2014.02.14 宇宙科学情報解析シンポジウム

超高層物理学分野における データ集約型の科学に関する調査

(情報・システム研究機構 H25年
第一回 融合シーズ探索提案)

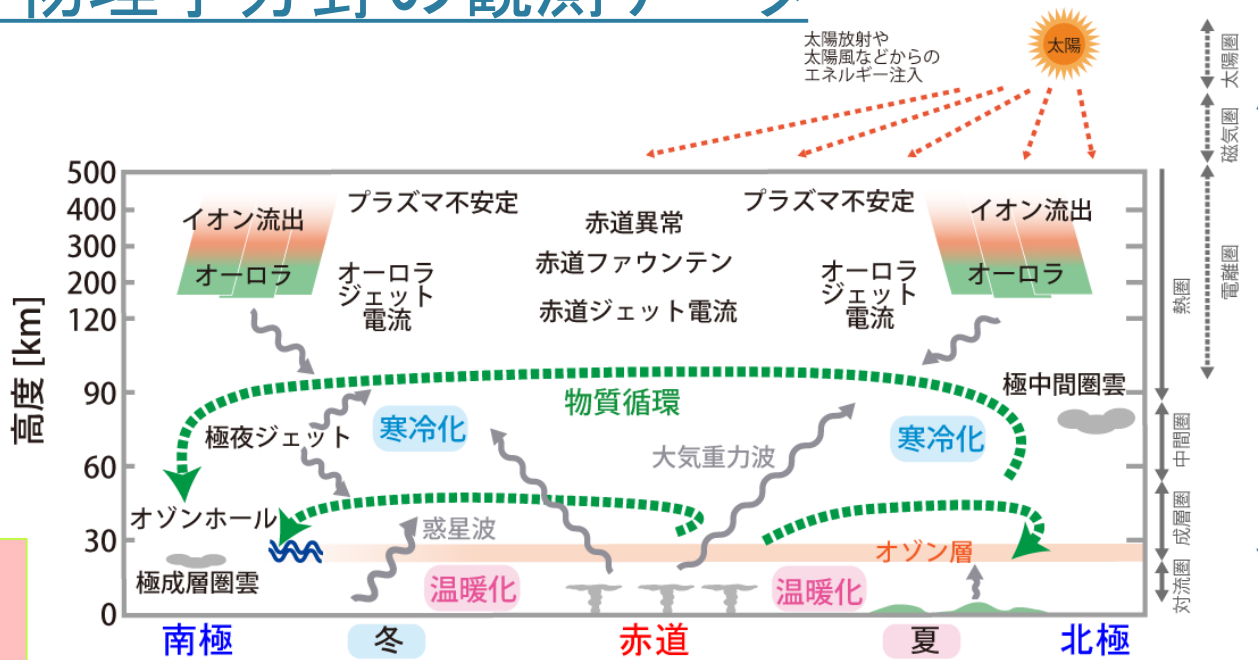
小山幸伸（京都大学）
蔵川圭（国立情報学研究所）
佐藤由佳（国立極地研究所）
田中良昌（国立極地研究所）
阿部修司（九州大学）
池田大輔（九州大学）

This document is provided by JAXA.

超高層物理学分野の観測データ

- 地表から宇宙空間に渡るグローバルな現象を反映
- 様々な時空間変動の現象を含む
- 取り扱う物理量が極めて多種

長期データを用いた統合解析が必須



- 日本では、様々な観測データは、<u>各研究機関において分散管理されている。</u>

IUGONET

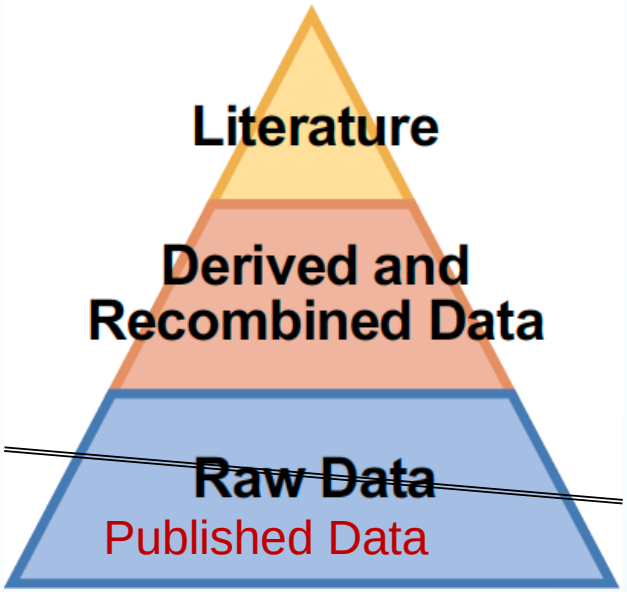
•メタデータ・データベース(データ流通向上の取り組み)
•データ解析ソフトウェア(データ利活用の取り組み)

This document is provided by JAXA.

