

# Dilemma in Open Science オープンサイエンスのジレンマ

Incentive in Scientific Communities and Motivation toward  
Eventual Openness

研究者共同体のインセンティブと結果オープン  
性へのモチベーション

National Institute of Informatics / SOKENDAI

Asanobu KITAMOTO

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

# Try to Understand Open Science...

**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

「巨人の肩」と  
データサイテーション

国立情報学研究所／総合研究大学院大学  
北本 朝展（きたもとあさのぶ）  
<http://agora.ex.nii.ac.jp/~kitamoto/>

2015/10/02 地球科コアプロジェクトS 1

「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/>

2015/10/21 SPARC Japan Seminar 2015 1

- **Convergence to Open Science – Formation of a Community to Foster Shared Perception from Different Dreams**
- 1<sup>st</sup> Open Science Data Workshop
- Sep. 17, 2015.

- **'Shoulders of Giants' and Data Citation**
- The Open Science Core Workshop
- Oct. 2, 2015

- **Design of Research Infrastructure and Utilization of Research Data for Breaking through Research Barriers**
- 2<sup>nd</sup> SPARC Japan Seminar 2015
- Oct. 21, 2015

# 5W1H of Open Science

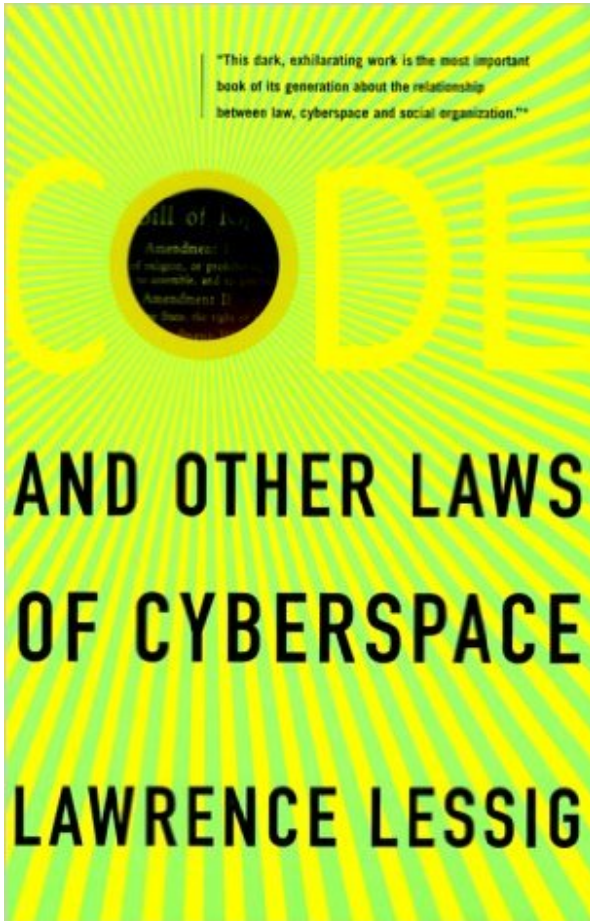
- **WHO** = Scientists and citizens
- **WHAT** = Make it open!
- **WHERE** = World, Japan, institutes
- **WHEN** = Near future?
- **WHY** = ?
- **HOW** = ?



Narratives for persuasion are missing

# WHY of Open Science

# Four Modes of Regulation



- **Lawrence Lessig (Founder of Creative Commons)** proposed four constraints to regulate in *Code: And Other Laws of Cyber Space* (first edition 1999).
- Four constraints are **the law, social norms, the market, and architecture**.
- Regulation can be designed as the sum of four constraints.

# WHY of Open Science

1. **Norm** : That is a good thing to do (otherwise purged from the community).
2. **Law** : Someone asked (forced) me to do it (otherwise punished by the authority).
3. **Market** : That is beneficial for me (by counting the profit)
4. **Architecture** : Well, it's difficult to behave in other ways (due to the environment imposed on me).

