

Talk at CODATA-ICSTI Data Citation Workshop, October 29, 2015

Research Data-DOI Experiment in Japanese DOI Registration Agency (Japan Link Center, JaLC)

Hideaki Takeda

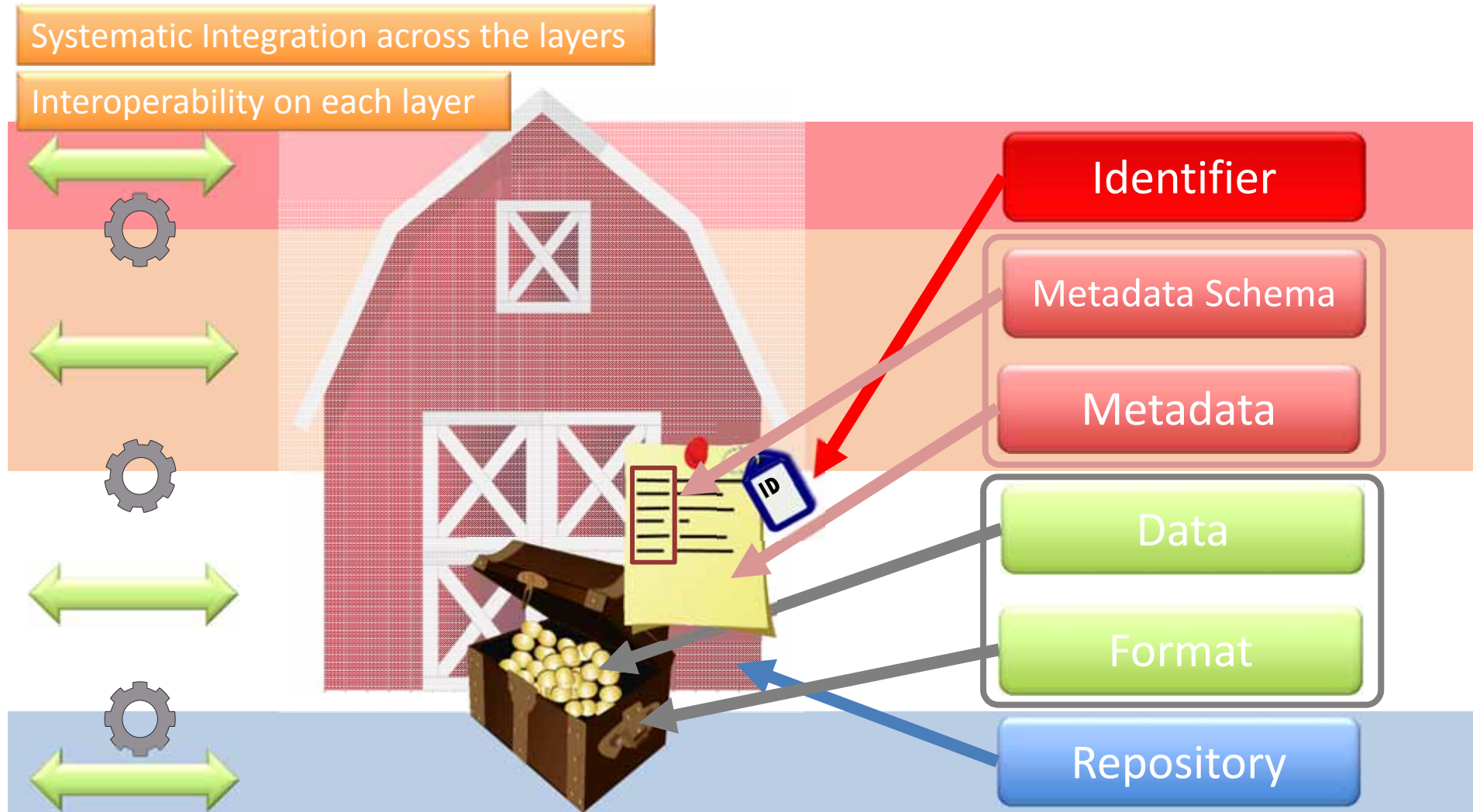
National Institute of Informatics

(Joint Steering Committee Chair, Japan Link Center)

