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

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

https://rd-alliance.org/
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/

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/

• 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.
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

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

• 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

Text or HTML for input into JSON-RPC API

Gazetteer

Morph. Anal.

Test Sample

Morph. Anal.

Natural Language

Morphologic. Analysis

Pattern Matching

Placename Extraction

Placename Resolution

Metadata Annotation

Update costs

Response in JSON format or use it as CMS modules

Failed!
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
  - Professional Infrastructure

- Citizens
  - Outreach Infrastructure

Inclusive Design

- Scientists
  - Universal Infrastructure with better experience for everyone

- Citizens
Case Study: “Digital Silk Road” (Humanities)
Digital Silk Road

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

http://dsr.nii.ac.jp/
Variety and Heterogeneity of Data

Text

Map

Photograph

Gazetteer
Stein’s map “Innermost Asia” was registered and displayed on Google Earth satellite images.
Problem of “Missing” Ruins

- Yārk̠ło-to
- Bezeklik
- Kara-khōja
- Toyuk
- Sirkip
- Lukchun
- Chong-hassār, Kichik-hassār

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

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

2015/10/21 SPARC Japan Seminar 2015
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

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
Mission: public benefit = open by default

Research
Mission: innovation = eventually open

Business
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

Closed Data → Barrier-Less Data → Open Data

Focus on actions to lower the barrier

Final state w/o the concept of improvement.

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/