoratives Bei-

Map



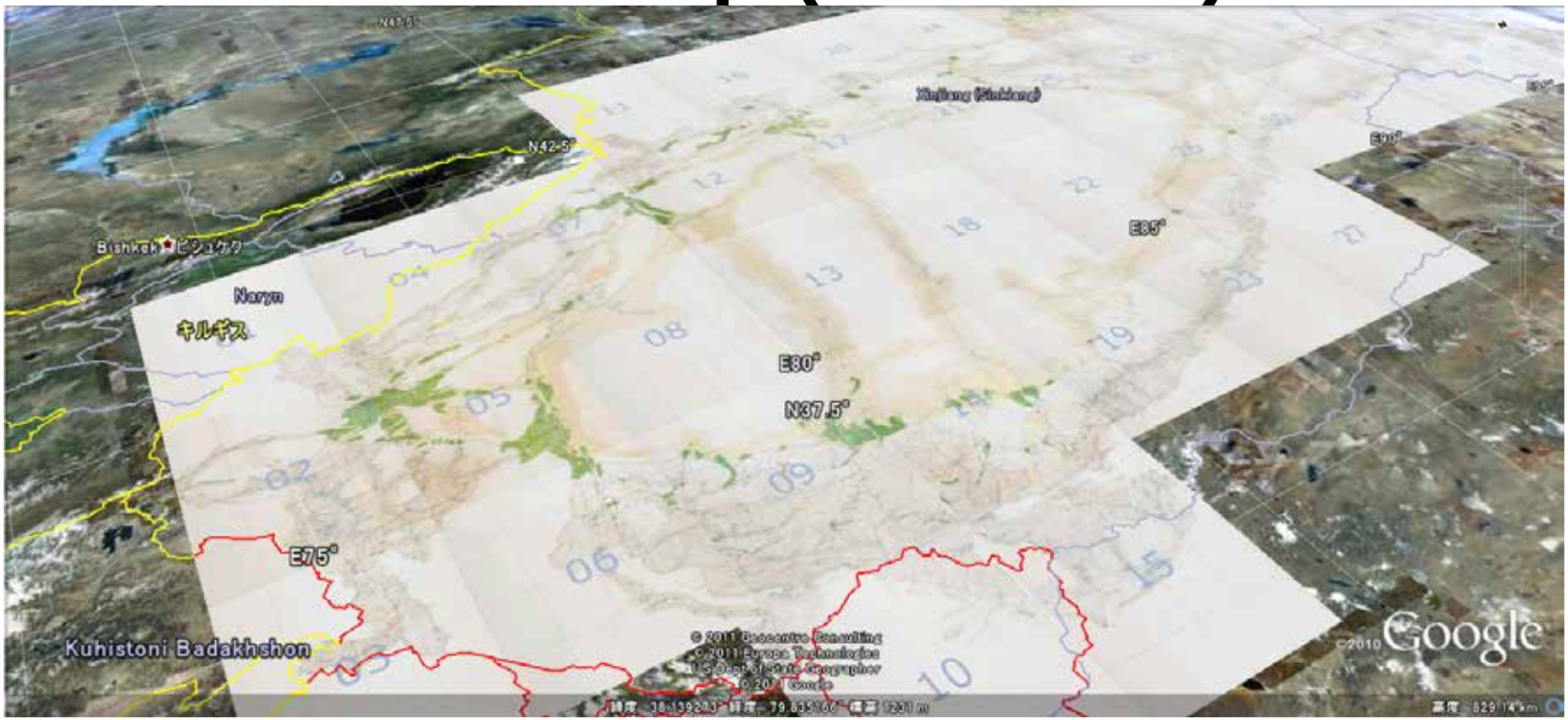
Photograph

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

Gazetteer

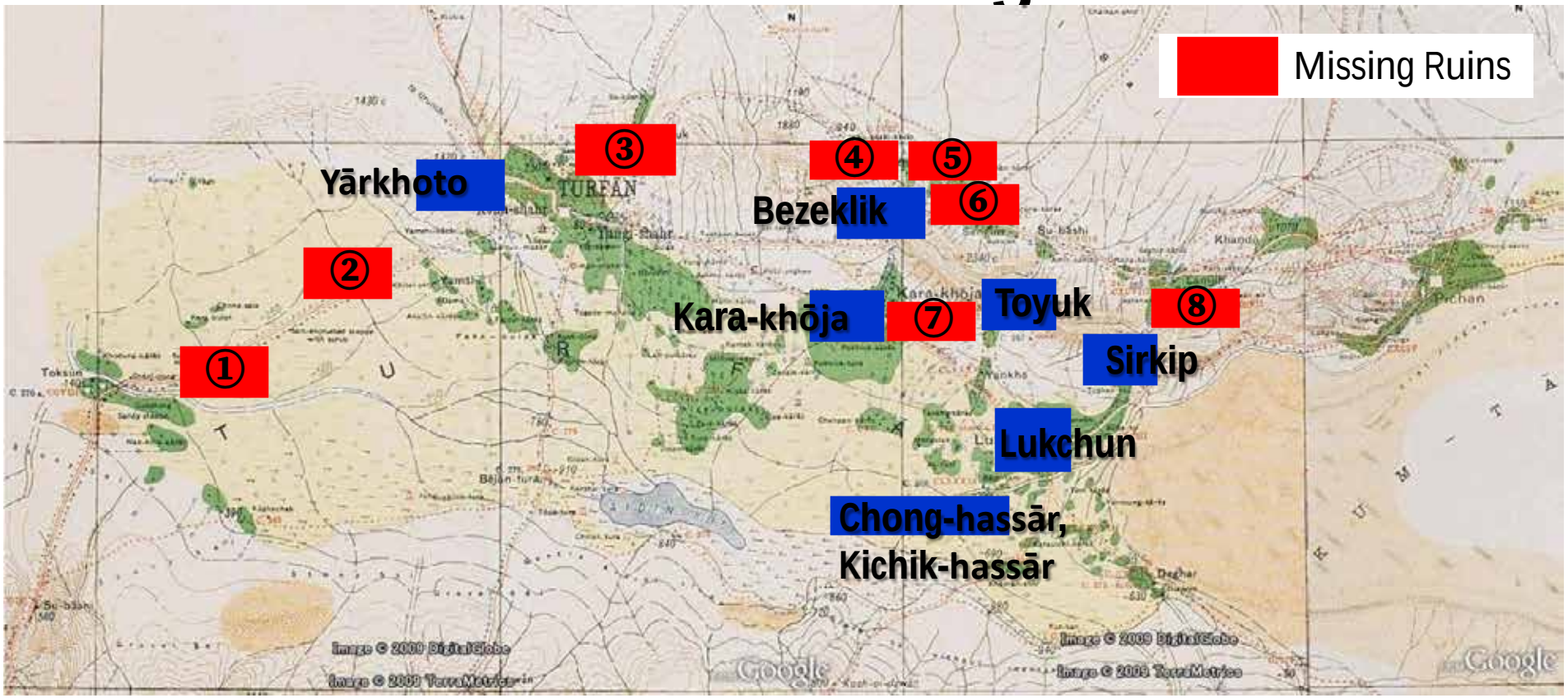
Abāb-langar, habit., 14. B. 3.
Ābād (of Ak-su), market-town, 12. A. 3.
Abād (of Kara-yulghun), vill., 12. B. 1.
Abād (of Karghalik), vill., 5. C. 4.
Ābād (of Kāshgar), vill., 5. A. 2.