# Openness by Norms



<https://www.icsu-wds.org/>

- “Science” has been open throughout the history.
- It is the obligation of good scientists to share results.
- Scientists may not be persuaded only by a good purpose.

# Openness by the Law



The screenshot shows the NSF website's 'Dissemination and Sharing of Research Results' page. The header includes the NSF logo and the tagline 'WHERE DISCOVERIES BEGIN'. A navigation bar lists categories like FUNDING, AWARDS, DISCOVERIES, NEWS, PUBLICATIONS, STATISTICS, ABOUT NSF, and FASTLANE. The main content area is titled 'Dissemination and Sharing of Research Results' and contains sections for 'NSF Data Sharing Policy', 'NSF Data Management Plan Requirements', and 'Requirements by Directorate, Office, Division, Program, or other NSF Unit'. The 'Requirements by Directorate...' section lists various directorates with links to their specific guidance documents.

**Office of Budget, Finance and Award Management (BFA)**

**Dissemination and Sharing of Research Results**

NSF Data Sharing Policy

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See [Award & Administration Guide \(AAG\) Chapter VI.D.4](#).

NSF Data Management Plan Requirements

Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See [Grant Proposal Guide \(GPG\) Chapter II.C.2.i](#) for full policy implementation.

Requirements by Directorate, Office, Division, Program, or other NSF Unit

Links to data management requirements and plans relevant to specific Directorates, Offices, Divisions, Programs, or other NSF units, are provided below. If guidance specific to the program is not provided, then the requirements established in [Grant Proposal Guide, Chapter II.C.2.i](#) apply.

Please note that if a specific program solicitation provides guidance on preparation of data management plans, such guidance must be followed.

- Biological Sciences Directorate (BIO)
  - [Directorate-wide Guidance](#)
- Computer & Information Sciences & Engineering (CISE)
  - [Directorate-wide Guidance](#)
- Education & Human Resources Directorate (EHR)
  - [Directorate-wide Guidance](#)
- Engineering Directorate (ENG)
  - [Directorate-wide Guidance](#)
- Geosciences Directorate (GEO)
  - [Directorate-wide Guidance](#)

NSF Data Management Plan

- Data Management Plan (DMP) for grants.
- Data preservation for reproducible research.
- Not easy to reach consensus within communities.



# Openness by the Market

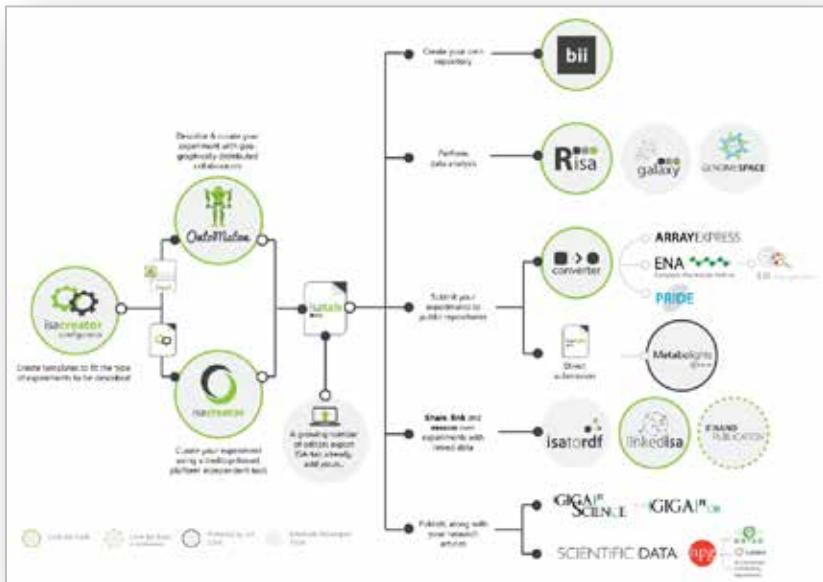
The screenshot displays the Scientific Data journal homepage. At the top, the journal title 'SCIENTIFIC DATA' is prominent, along with navigation links for Home, Archive, About, For Authors, For Referees, Data Policies, and Collections. A search bar is located in the top right corner. The main content area is divided into several sections:

