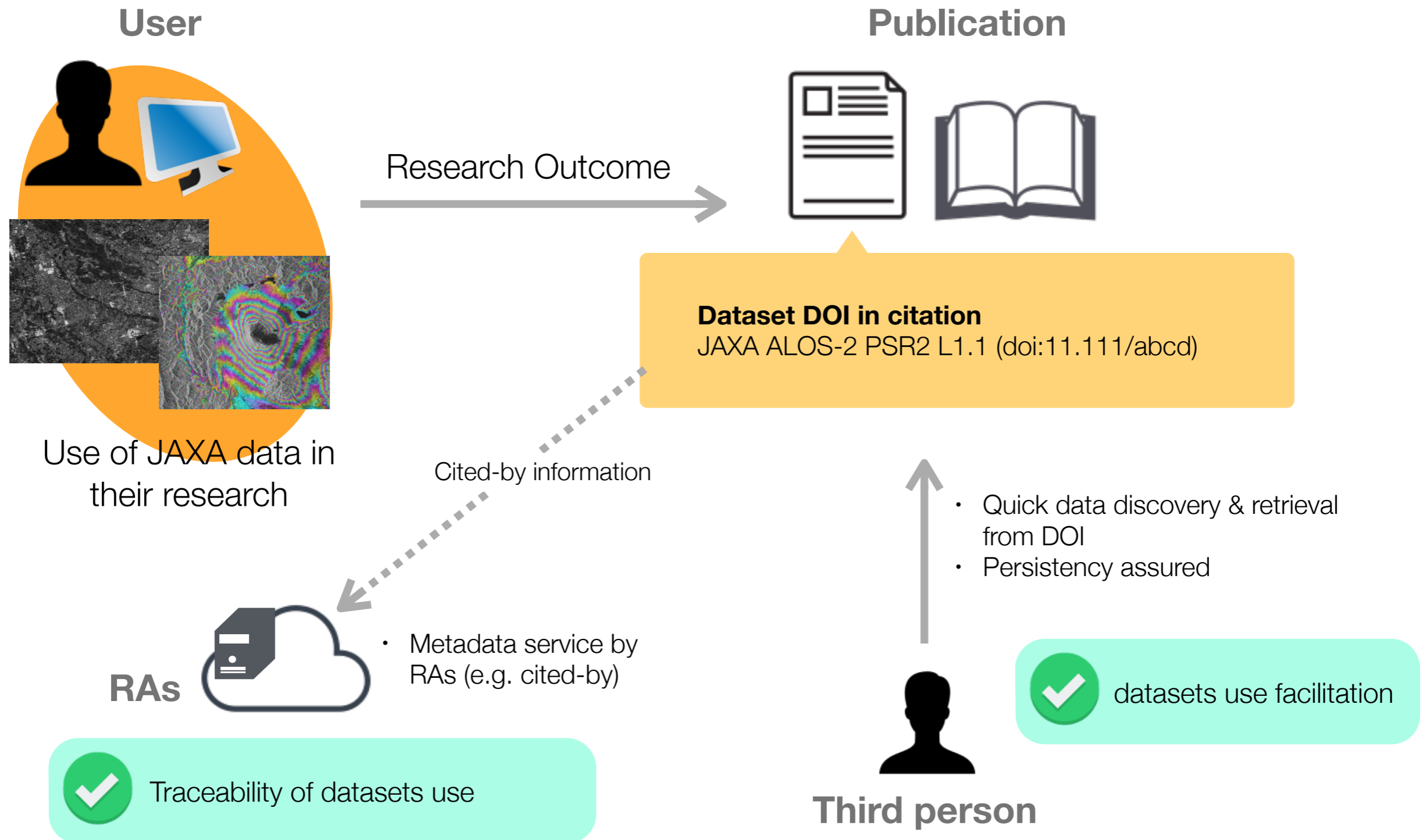# JAXA Status on
Persistent Identifier Activity