Ābād (of Turfān), vill., 28. C. 3.
Ābād (of Yārkand), vill., 5. C. 2.
Ābād-jilga, valley, 12. B. 2.
Abdal, vill,. 30.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

Stein Map (Silk Road)



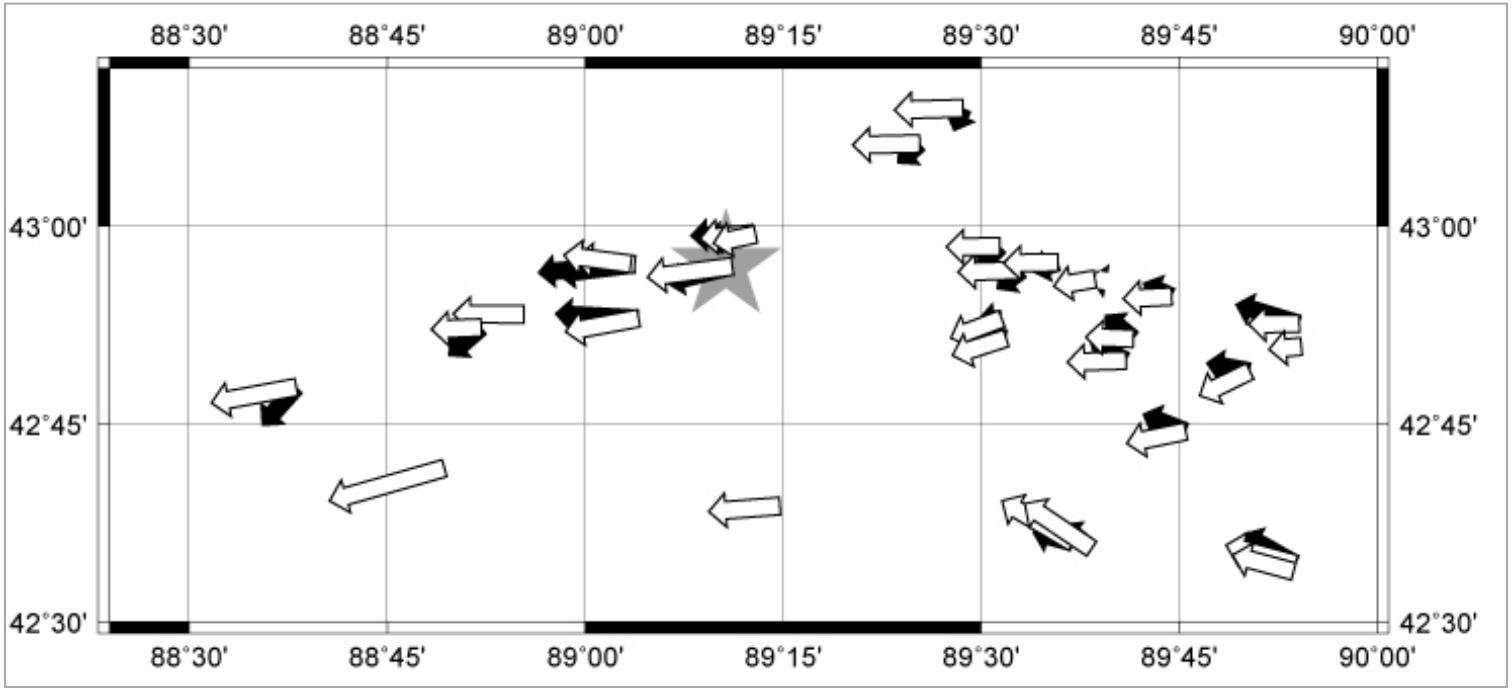
- Stein's map "Innermost Asia" was registered and displayed on Google Earth satellite images.