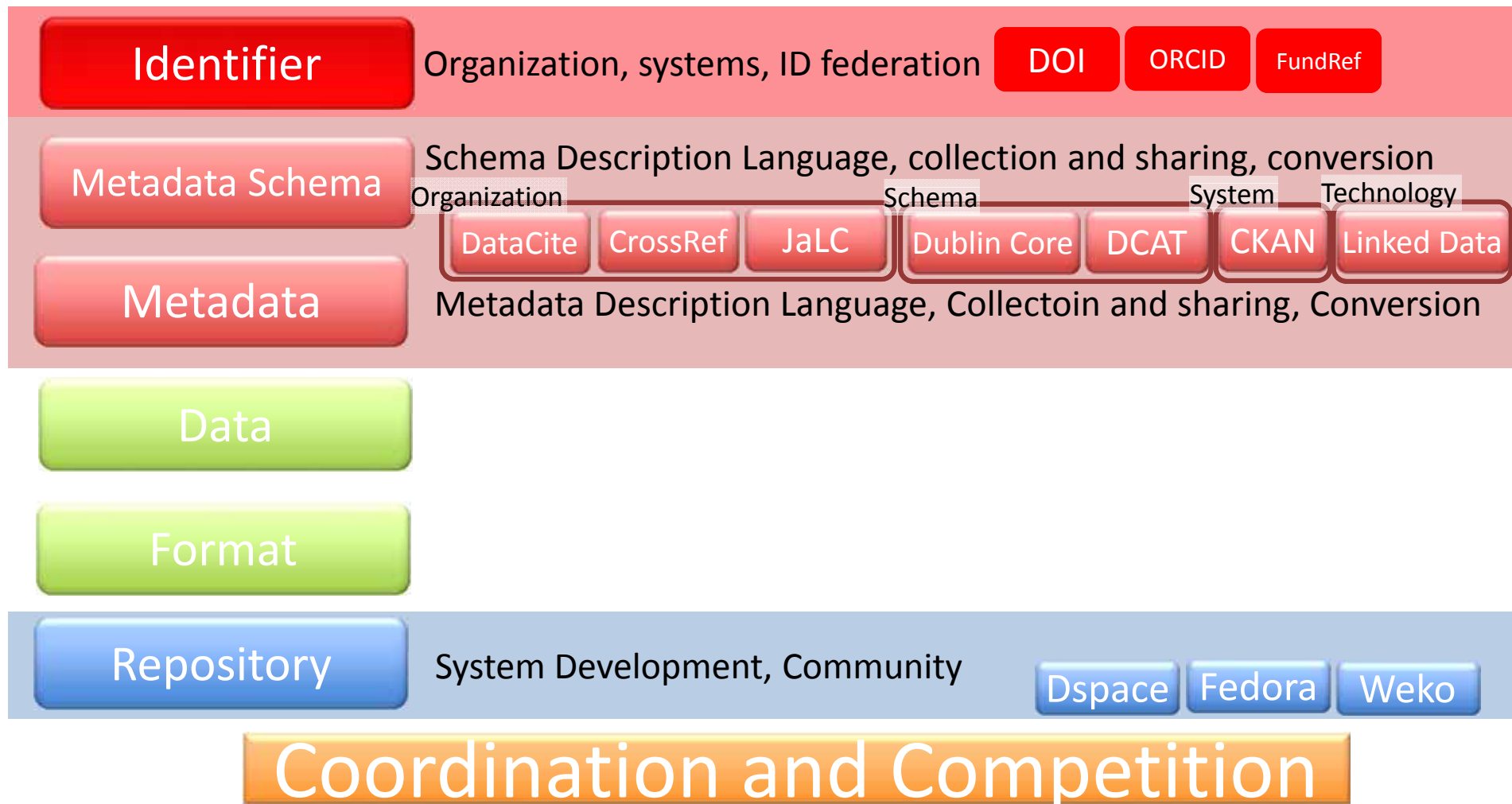
takeda@nii.ac.jp

ORCID: 0000-0002-2909-7163

Architecture of data sharing

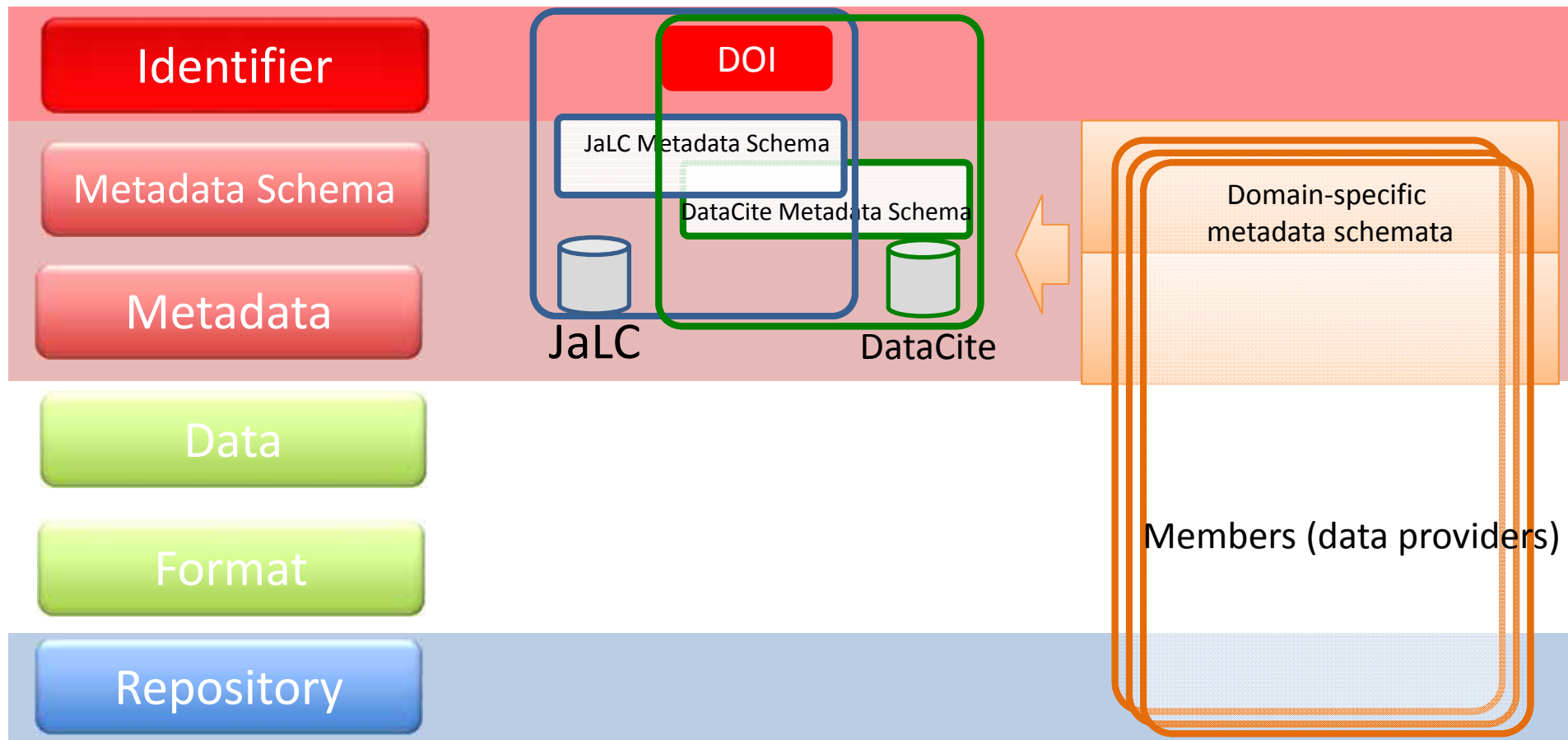


Architecture of data sharing



DOI

DOI in Architecture of Data Sharing

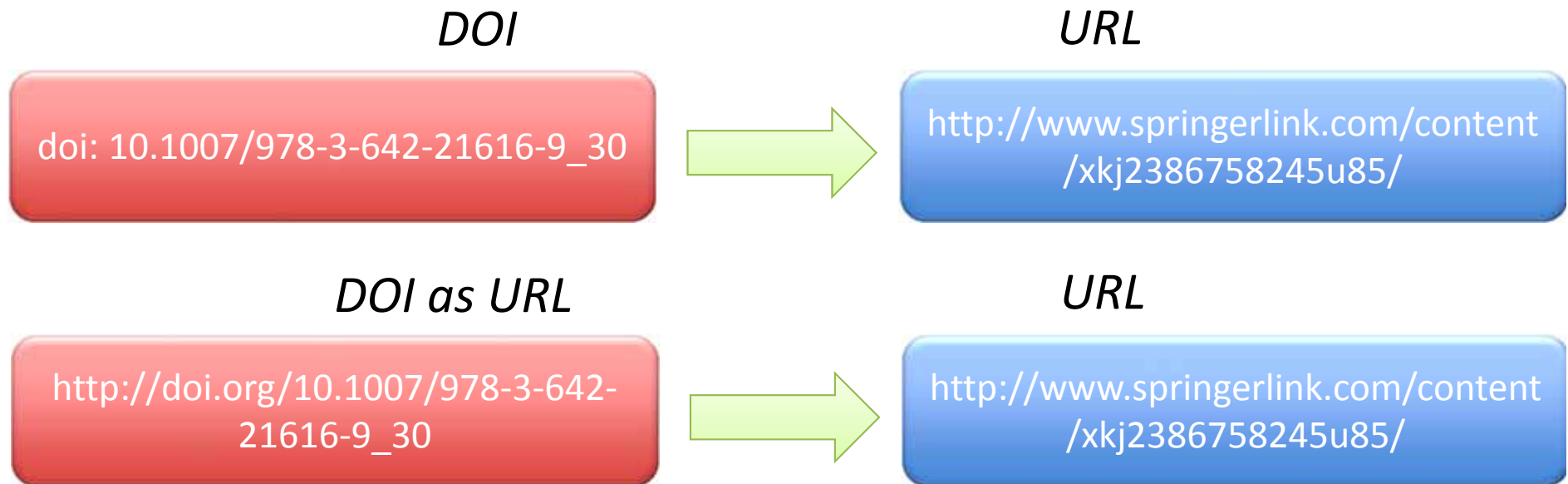


DOI (Digital Object Identifier)