原点に立ち返って、学術情報一般を考える。

科学的データの階層

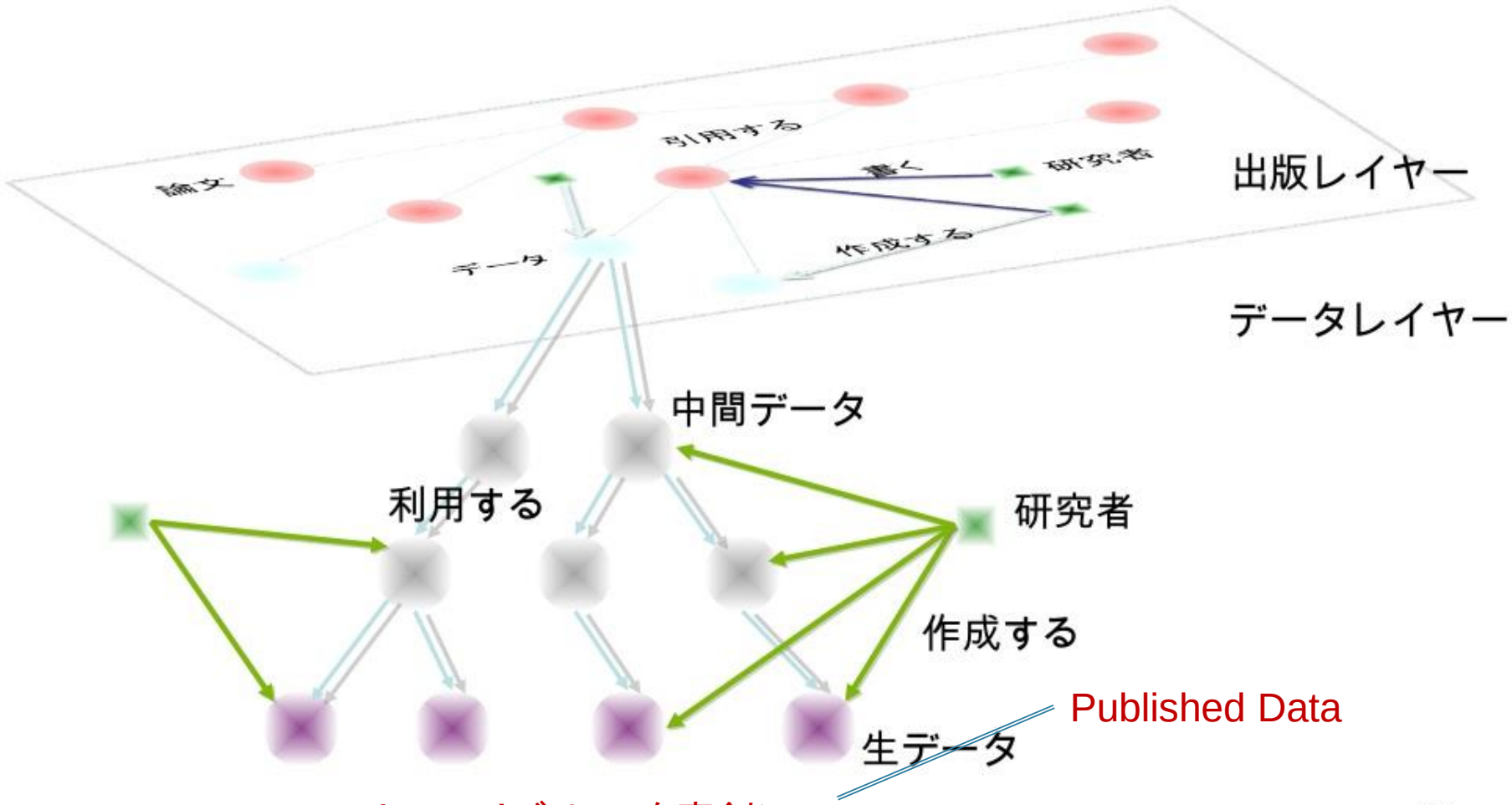
- 3つの階層
  - Literature
  - Derived and Recombined Data
  - ~~Raw Data~~ Published Data
- 階層の要素
  - 同一の階層の要素は互いに関係を持つ
  - 隣り合う階層を構成する要素が互いに関係を持つ
- ネットワークが分野をまたいだ要素の統合と関連を可能にする。



Tony Hey, Stewart Tansley, & Kristin Tolle (Eds.). (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research.
Retrieved from http://research.microsoft.com/en-us/collaboration/fourthparadigm/default.aspx

This document is provided by JAXA.