Problem of “Missing” Ruins



① Oi-tam, ruined fort
② Bögan-tura
③ Buluyuk (Shipang, Sassik-bulak, Kazma)
④ Murtuk-ruins
⑤ Yoghan-tura
⑥ Chikkan-köl
⑦ Bedaulat’s town, Bēsh-kāwuk, Kosh-gumbaz
⑧ Yutōgh

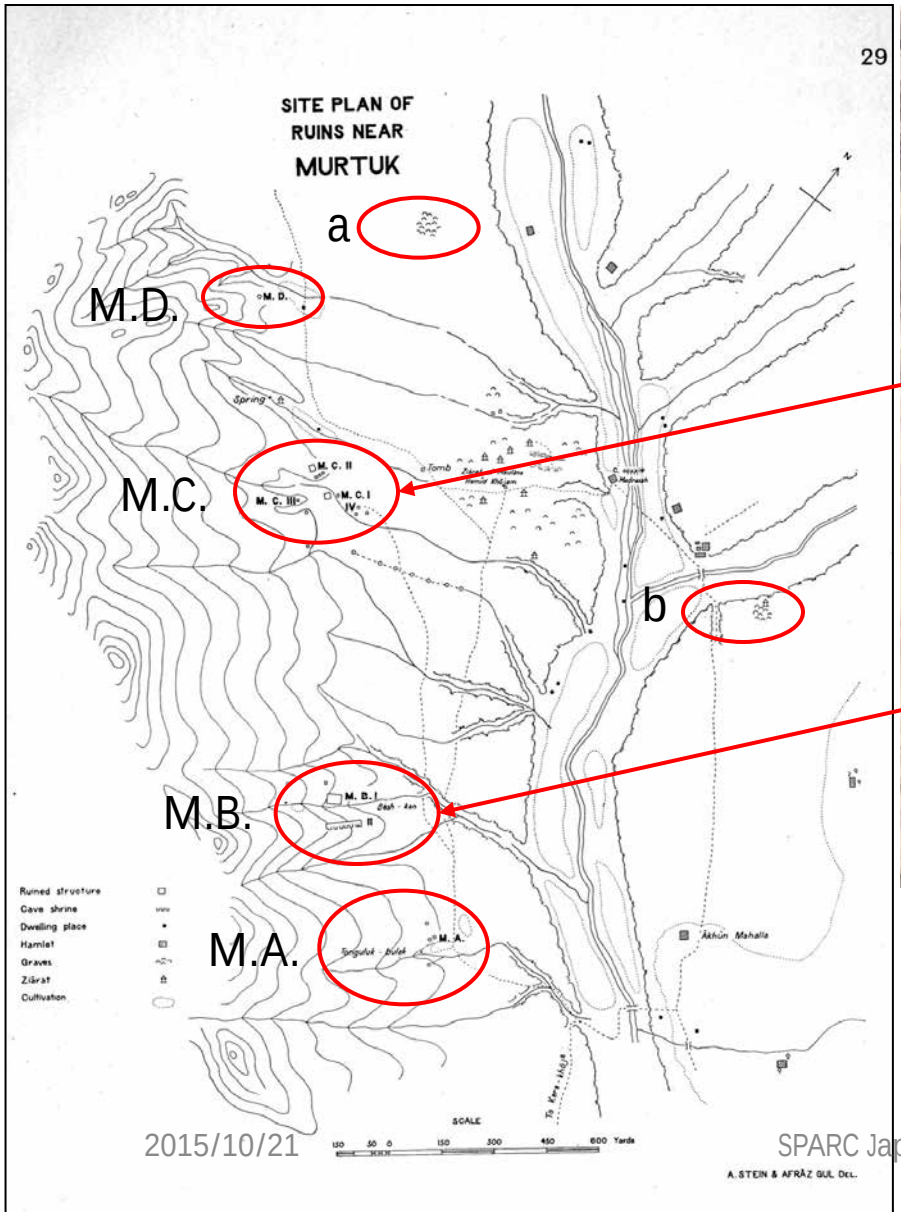
Error Distribution in Turfan

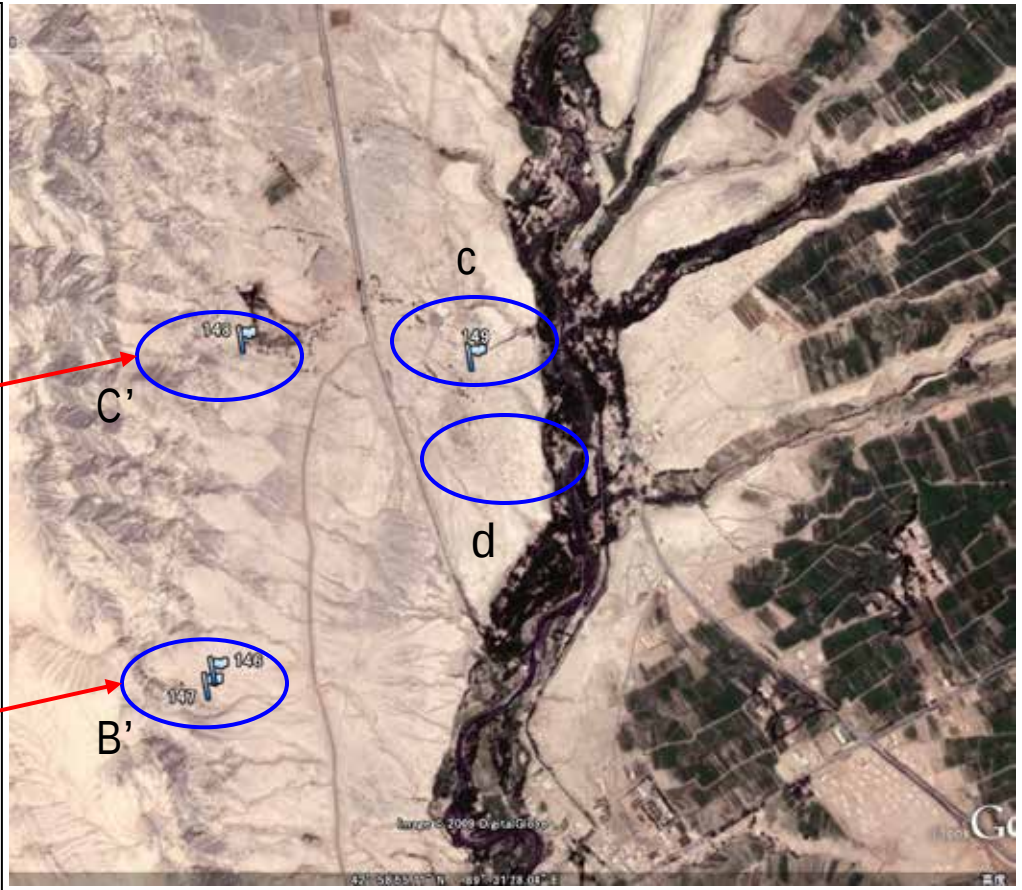


Error Distribution in Turfan Basin / White: Innermost Asia / Black: Serindia

- Some ruins were reported by 20th 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