- Service to translate DOI names to URIs containing digital objects
- Service managed by International DOI Foundation (IDF)
- Initially started by STM publishers to share identifiers for digital publications
- Distributed management
 - Delegation of registration tasks to Registration Agencies (RAs)

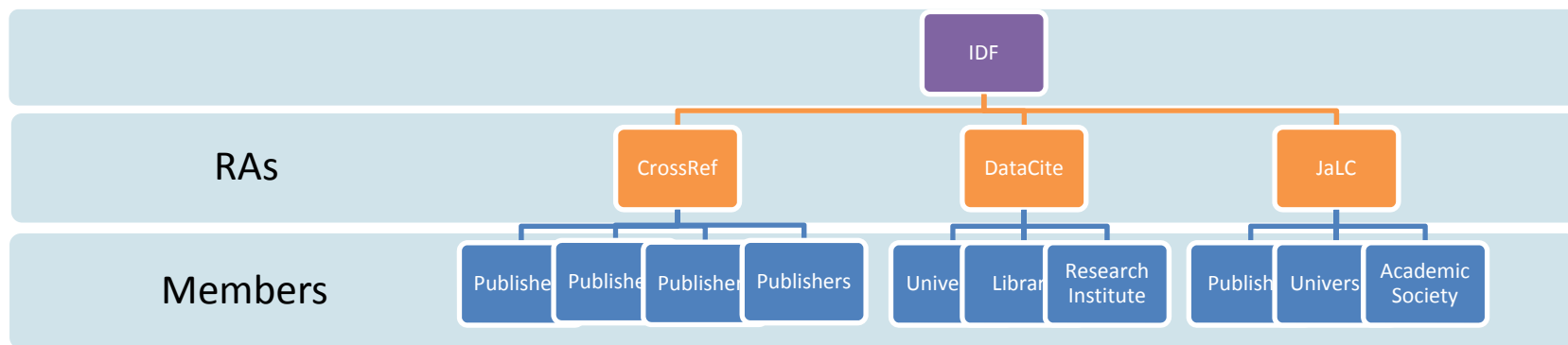
DOI (Digital Object Identifier)

- Service to translate DOI names to URIs containing digital objects



Management Structure of DOI

- There Layers: International DOI Foundation (IDF), Registration Agency (RA), members
- RAs contributes to IDF by registration to Registry DBs, management of Registry DBs, and members fees
- RAs offers services for DOI registration to their members
- Members can register DOIs to their digital objects through RAs



Roles of DOI

- Provide resolvable, persistent, interoperable links
 - Resolvable: standard syntax + mapping by handle system
 - Persistent
 - Technically: management of registry DBs
 - Socially: organizational operations and duties for members
 - Interoperability: sharing datamodel

Registration Agencies (RAs)

 airiti DOI	 crossref	 EIDR Entertainment Identifier Registry	 中文DOI
Airiti, Inc.	CrossRef	EIDR (Entertainment Identifier Registry)	ISTIC (The Institute of Scientific and Technical Information of China)
 cnki 中國知網 cnki.net	 DataCite	 Japan Link Center	 mEDRA
China National Knowledge Infrastructure (CNKI)	DataCite	JaLC (Japan Link Center)	mEDRA (Multilingual European DOI Registration Agency)
 Publications Office			
OP (Publications Office of the European Union)			

CrossRef

- Ensure accessibility and citation of articles and books in STM publications
- Started in 1999
- Largest and oldest RA of IDF
 - Most of DOI registered are via CrossRef
 - Members over 70 countries, most are publishers
- Functions
 - DOI Registration
 - Metadata Management
 - Bibliographic metadata
 - Citation
 - Services with metadata
 - Search for bibliographic metadata and citation
 - Reverse look up

DataCite

- IDF RA for research data
- a not-for-profit organization since 1 December 2009



The screenshot shows the DataCite website homepage. At the top, there is a navigation menu with links: ABOUT DATACITE, DATACITE FOR YOU, PARTICIPATE, EVENTS, NEWS, and JOIN DATACITE. The main content area features a blue header with the DataCite logo and a navigation menu. Below this, there is a section titled "DataCite & 'figshare for Institutions'" with a sub-heading "DataCite has been working with Figshare through EZID at the California Digital Library (CDL) for closing on 2 years now, and we are happy to extend that relationship with Figshare's recent institutional offering." A "Read more" button is visible. To the right of this text is a graphic featuring a colorful circular pattern of dots, a red heart, and a silhouette of a building. Below the main content area, there is a section titled "WHAT CAN DATACITE DO FOR YOU?" with five icons and corresponding text boxes:

- CITE YOUR DATA**: Data citation is fundamental as it enables easy reuse and verification of data, making it possible to track and quantify the impact of data. CiteSpace creates a...
- FORMAT YOUR CITATION**: Use the DOI Citation Formatter, a service created in collaboration with CrossRef, to format your citation, ensuring you adopt the correct format for your needs...
- FIND A REPOSITORY**: DataCite supports all researchers looking to deposit and/or find data, in collaboration with m4data.org
- FIND A DATASET**: DataCite Metadata Search is a service that allows people to search for datasets registered with DataCite, via the metadata associated with the datasets.
- GET YOUR DOI STATISTICS**: DataCite provides statistics for members on DOI registrations and DOI resolutions, filtered by Allocator, Data Center or Prefix.