- Featured Data Descriptor:** The top featured article is 'The Icelandic 16-electrode electrohysterogram database' by Alexander et al. (2015 April 2015). The abstract discusses the prevalence of preterm births and the challenges of predicting patients' labor, highlighting the use of 122 4-by-4 electrode EHG recordings from 45 pregnant women.
- About Scientific Data:** A section explaining the journal's mission as an open-access, peer-reviewed publication for descriptions of scientifically valuable datasets. It includes a 'Submit manuscript' button.
- Latest Research:** A grid of recent articles, including 'A global compendium of human Crimean-Congo haemorrhagic fever virus occurrence', 'Sequence data for Clostridium autohologenum using three generations of sequencing technologies', and 'A comprehensive map of genome-wide gene regulation in Mycobacterium tuberculosis'.
- Announcements:** A section for updates, including 'Author's Corner: Mapping Crimean-Congo haemorrhagic fever' and 'Repository highlight: Tgshare and the crucial service of generalist repositories'.

- Accumulating credits through open data and data citation.
- Research grants can focus more on projects' openness.
- The outcome of a market design is not easy to predict.

Scientific Data (Nature publishing group)

# Openness by Architecture



<http://www.isa-tools.org/software-suite/>

- Research workflow for data sharing and management.
- Software code can reduce the barrier toward openness.
- **Commercialization may be advanced by commercial services.**

# Narratives for WHY

- Open science needs narratives to explain why we should move toward that direction.
- **The market and architecture** : interesting areas for exploration, but with a risk of more commercialization by powerful players.
- **Law and norms** : necessary areas for exploration, but with a risk of emphasizing transparency over benefit.

# HOW of Open Science

# HOW of Open Science

- How can we move from traditional science to open science? Dreams without plans and actions will never be realized.
- Solutions should be adapted to each constraint. Architecture needs better data infrastructure; the law needs better rules.
- Each solution is affected by its own dilemma, or trade-off, between multiple stakeholders. A solution should clarify for whom.

# Games in Scientific Communities

- **Ideal scientists** : motivated by pure interests in science and discovery = not influenced by other scientists.
- **Real scientists** : motivated by a desire to get more funding, to publish more (good) papers, to receive more citations, with the final goal of winning a position or promoted to a higher position = always influenced by others.
- **A reputation game in a scientific community.**

# Game Theory

- Game theory is a study of interaction in terms of decision making between actors who make decisions.
- A decision of one actor is influenced by a decision of other actors.
- Simplify the real world by a model to understand the mechanism behind competition and cooperation.

# Prisoner's Dilemma

| Reward      |               | Scientist B |               |
|-------------|---------------|-------------|---------------|
|             |               | Cooperate   | Not cooperate |
| Scientist A | Cooperate     | (4, 4)      | (-6, 10)      |
|             | Not cooperate | (10, -6)    | (0, 0)        |

If both scientists cooperate, both will receive benefits. But one can receive more benefit by not cooperating with the other.



# Open Data's Dilemma

| Reward      |          | Scientist B |          |
|-------------|----------|-------------|----------|
|             |          | Open        | Not open |
| Scientist A | Open     | (4, 4)      | (-6, 10) |
|             | Not open | (10, -6)    | (0, 0)   |

You can take the largest advantage by using open data by other scientists and still hide the data of yourself.

# Social Dilemma

- Each person chooses to cooperate or not cooperate. For each person, non-cooperation results in larger benefit than cooperation.
- If everyone chooses not to cooperate because it is more beneficial, the result is worse than that obtained by cooperation.
- The result is desirable for everyone, when everyone chooses to cooperate even if it is not beneficial.

# Open Access's Dilemma?

| Priority  |               | Publisher |           |
|-----------|---------------|-----------|-----------|
|           |               | Cheap     | Expensive |
| Scientist | Subscribe     | (1, 3)    | (2, 1)    |
|           | Not subscribe | (4, 4)    | (3, 2)    |

For a scientist, a subscription is always a better choice. For a publisher, expensive is always a better choice. This is bad for a scientist, but a publisher is satisfied; not dilemma.