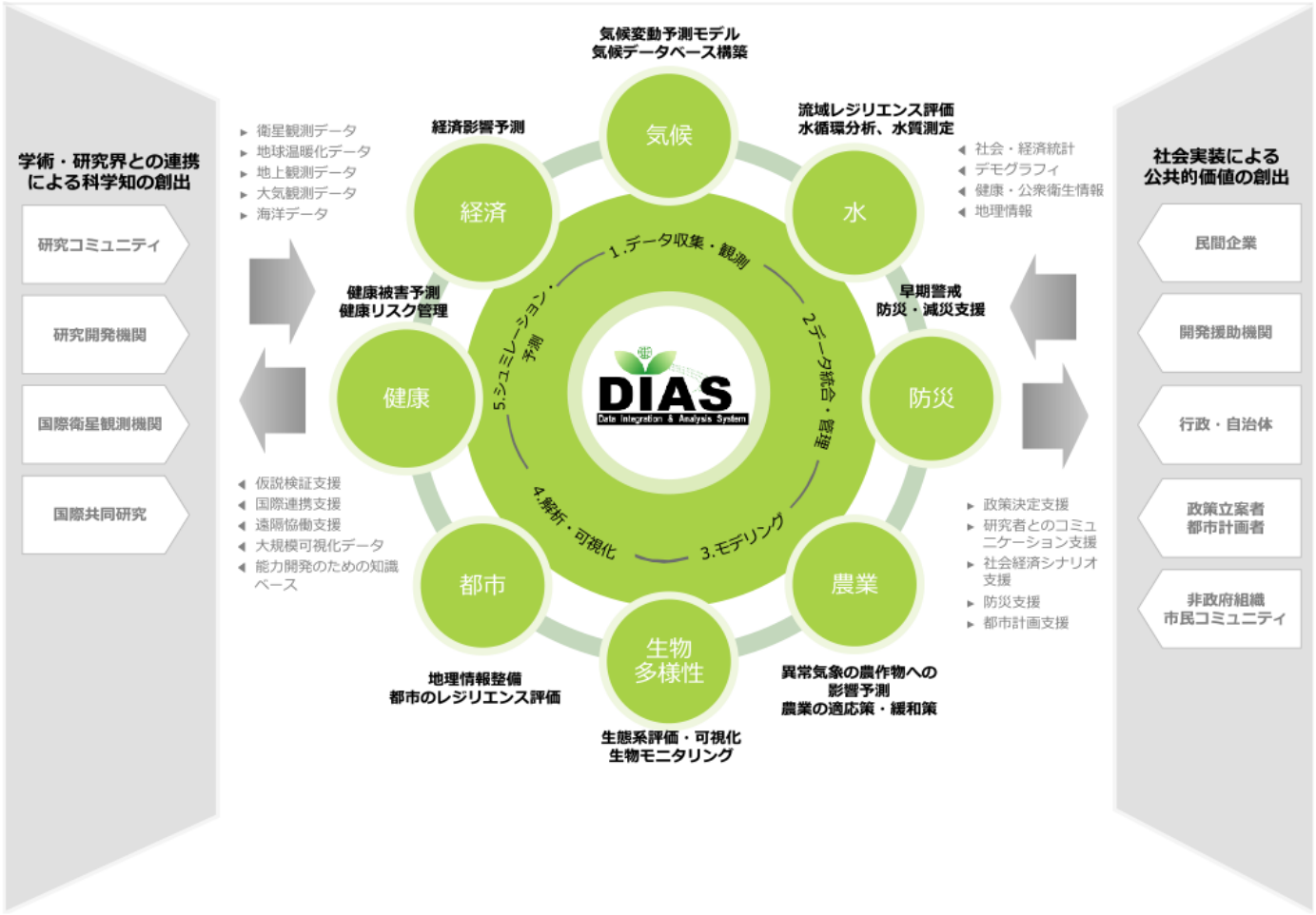
- Take data prepared by another community.
- Apply existing methods that you are already familiar with.
- Obtain results which are already known, or trivially correct.

Innovative Results