Japan Link Center (JaLC)

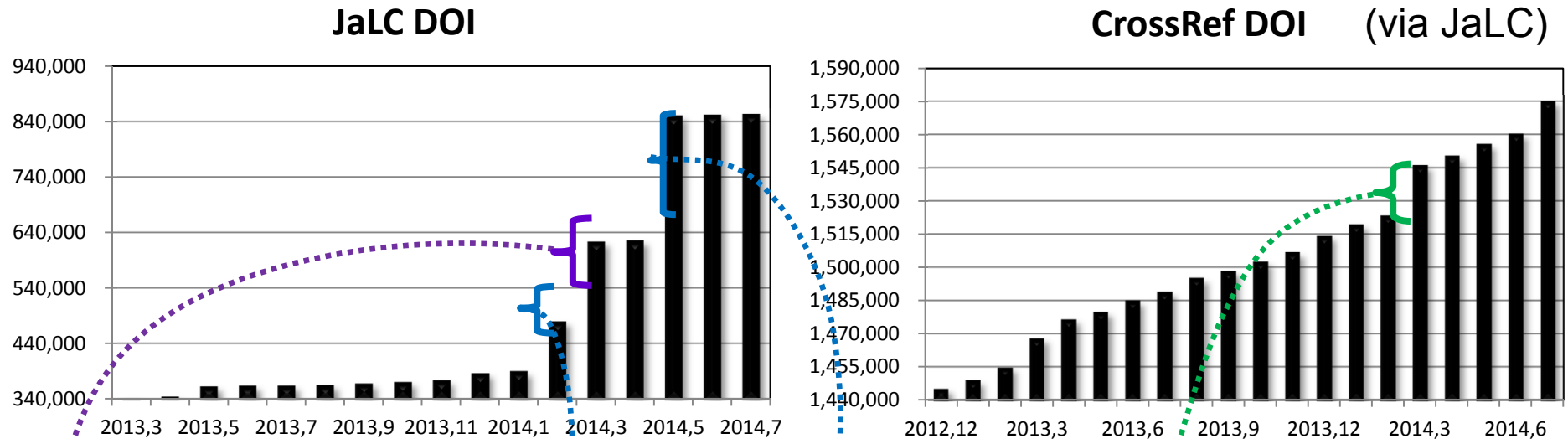
- Founded in March 2012
- Aimed to register DOIs for academic contents produced in Japan or in Japanese, to circulate information in Japan and overseas.
- Controlled by four national organizations:
 - Japan Science and Technology Agency (JST)
 - National Institute for Materials Science (NIMS)
 - National Institute of Informatics (NII)
 - National Diet Library (NDL)
- Operated by JST
- Membership system (Academic societies, Publishers, University libraries, etc)
 - 24 Members, 950 Associate members
- External coordination

JaLC is a member of CrossRef and DataCite(Mar. 2014)

Over 1,300,000 DOI registered



Number of DOIs



J-STAGE Article : 90,000 + 200,000 (JaLC DOI)

+ 20,000 (CrossRef DOI)

(Collective registration for the articles which had not been registered DOI)

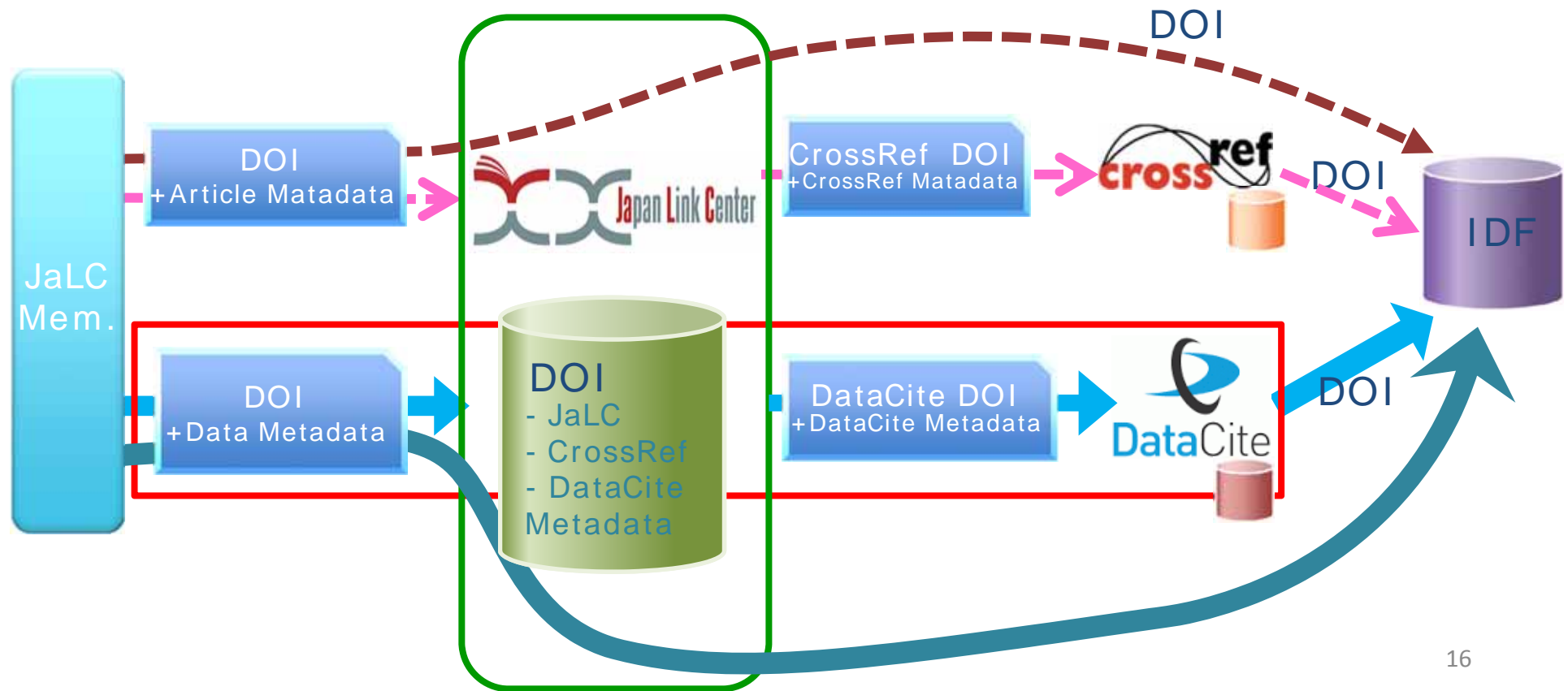
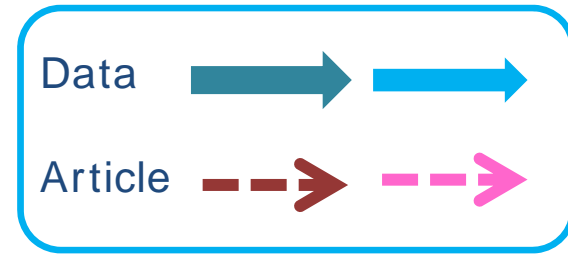
National Diet Library s Doctoral Thesis : 140,000

Total amount of DOIs: 2,427,391 (at the end of July 2014)