# Narratives for HOW

- We need an alternative narrative than “publish or perish” of peer-reviewed articles.
- **Open data's dilemma** about hiding data is due to the fear of losing competition.
- **Open access's dilemma?** about expensive journals is due to the fear of excluding from competition.
- Commercial actors want to fuel this narrative for their sustainable business and enclosure.

# Possible Solutions

# Incentive and Motivation

## Incentive (extrinsic)

- Change a reward for a choice.
- A solution closer to the market.
- A rational person is expected to respond to incentive.

## Motivation (intrinsic)

- Change the mind for a choice.
- A solution closer to social norms.
- Extrinsic incentive gives bad impact on intrinsic motivation.

# Control the Incentive

| Reward      |          | Scientist B |            |
|-------------|----------|-------------|------------|
|             |          | Open        | Not open   |
| Scientist A | Open     | $(X, X)$    | $(-6, 10)$ |
|             | Not open | $(10, -6)$  | $(0, 0)$   |

Increase  $X$  so that open scientists can take more benefit from cooperating with other open scientists.

# Giant Shoulder Index (GSI)

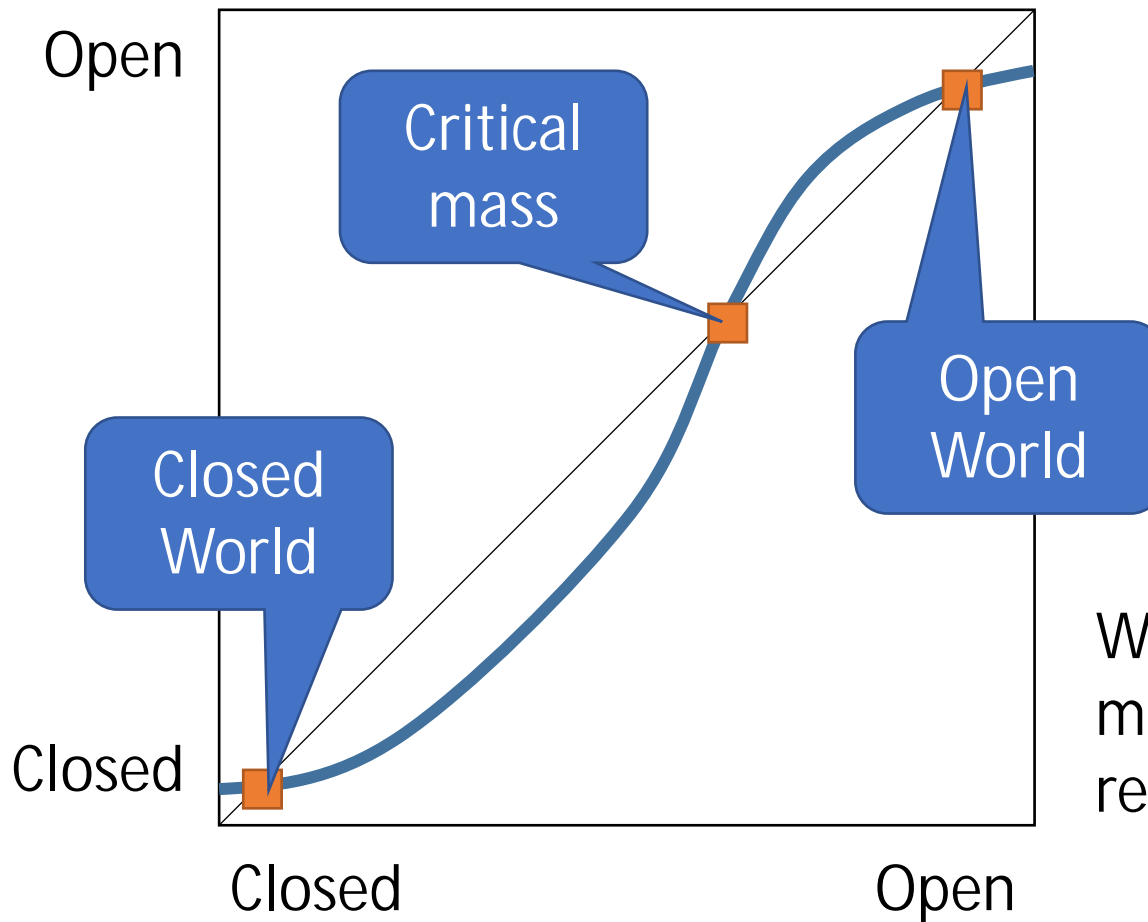
- **Giant Shoulder Index (GSI)** : the holy grail of research index for evaluating the true contribution of a scientist to scientific communities and the world.
- Question: an impact factor or a citation index is a good approximation of GSI?
- **GSI can be designed to reduce the dilemma** by considering other types of contribution.