データレイヤーとの相互展開



トレーサビリティを真剣に
考える必要性

This document is provided by JAXA.

The Loyal Society of London Philosophical Transactions

４代　家綱時代

Boyleの実験結果には到達しない。

(10)

An Account of a very odd Monstrous Calf.

By the same Noble person was lately communicated to the *Royal Society* an Account of a very Odd Monstrous Birth, pro- ... *Limmington* in *Hamshire*, where a Butcher, having ... Cow (which cast her Calf the year before) to be co- ... might the sooner be fatted, killed her when fat, ... Womb, which he found heavy to admiration, ... which had begun to have hair, whose hinder ... ynts, and whose Tongue was, *Cerberus*-like, tri- ... his Mouth one, and one in the midst-Between ... and the Elinder-leggs was a great ... id: The *Sternum*, or that part of the P ... was also perfect Stone; and the Stone ... twenty pounds and a half, the outside ... reenish colour, but some small parts being ... appeared a perfect Free-stone. The Stone, acco ... Letter of Mr. *David Thomas*, who sent this Account ... is with Doctor *Haughteyn* of *Salisbury*, to whom he al ... for further Information.

Of a peculiar Lead-Ore of Germany, *and the Use*

... since, sent hither out of *Germany* from an ... List of several *Minerals* and *Earths* of that ... y, together with a *Specimen* of each of ... re was a kind of *Lead-Ore* which is more ... rest, because of its singular use for *Es-* ... ng that there is not any other *Metal* ... d in the *Upper Palatinate*, at a place called ... two sorts of it, whereof one is a kind of ... e, and a ... ll good Leads the other not so ... farinace ... y information, coming along ... y are fetcht, not from under the ground, but, the Mines of that place having lain long neglected, by reason of the Wars of *Germany* and the increase of Waters, the people, living there-

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions* (1665-1678) www.jstor.org

doi:10.1098/rstl.1665.0007

R. Boyle

タイトルに加え、著者もセンタリングされた！

(683)

Some Observations

Concerning the Comet, *that hath lately appear'd in foreign parts, communicated from* Italy *and* Portugal.

The Italian *Account, given by* Gio. Domen. Cassini, *we thus English.*

ANno 1668. the 10th of *March* h. 1. of the following night, (after the *Italian* way of counting) at *Bononia*, I observed a path of Light extended from the *Whale* through *Eridanus*; which I judged to be the Train of a *Comet*, both by the Figure and Colour, as also because that the direction of it, being by the Fancy continued, seem'd to proceed unto the 21. degr. of *Pisces*, where the *Sun* then was, and so tended to the part opposit to the *Sun*, like other Comets. By its extreme point it reached to that Star in *Eridanus*, which is called the Fourteenth by *Bayerus*: But it issued out of the Horizontal Clouds, so that I thought, the Head of the Comet was either vailed by them, or hid under the *Horizon*. It followed the Motion of diurnal revolution West-ward, and it was to be seen about the second hour of the night; for then it was demersed in the mists of the *Horizon*.

Not far from its pointed end Eastward, there appeared a *Star*, equal to the brightest of the *Fourth magnitude*, almost in the same place, where was observ'd the *Comet* of A. 1664. *Decemb.* 31. which Star was not *then* seen, nor at other times elsewhere, nor is described in any Catalogue, on any Globe or Map, that I can learn; which therefore I deem to be a *New* one, that is, of new appearance.

March 11. at even the *Horizon* was in the *West* overcast with thin Clouds; among which after one hour of the night, there was seen a Brightness in the *Whale*, at least for half an hour, which was very like the splendour of *Venus*, likewise vailed by thin clouds.

March 12. at night, the lower parts of the Heavens in the *West*

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions* (1665-1678) www.jstor.org

約３５０年前に論文スタイルが確立！

doi:10.1098/rstl.1668.0016

G. D. Cassini

This document is provided by JAXA.