Content categories

Category	
Journal articles	
Journal articles	Dec.2012 -
University bulletins	Sep.2014 -
Conference proceedings	Mar.2012 -
Books	
Books	Jan.2015 -
Doctoral theses	Mar.2014 -
Reports	
Technical reports	Jan.2015 -
Governmental reports	Jan.2015 -
Research data	Jan.2015 -
e-learning resources	Jan.2015 -

DOI Registration Flow



Metadata Schemata

No	Contents type	Additional Metadata	External Deposit
1	Journal Article	journal name, ISSN, volume, number, title, page, etc.	CrossRef
2	Book	series title, chapter , ISBN, etc.	CrossRef
3	Research Data	size, geolocation, rights, signature, etc.	DataCite
4	E-learning	Learning Resource type, rights, etc.	
5	Other	(Basic metadata only)	

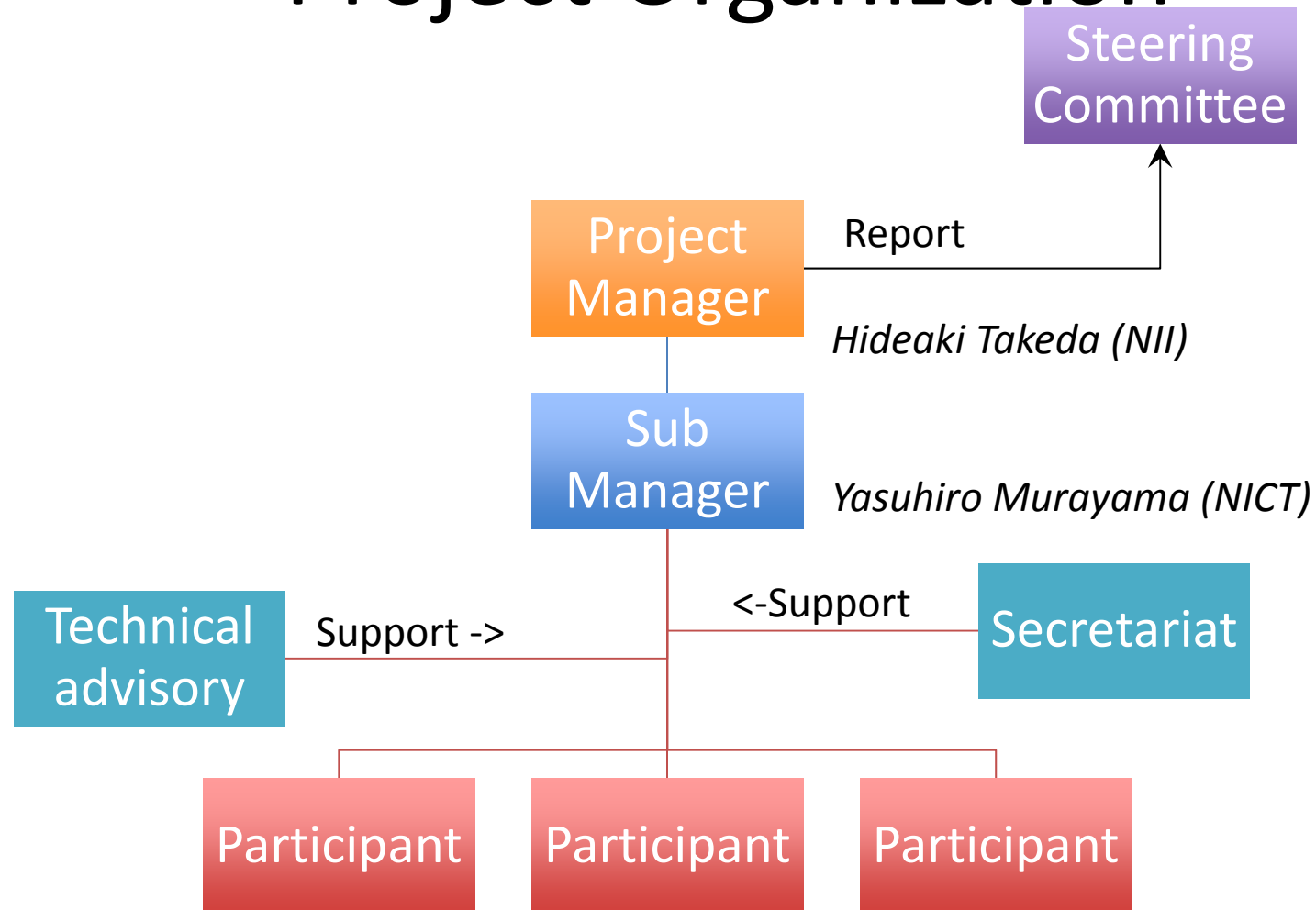
Basic metadata:

DOI, URL, title, author information, affiliation, researcher id, publication date, publisher, edition, related contents, funder

Experiment Project to register DOIs for Research Data

- Goal
 - Establish operation flows to register DOIs for research data
- Objectives
 - Set policies in registering DOIs for research data
 - Establish operation flows to register DOIs for research data with JaLC system.
 - Test Data DOI registrations
- October 2014 – October 2015

Project Organization



Participants are supposed to be research institutes, universities, etc. and chosen by public invitation.