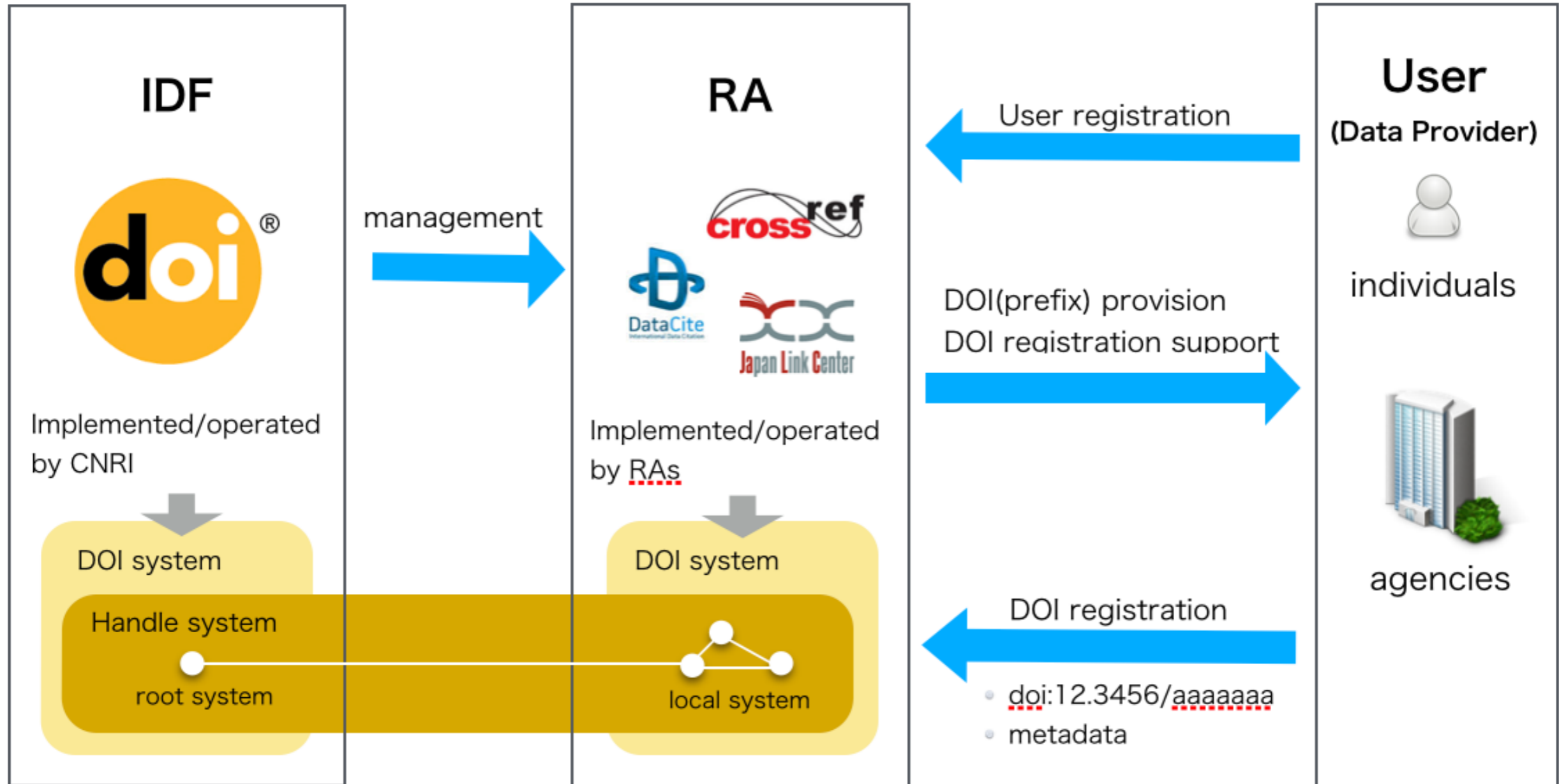WGISS-39, Tsukuba, Japan
Yoshiyuki Kudo (JAXA)

Status

- We have done a study on PID towards the end of our fiscal year (- Mar 2015)

  - ABCs of PID

  - PID implementations at CEOS agencies
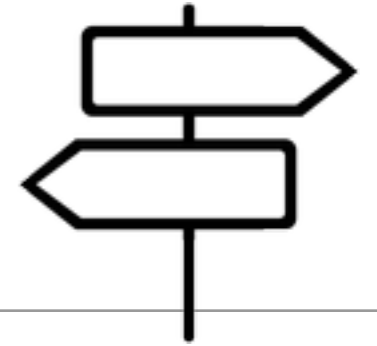
  - PID implementation impact on JAXA systems