論文の変革の兆し

電子化

- PEPS 誌、 EPS 誌　→　オープンアクセス

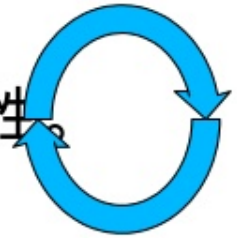
- PLOS ONE, altmetrics
  - 雑誌単位の IF 評価　→　論文単位の評価

- PANGAEA & ELSEVIER
  - 地球科学のデータ出版とジャーナル出版の融合

This document is provided by JAXA.

持続可能なデータ公開活動の環境構築が、地球物理の課題

- 共通認識
  - 観測・データ整備等のデータ公開活動の重要性。



- 現状
  - 研究者が教育・研究の合間にボランティアで活動。
  - データ公開活動<u>そのもの</u>は評価の対象でない。

DOIを付与し、Data Publication, Data Citationの仕組みを確立することで、どれだけデータが科学論文に引用されているか、定量化可能になる。しがらみ（特許、倫理、国防、産業）の少ない超高層物理学分野に

This document is provided by JAXA.

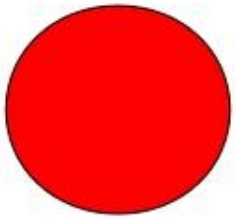
データ公開活動の収益構造の確立

1. DOI をデータセットに付与する。

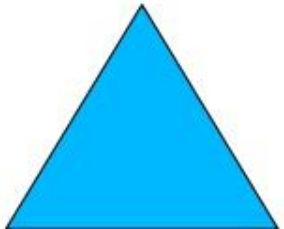
2. ジャーナルから引用する。

データ提供者側の

- DOI と ORCID がポイント … 論文の世界で使われているものと同じ識別子を、データ用に意図的に使う！



データ提供者側



データ利用者側

資金提供者側

This document is provided by JAXA.

DOI付与の流れ（予定）の一部 (他の流れもある)