# Self-Reinforcing Dynamics

- An open scientist may be exploited by a closed scientists.
- An open scientist can obtain larger benefit by cooperating with another open scientist.
- **A critical mass of open scientists** start a self-reinforcing dynamics toward an open world.
- **Chain reaction of actors** under social dilemma arrives at one of two attractors.

# Critical Mass



Where is the critical mass to start self-reinforcing dynamics?

# Mission and Openness



Government

Research

Business

Mission:  
public benefit  
= open by  
default

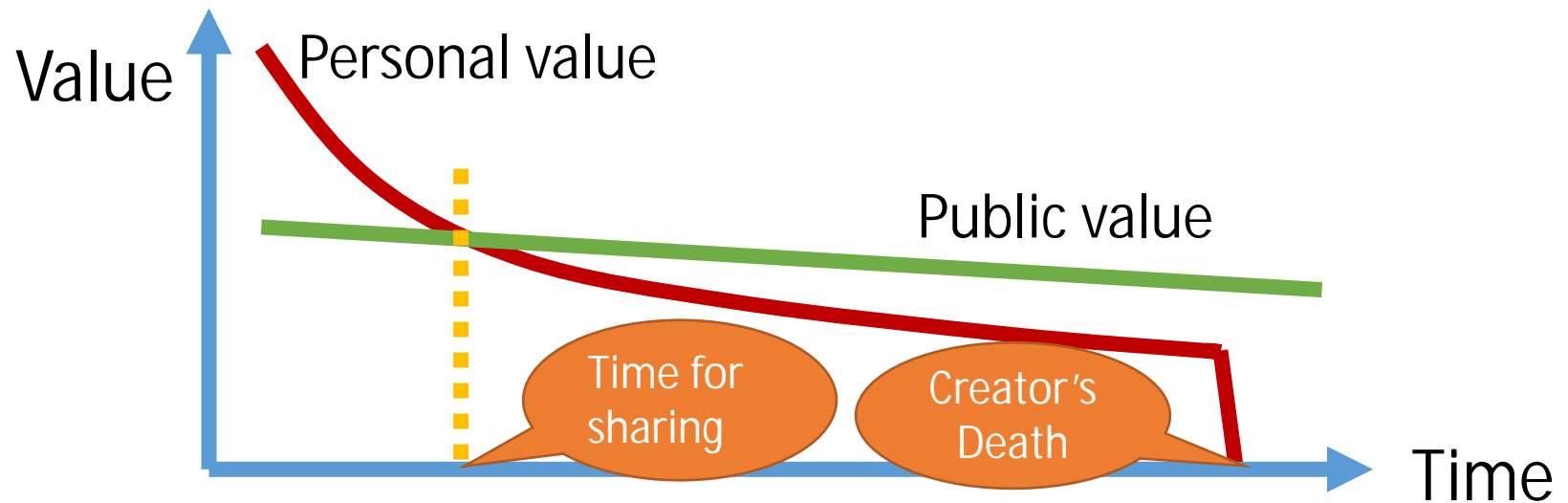
Mission:  
innovation =  
open by  
strategy