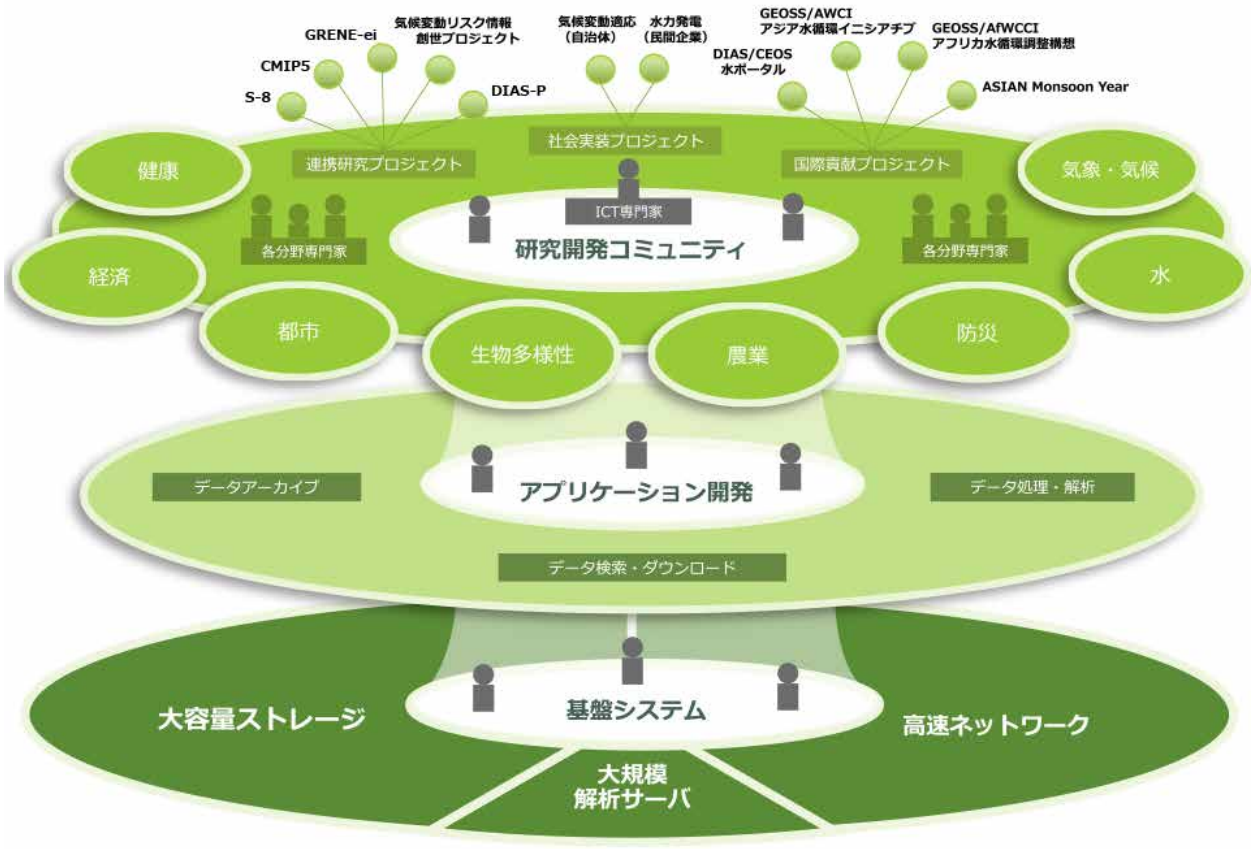
- Deepen knowledge about the problem.
- Propose a new concept for innovating the viewpoint of research.
- Obtain results which cannot be realized without collaboration.