Members of the project

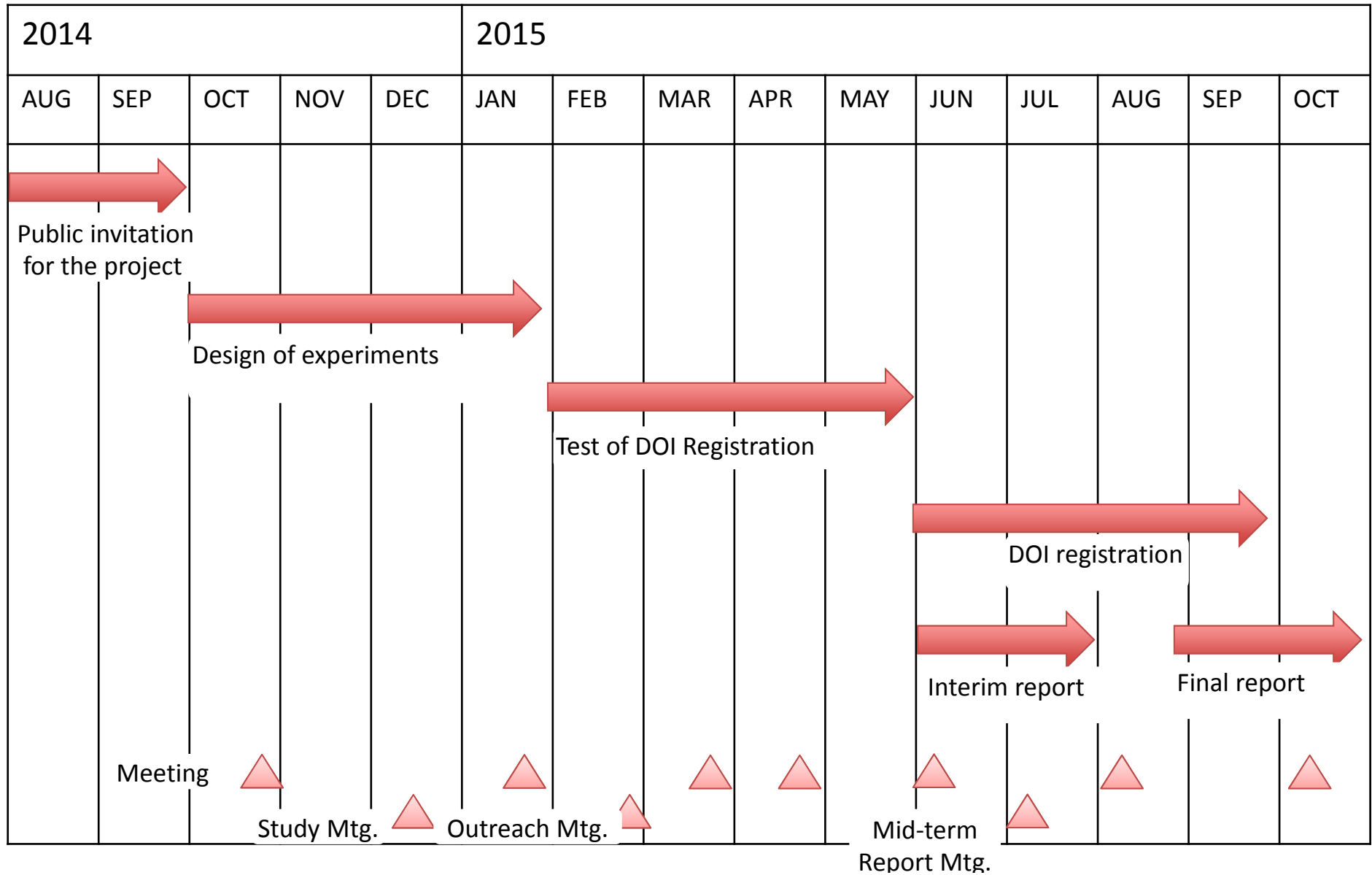


9 projects with 14 organizations

Members of the project

- National Bioscience Database Center (NBDC), Japan Science and Technology Agency (JST)
- National Institute of Polar Research (NIPR)
- National Institute of Informatics (NII)
- DIAS-P Project (National Institute of Informatics (NII))
 - Japan Agency for Marine-Earth Science and Technology (JAMSTEC)
 - University of Tokyo
 - Kyoto University
 - National Institute for Environmental Studies (NIES)
- National Institute of Advanced Industrial Science and Technology (AIST)
- National Institute of Information and Communications Technology (NICT)
 - Kyoto University
 - National Institute of Informatics (NII)
 - InfoProto Co.,Ltd.
 - Japan Aerospace Exploration Agency (JAXA)
 - National Institute of Polar Research (NIPR)
- Chiba University Library
- National Institute for Materials Science (NIMS)
- Neuroinformatics Japan Center, Brain Science Institute (BSI), RIKEN

Schedule



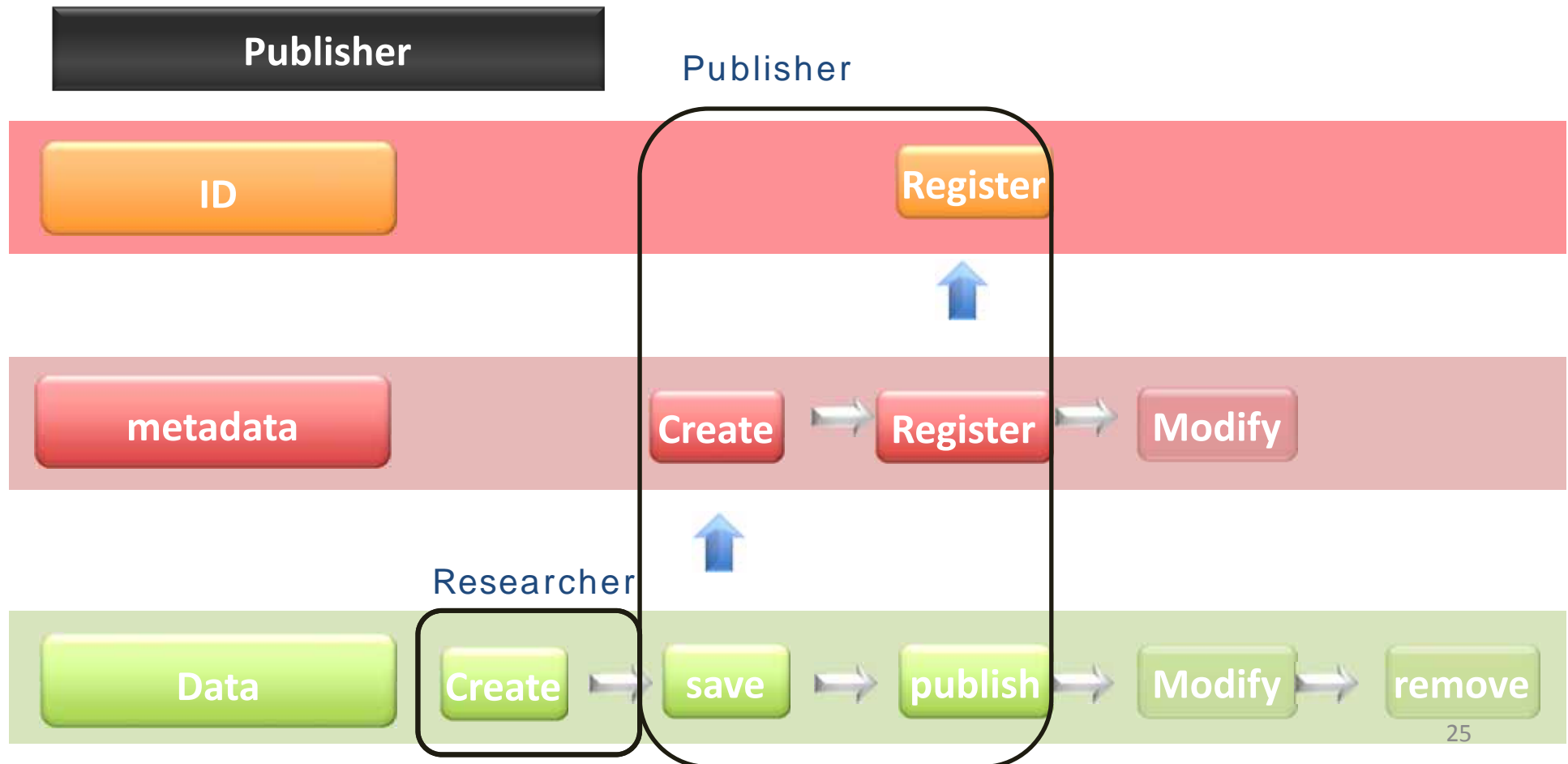
Issues in Data DOI

- Flow of operations
- Persistent access
- Granularity of data in registration
- Dynamics of data
- Landing page
- Quantity of data
- Applications