NICT

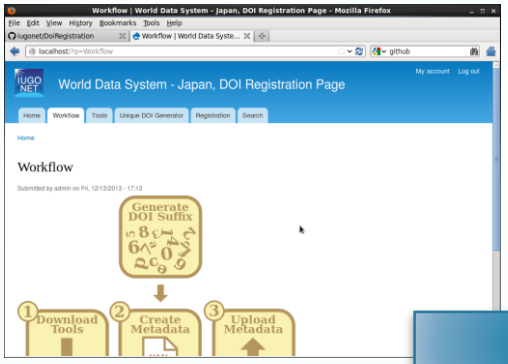
WDCs in Japan

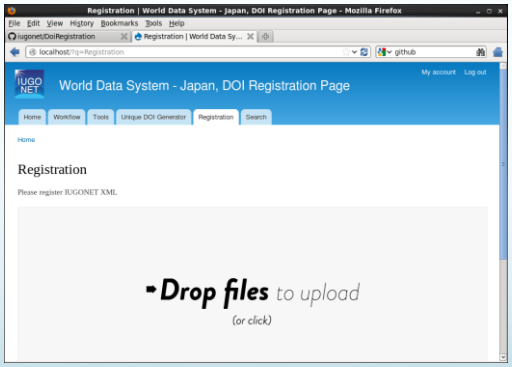
地磁気, Kyoto

電離層, NICT

オーロラ, NIPR

宇宙線, NAGOYA





IUGONET
メタデータ

日本にあるRA

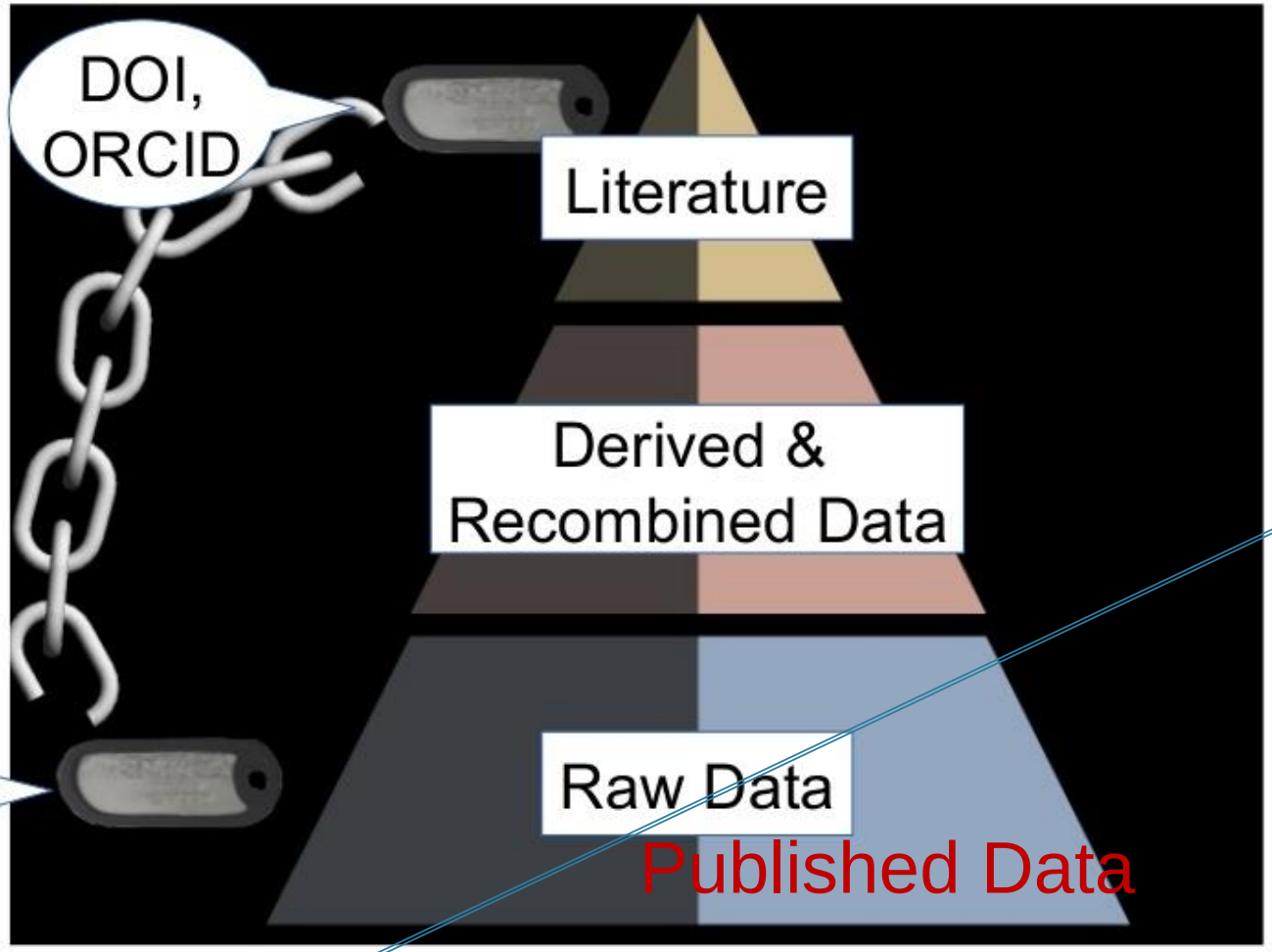
Japan Link Center

Japan Link Center（通称：JaLC（ジャルク））へようこそ！

2014年夏ごろには、この流れでDOI付与されるか！？

This document is provided by JAXA.

文献とデータセットの結合



DataCite

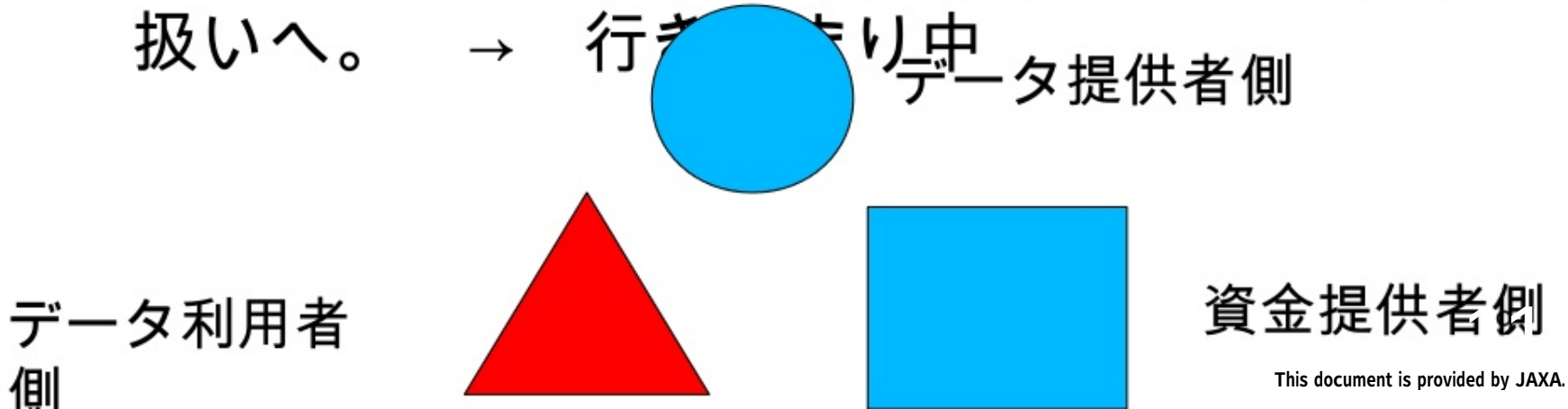
Japan Link Center

DOI,
ORCID

This document is provided by JAXA.