Forces to Break through Barriers

DIAS : Data Integration and Analysis System



DIAS = Data + Community

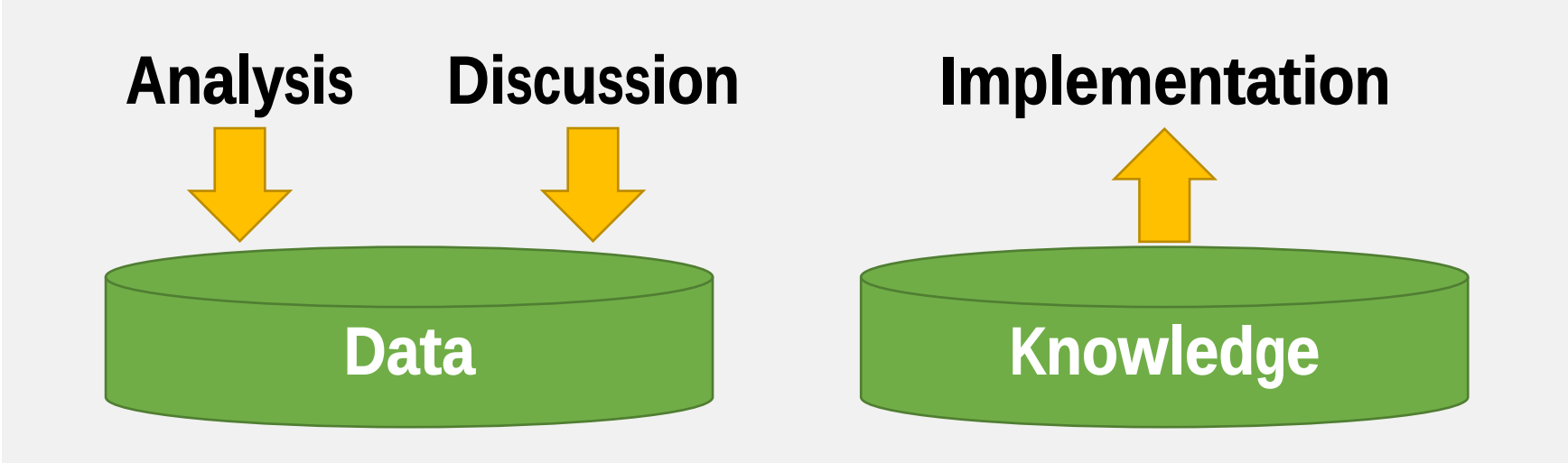


Community Infrastructure

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