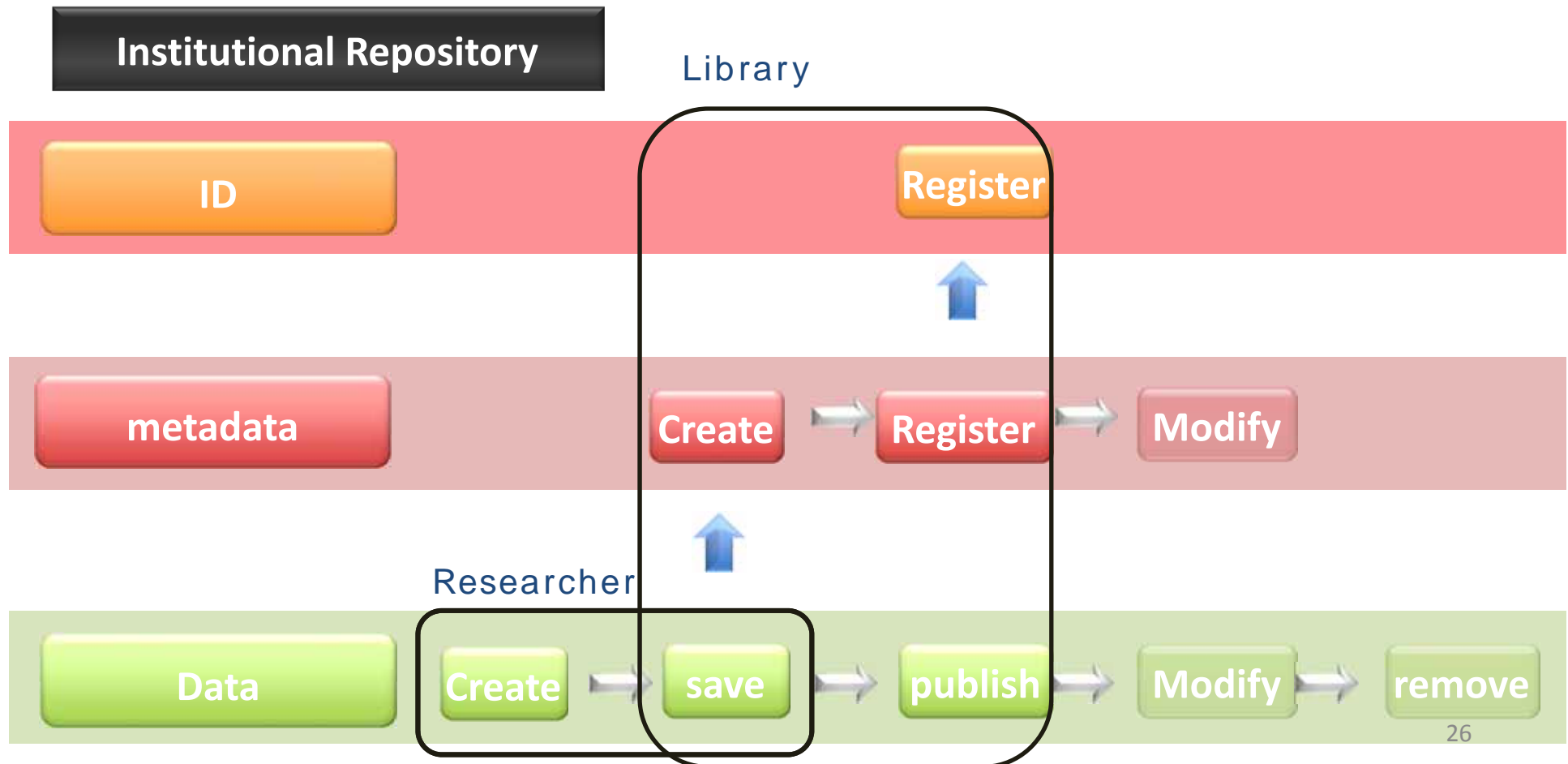
Issues in Data DOI

- Flow of operations: Who, When, How
 - Who registers data?: Researcher/Project manager/Librarian
 - When is data registered?
 - How is metadata provided for data?
 - Persistent access
 - What persistency can we expect for data?
 - Can time-limited projects participate? Who will ensure the persistency of the data?
- (ex.)
- ✓ The representative institute takes over all of the data
 - ✓ Registering DOIs only for data managed by real organizations among the members of the project