文献とデータセットの不完全

- 現状の Data Publication & Citation は、 Journal Publishing 目線。　→　科学的な REUSE×
- （海外の）図書館業界（教育 + 研究支援）
  - ビッグディール問題の次は、科学データ取り扱いへ。　→　行き詰まり中



This document is provided by JAXA.

DataCite XML Schema Ver.3

ID	Property	Obligation
1	Identifier	M
2	Creator	M
3	Title	M
4	Publisher	M
5	Publication Year	M
6	Subject	R
7	Contributor	R
8	Date	R
9	Language	O

ID	Property	Obligation
10	ResourceType	R
11	AlternateIdentifier	O
12	RelatedIdentifier	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description	R
18	Geolocation	R

科学的な用途としては、
データセットの詳細情報が不足

This document is provided by JAXA.

科学的な用途には？

- ドメインに特化したメタデータを用意する必要がある。

↓

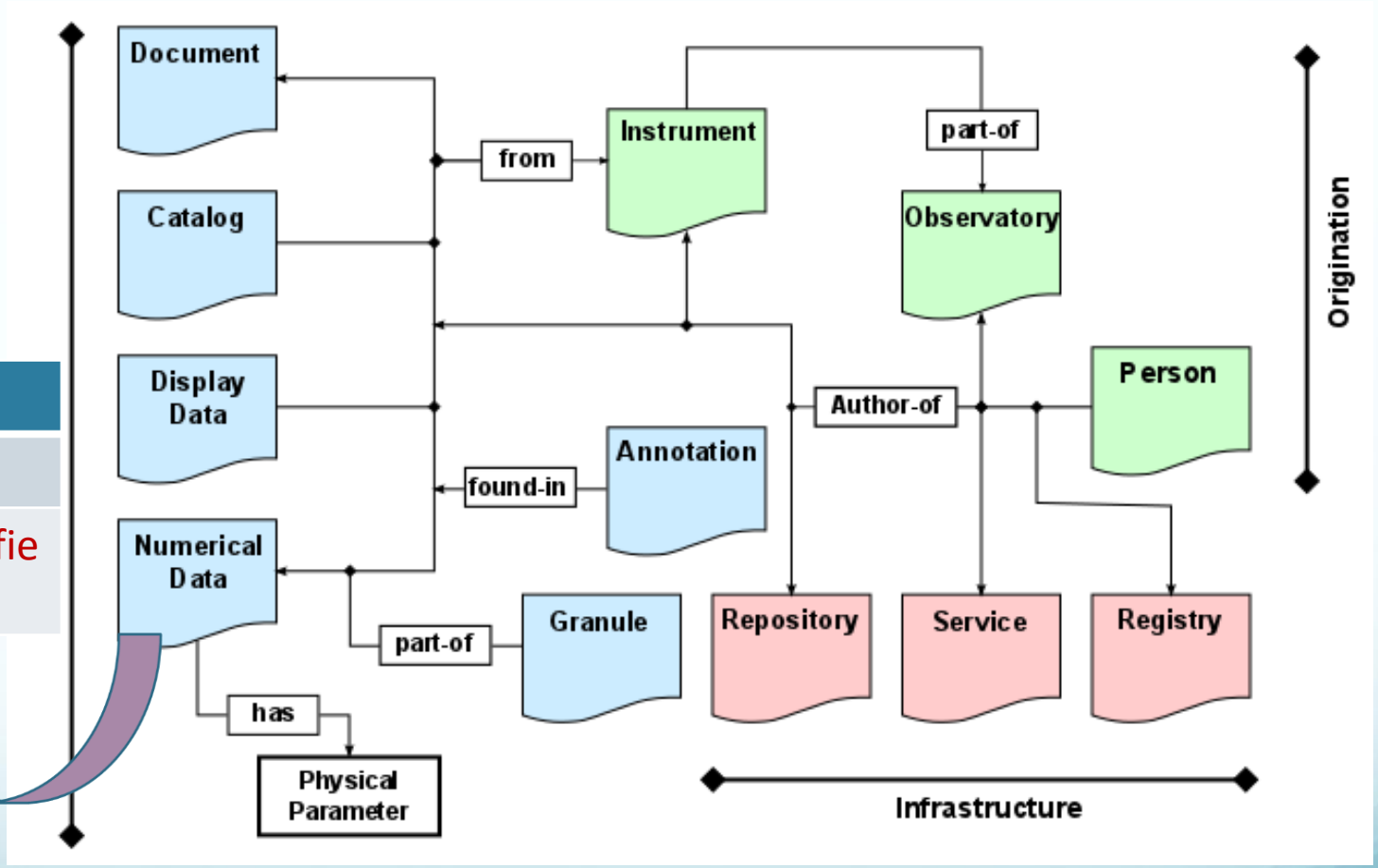
- 超高層物理学ドメインでは、DOI Registration Agencyが管理しているメタデータと、IUGONETメタデータの連携が◎