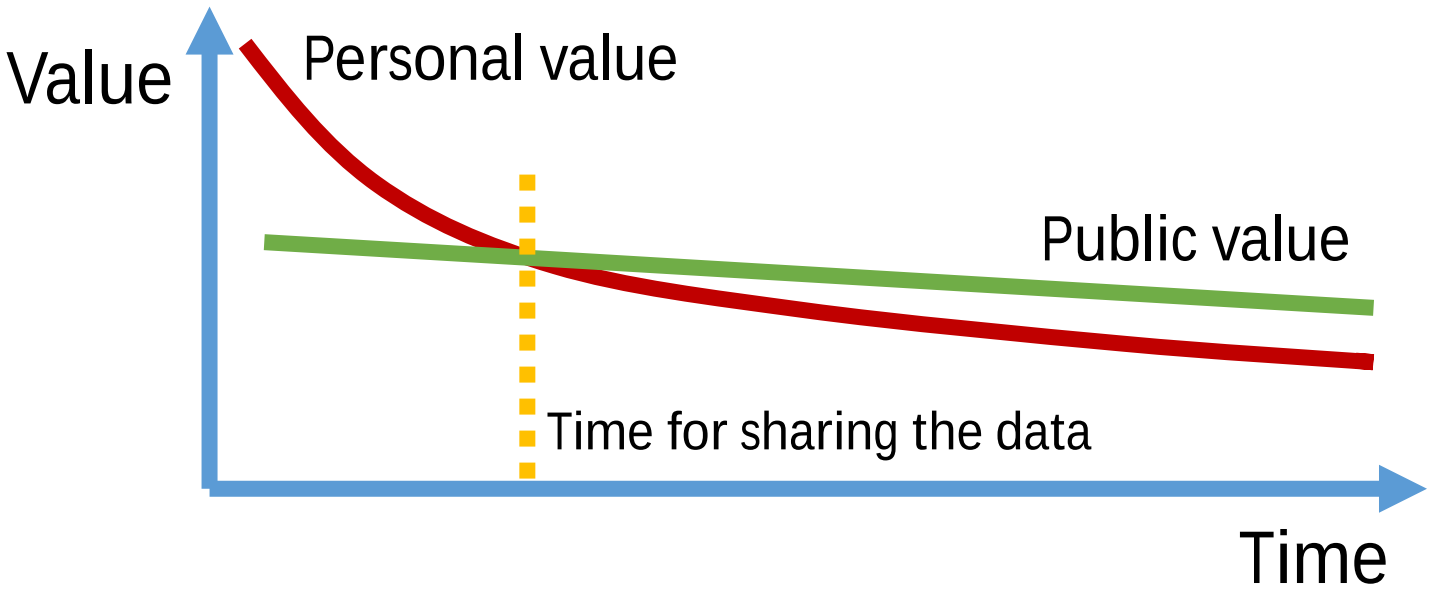
Business

Mission:
public benefit
= open by
default

Mission:
innovation =
eventually
open

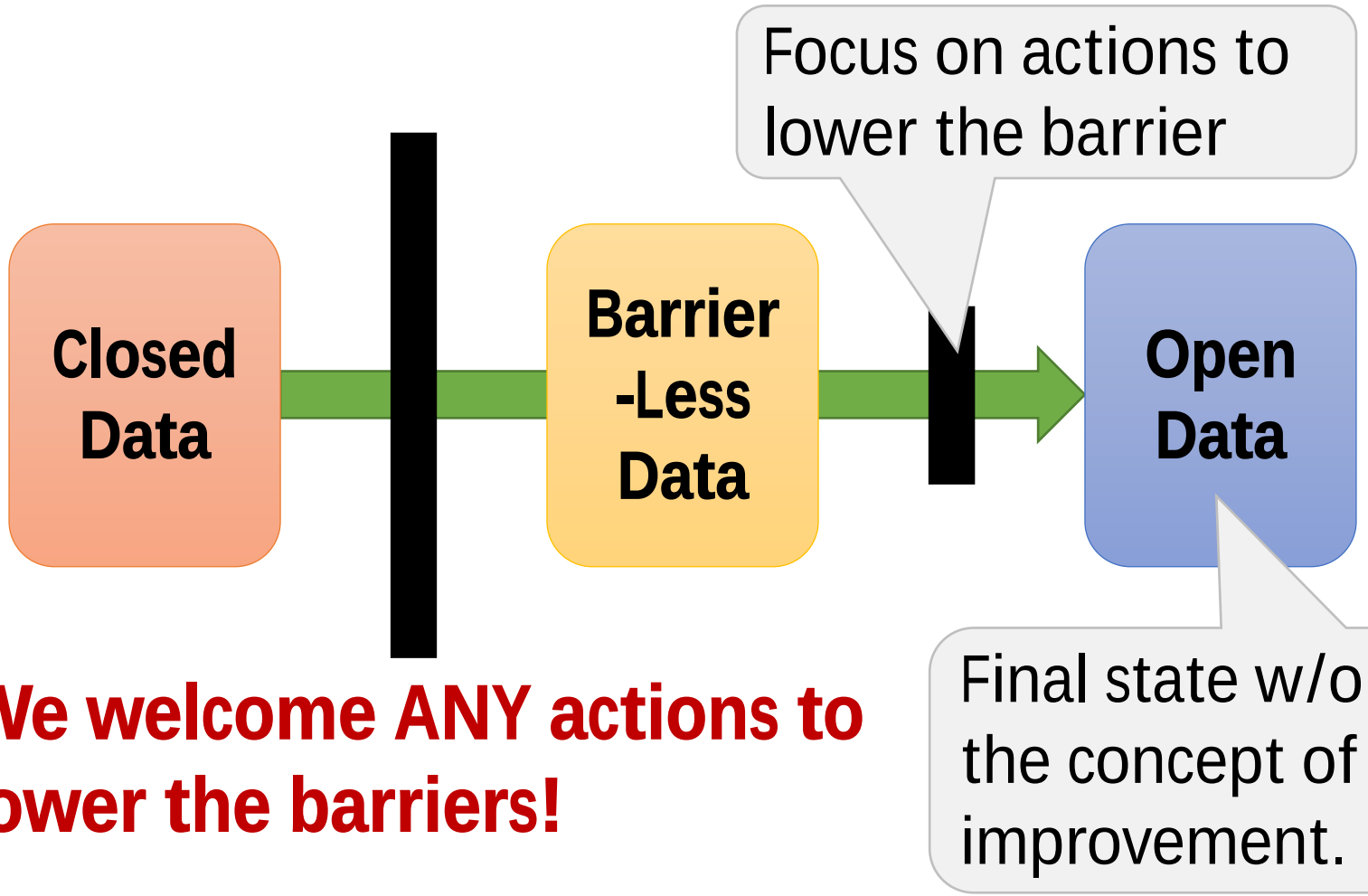
Mission:
(internal) profit
= open by
strategy