EO data with PID

User

Use of JAXA data in
their research

Research Outcome →

Publication

Dataset DOI in citation
JAXA ALOS-2 PSR2 L1.1 (doi:11.111/abcd)

Cited-by information

RAs

- Metadata service by
  RAs (e.g. cited-by)

✓ Traceability of datasets use

- Quick data discovery & retrieval
  from DOI
- Persistency assured

Third person

✓ datasets use facilitation

Big picture　(IDF, RA, User)

Recent background concerning dataset PID

🇯🇵 Japan

- Dataset PID (DOI) discussion **officially started in 2014**

- Japan Link Center (**JaLC**) nominated as a **DataCite member** in Japan (2014-)

  - JaLC is a single choice for dataset DOI RA in Japan as of today

🌐 World wide

- Increasing interests and implementations within **CEOS agencies**

- **CEOS PID Best Practices**

- **GEO Data Management Principles** (final draft) has clear statements regarding PID

JaLC and dataset PID (DOI) status

- **"DOI assignment experiment on scientific datasets" is now ongoing led by JaLC**

  - Term : 2014-10 ~ 2015-09

  - 9 agencies participating, including DIAS
    (EO datasets in the order of peta-bytes)

- Wide open discussion with potential
  users/data-providers at JaLC in Feb 27 2015

(image by JaLC)

JaLC and dataset PID (DOI) status

- The Feb discussion **revealed 2 main users** (data providers) with **different motivation** for being part of the dataset PID assignment

  - **Individuals** (researchers, scientists)

    - Metadata repository + search service

    - cited-by metrics for justifying their research

  - **Data Centers**

    - More chance of datasets exposure (discovery and use by users)

    - cited-by metrics for tracing the datasets use

JaLC and dataset PID (DOI) status

- The experiment is still in progress

- JaLC will establish the PID assignment policy upon the end of the experiment (Sep 2015)

Survey of CEOS agencies' PID implementation status

Agency	PID	RA	PID to what ?	Implementation status
NASA	DOI	DataCite	Collection level dataset	Operational
NOAA	DOI	DataCite	Collection level dataset	Test phase
esa	DOI	DataCite	Collection level dataset	Test phase
DLR	DOI	DataCite	Collection level dataset	Test phase
EUMETSAT	DOI	DataCite	Collection level dataset	Test phase
CNES	ARK	Unknown	MIssion, Collection, Granule	Test phase

(survey by JAXA as of Dec 2014)

Key points for JAXA's PID implementation

(1) Which PID to adopt ?   (DOI ? ARK ? URN ? …)

(2) PID assignment to which digital object ? And granularity ?

(3) PID string syntax (naming)

(4) Landing page

(5) RA selection and contract

(6) What needs to be done to the existing systems to support PID operation ?

(7) Policy for the long-term (persistent) commitment

(8) Define operation procedure wrt PID

📝 Preliminary Study Conclusion

- **DOI for PID**

  - CEOS Best Practice recommendation

  - No clear advantages of using other (e.g. ARK)

  - JaLC is promoting DOI assignment for datasets

- **Opaque character string for DOI**

  - **"Identifier"** should have no meaning by itself

  - CEOS Best Practice recommendation

- **DOI assignment to EO dataset collection**

  - CEOS Best Practice recommendation

  - Not on browse, ancillary data, or NRT data

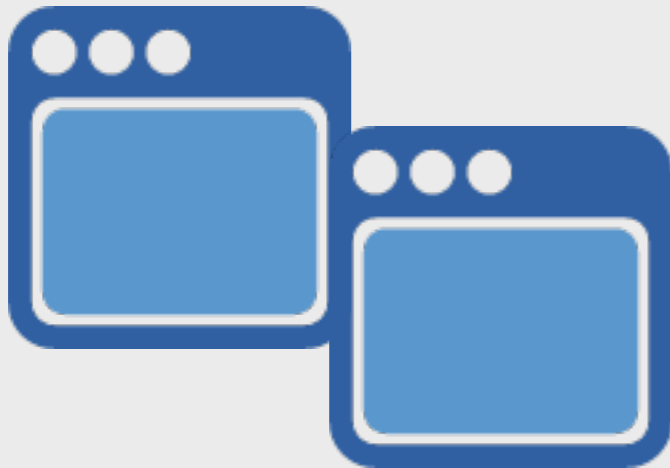📝 # Preliminary Study Conclusion

- **Landing page**

  Well description of the dataset desired, navigation for the data access should appear in the page.
  2 possible options :