Mission:  
organization's  
profit = closed  
by default

# Open by Strategy

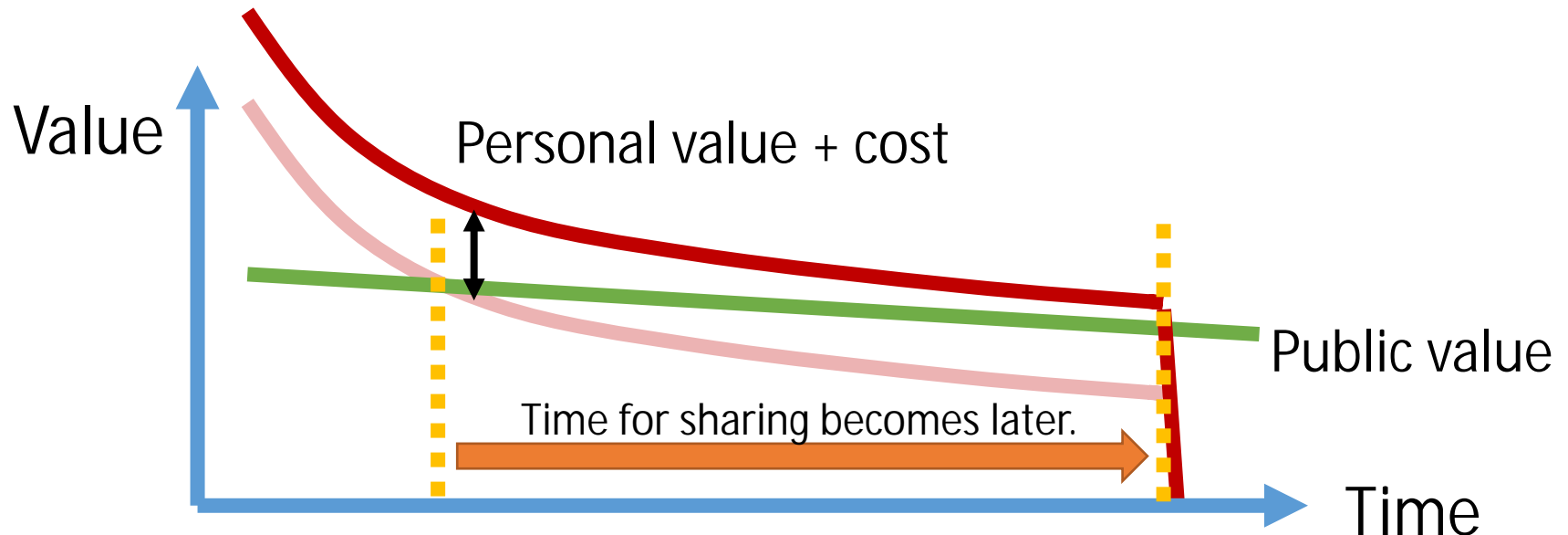
- **Law** : embargo, policy, and enforcement
- **Norms** : trans-disciplinary science and eventual openness
- **Market** : credit mechanism and open innovation
- **Architecture** : research infrastructure and identifier systems