Eventual Openness of Data



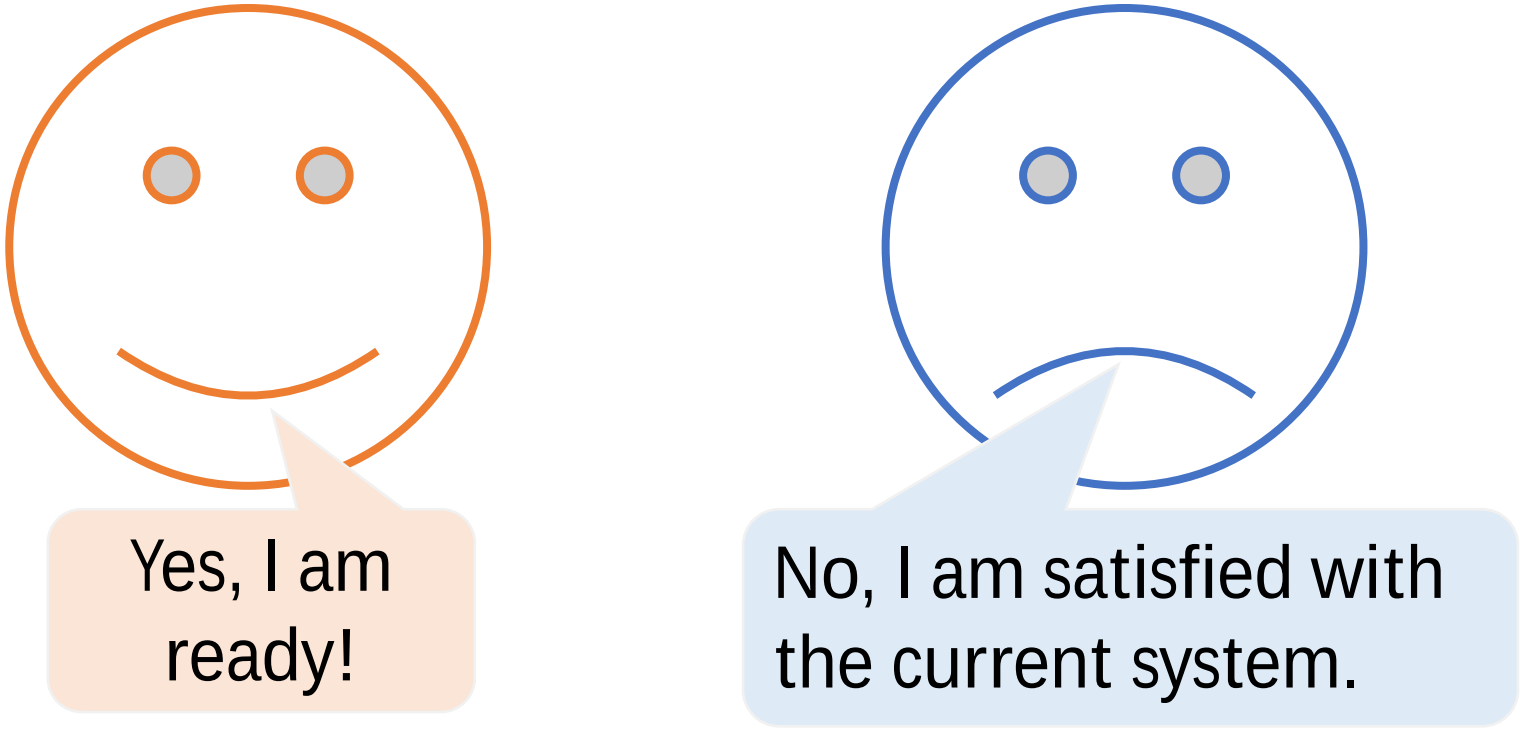
- Raise public value by increasing citation, evaluating open innovation, reducing data management cost.
- At the end of your life, you want to share the data!

Barrier-Less Data



We welcome ANY actions to lower the barriers!

Are You Ready for Actions?



- 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/