#1

Product page by each mission

Each mission (e.g. ALOS2) has their product description page

👍 Can be descriptive as needed (in Japanese)

👎 • Cost
    • MMI consistency from mission to mission

#2

CEOS/IDN dataset page

Platforms>Earth Observation Satellites>GOSAT

GOSAT Level 3 global CH4 distribution (SWIR) product
Entry ID: GOSAT_Level3_of_TANSO-FTS_SWIR_CH4

[Get Data]   [Update this Record]

Summary
Abstract: The FTS SWIR Level 3 data products are generated by interpolating, extrapolating, and smoothing the FTS SWIR Level 2 column-averaged mixing ratios of CO2 and CH4 on a monthly basis. A geostatistical calculation technique called Kriging method is applied. The values are gridded to 2.5-degree cells. The FTS TIR Level 3 data products are global maps of CO2 and CH4 at several pressure levels that are processed in the same manner as above. These data products can be utilized for visualizing local highs and lows of the greenhouse gases on a global scale.L3 global CH4 distribution (SWIR) product is the CH4 column-averaged mixing ratio data projected on global map.

Related URL
Link: GET DATA
Description: GOSAT User Interface Gateway (GUIG)

Link: VIEW PROJECT HOME PAGE

Geographic Coverage

👍 Registered and available already - no cost

👎 English page only

📝 Preliminary Study Conclusion

- **RA selection and contract**

  - **DataCite** is the #1 candidate

    - Seen implementation practices by NASA、ESA etc.

    - CEOS Best Practices Recommendation

  - DataCite member in Japan ＝ **Japan Link Center (JaLC)**

    - Membership basis, annual fee

    - No direct registration to DataCite (Germany) allowed

📝 Preliminary Study Conclusion

- **What needs to be done to the existing systems to support PID**

1 Prepare Landing Pages

2 Put in place "DOI assignment system"

archive data
Bulk DOI generation in the catalog system

In-flight mission data
DOI assignment at data processing systems
- each process result to include DOI

📝 Preliminary Study Conclusion

- **What needs to be done to the existing systems to support PID**

3 DOI (and metadata) ingestion to RA and maintenance (keeping everything current)

- DOI registration upon new product (collection) release
  - Including metadata, landing page
- The entity (system) that assigns DOI should be in charge

4 Metadata addition to the Catalog system

- DOI as a new metadata field

🖉 Preliminary Study Conclusion

- **What needs to be done to the existing systems to support PID**

5   Portal should be enabled to :

- Show DOI along with other metadata
- Show instruction on how to cite the datasets using DOI
- allow the download files to include above (DOI and citation instruction)

📝 Preliminary Study Conclusion

- **Policy for the long-term (persistent) commitment**

  - Long-term commitment
    - Annual membership fee to RA
    - DOI generation and registration
    - Maintenance of DOI and landing pages
  - With this commitment in mind, the policy should be explicitly stated in a document, and should be shared among relevant missions/projects.

📝 Preliminary Study Conclusion

- **Define operation procedure wrt PID**

    - should be incorporated in operation manuals

    - should be shared among related mission systems

Findings

- Increasing PID use/implementation in CEOS agencies

  - NASA, ESA, NOAA, CNES, EUMETSAT, …

  - CEOS PID Best Practices

- Dataset-DOI gathering momentum in Japan with the leadership of JaLC

- DOI implementation requires cost

- DOI-included citation should become a firm practice to justify the involvement and find clear benefits (cited-by traceability)

- Right timing to have PID systems ? - Still looking closely at ongoing related activities for the final judge :

  - GEO, CEOS, DIAS, JaLC, …