This document is provided by JAXA.

その1: JaLCのメタデータ
← IUGONETのResourceID

DataCite/JaLC

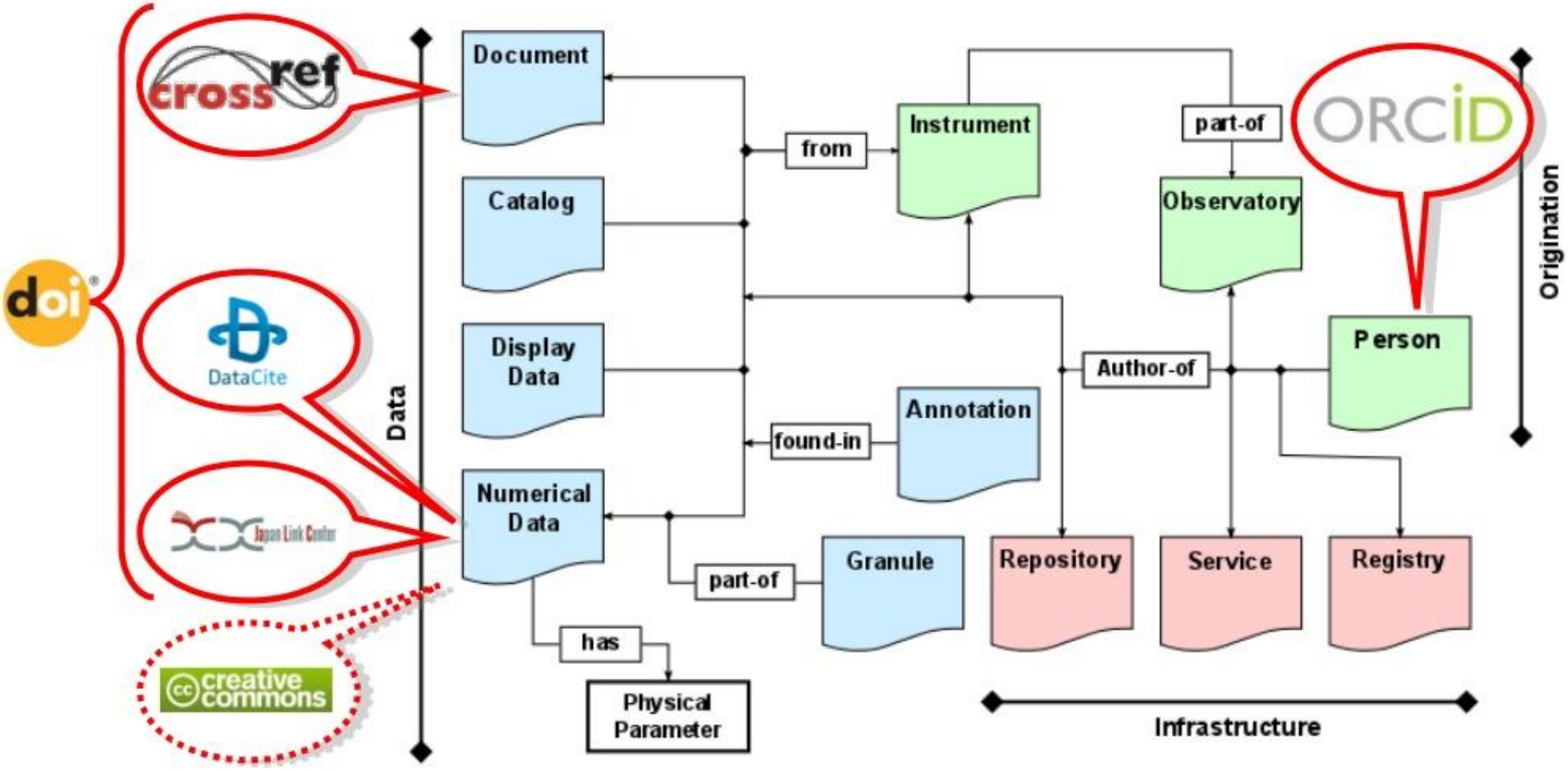
ID	Property
10	ResourceType
11	AlternateIdentifier