Life cycle of data and stakeholders - in case of literature -

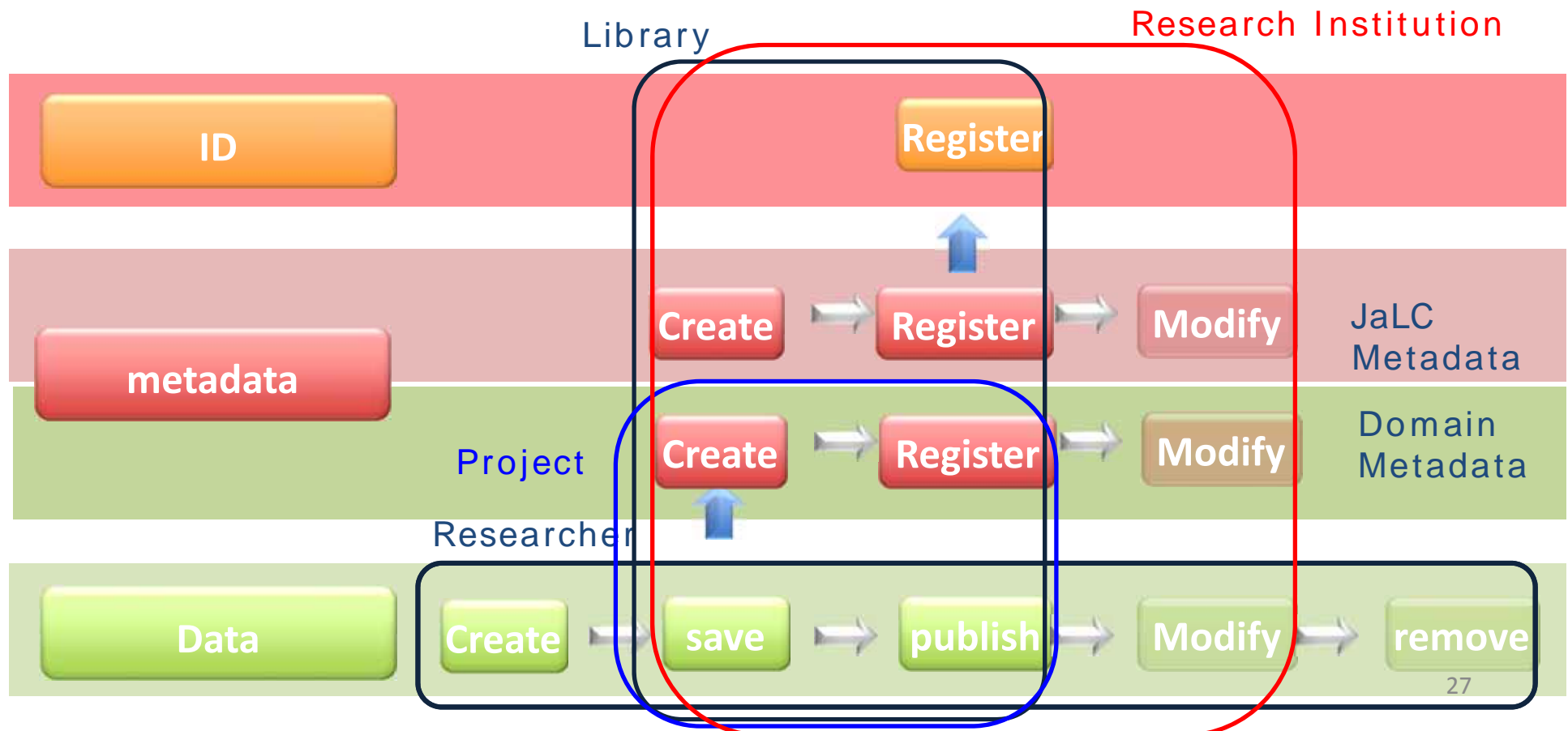


Life cycle of data and stakeholders - in case of literature -



Life cycle of data and stakeholders

- in case of data -



Issues in Data DOI

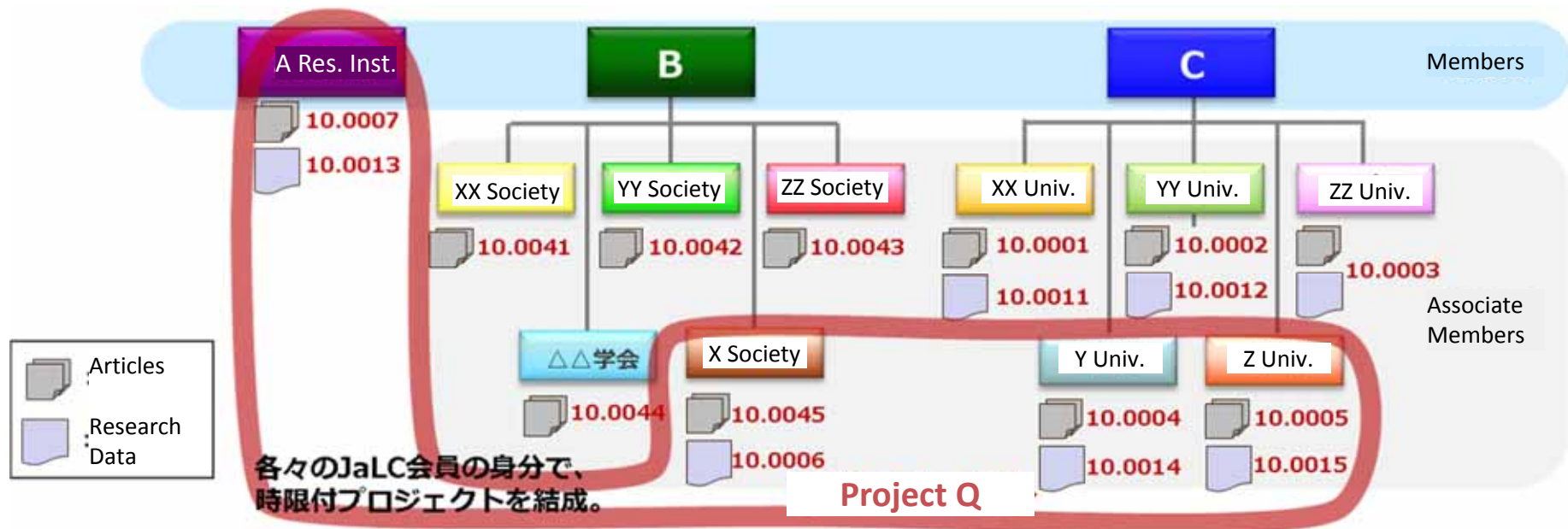
- Persistent access
 - What persistency can we expect for data?
 - Can time-limited projects participate? Who will ensure the persistency of the data?

(ex.)

 - ✓ The representative institute takes over all of the data
 - ✓ Registering DOIs only for data managed by real organizations among the members of the project

Projects and Institutions

- Research project may consist on members from multiple institutes



Issues in Data DOI (cont'd)

- Granularity of data in registration
 - Some aspects for granularity of data
 - Good for citation
 - Granularity of data itself
 - Observation data/Experiment data/Simulation data
 - Easy for access
 - Easy for management
 - Quantity of data

Issues in Data DOI (cont'd)

- Dynamics of data
 - Adding data after registration of DOI
 - Some options:
 - Different DOIs
 - Add relationship metadata to denote the relation to the original DOIs
 - Use the original DOI
 - Versioning: add the link to the new data while keep the link to the original data
 - History of changes in the single DOI
 - No descriptions (e.g., data in observing)

Issues in Data DOI (cont'd)

- Landing page
 - Metadata description
 - For open/closed data
- Quantity of data
 - Registering DOI for a large amount of data
- Applications
 - Citing DOIs for research data
 - Developing other applications

Recommendations for Data DOIs

- Recognition of variety of the nature of data
- Minimal Commitment
 - Persistency, Interoperability, Usability, manageability
- Design own “DOI Registration Policy”
 - Along with Institutional “Data Management Policy”
 - Define/recommend rules of DOI registrations on these issues.
 - Flow of operations
 - Persistent access
 - Granularity of data in registration
 - Dynamics of data
 - Landing page
 - Quantity of data
 - ...

Summary

- Data-sharing architecture
 - Interoperability should be guaranteed
 - Layers
 - ID/Metadata Schema/Metadata/Data format/Data/Repository
 - Cooperation and Competition
- DOI is the promising ID for data but different in use from one for literature
 - DOI Registration Policy is needed
- Data DOI is a good key with which various stakeholders can communicate practically to each other
 - “RDA Japan”?