# Eventual Openness



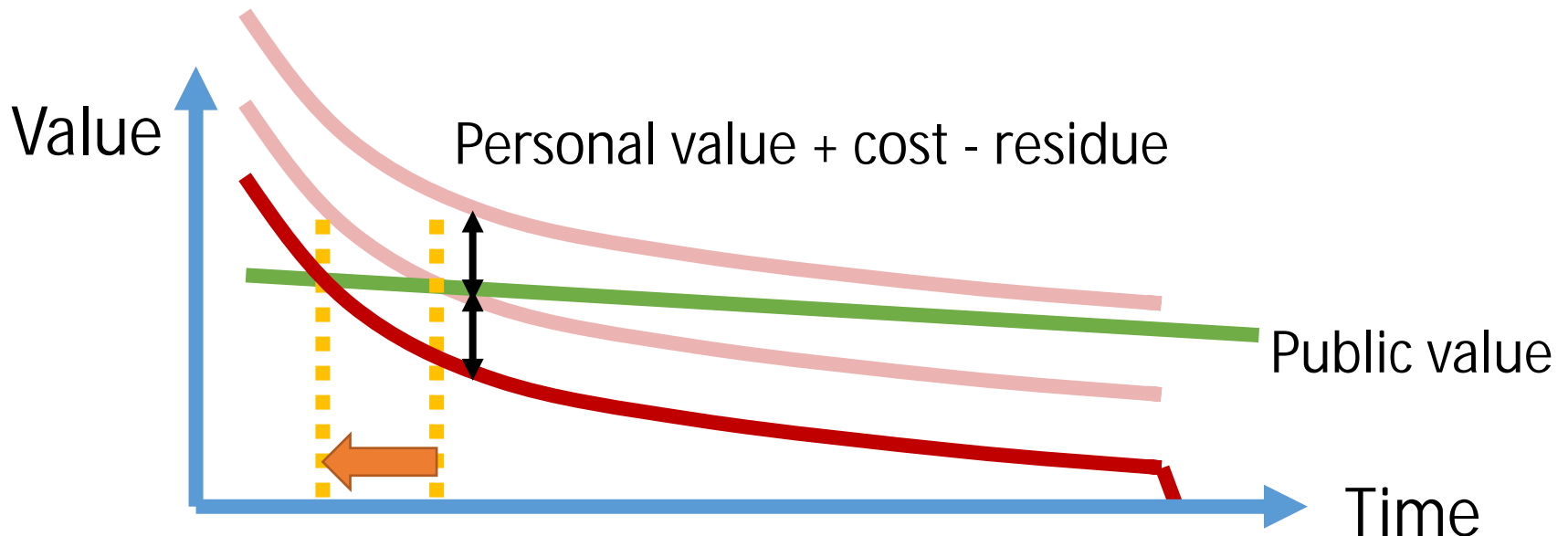
- Personal value decays rapidly than public value.
- At the end of life, personal value = 0.
- **When personal < public, it's the time for sharing.**

# Cost and Openness



- A scientist needs to pay cost for sharing the data (such as formatting, cleaning and documenting).
- To compensate the cost, **a scientist requires more public value that meets value plus cost.**

# Benefit and Openness



- Openness is not only about losing value, but also about gaining value (the residue) after opening.
- The residual value, such as receiving more credits, can reduce a barrier to openness.

# Solutions to Eventual Openness

- Three solutions for eventual openness.
  1. **Market solution** : Raise the public value or the residual value for larger incentive.
  2. **Architecture solution** : Reduce the cost by providing a good data infrastructure.
  3. **Norms solution** : Imagine the motivation of a scientist at the end of life, and change the personal value curve to a steeper one.



# Summary

# Summary

1. “Why open science” can be explained by four modes of constraints; the law, norms, the market, and architecture.
2. “How open science” should be analyzed under social dilemma to design incentive structure in scientific communities.
3. Social dilemma may be partially solved by the idea of Giant Shoulder Index (GSI) and eventual openness.

# Final Remark

- Social dilemma seems to be ubiquitous in open science. Why?
- This is because **multiple stakeholders (actors) have different dreams** and their dreams are sometimes competing or conflicting.
- **The road to hell is paved with good intentions.** We should be aware of social dilemma, and try to remove its structure, or at least make a globally better decision.

# Related Websites

- Open Science Page
  - <http://agora.ex.nii.ac.jp/~kitamoto/research/open-science/>
- Presentation Slides in the Past
  1. Convergence to Open Science - Formation of a Community to Foster Shared Perception from Different Dreams –
    - <http://agora.ex.nii.ac.jp/~kitamoto/research/publications/osd15.html>
  2. 'Shoulders of Giants' and Data Citation
    - <http://agora.ex.nii.ac.jp/~kitamoto/research/publications/corefs15.html>
  3. Design of Research Infrastructure and Utilization of Research Data for Breaking through 'Research Barriers'
    - <http://agora.ex.nii.ac.jp/~kitamoto/research/publications/sparc15.html>