SPASE Data Model

This document is provided by JAXA.

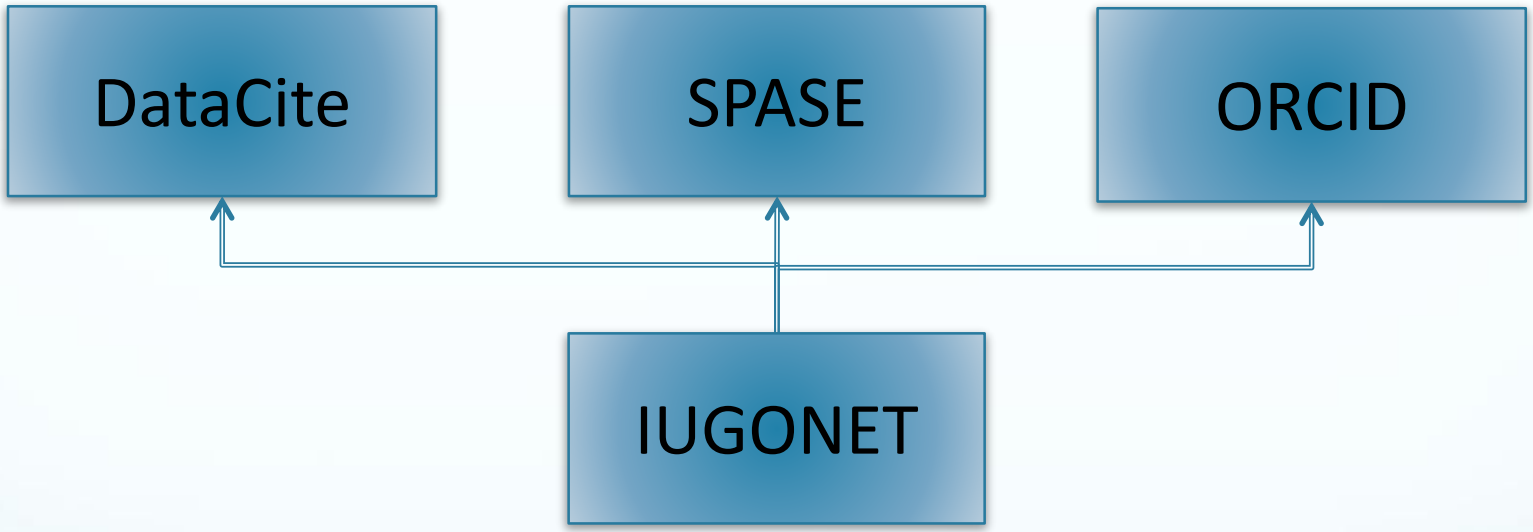
前述の動きに呼応した IUGONET XML Schema の再定義



学術出版を頂点とした科学データの REUSE の準備をした。

This document is provided by JAXA.

IUGONET XML Schema Ver. 2_2_2_5 (RC)



- CC等のライセンス情報も入力可能な用に拡張済み。
- JaLCのスキーマをincludeするかどうかはTBD。

This document is provided by JAXA.

まとめ

- 日本における、データセットへのDOI付与活動のお手伝いをした(JaLC, WDCs)。
- データ集約型科学に向けた第一歩として、超高層物理学のデータセットへDOIを付与する準備をしている。そして、超高層物理学分野におけるメタデータ・スキーマを整備している。
- JpGUジャーナル(PEPS誌)、SGEPSS等のジャーナル(EPS誌)の投稿規定に、Data Citationを含めるよう働きかけが必要。 → これを動かすためには、「データ引用した学術論文」のモックアップを機関リポジトリ等へアップロードし、そのデータセットへの到達性をデモする。
- 宇宙科学情報学解析論文誌も、Data Citationに関する投稿規定を整備するのはどうでしょう？

This document is provided by JAXA.