

超高層物理学のための メタデータ考察

Consideration of the metadata for upper atmospheric research

Morio Yamauchi **ORCID : 0000-0002-4077-8111**

Yusuke Miyanari **ORCID : 0000-0001-9064-7956**

Yukinobu Koyama **ORCID : 0000-0001-5363-3870**

Advanced Course for Electrical, Electronic and Information
Engineering, National College of Technology, Oita College

話題

0. 自己紹介

1. IUGONETメタデータについての考察

2. 小規模研究室における観測データ共有

自己紹介

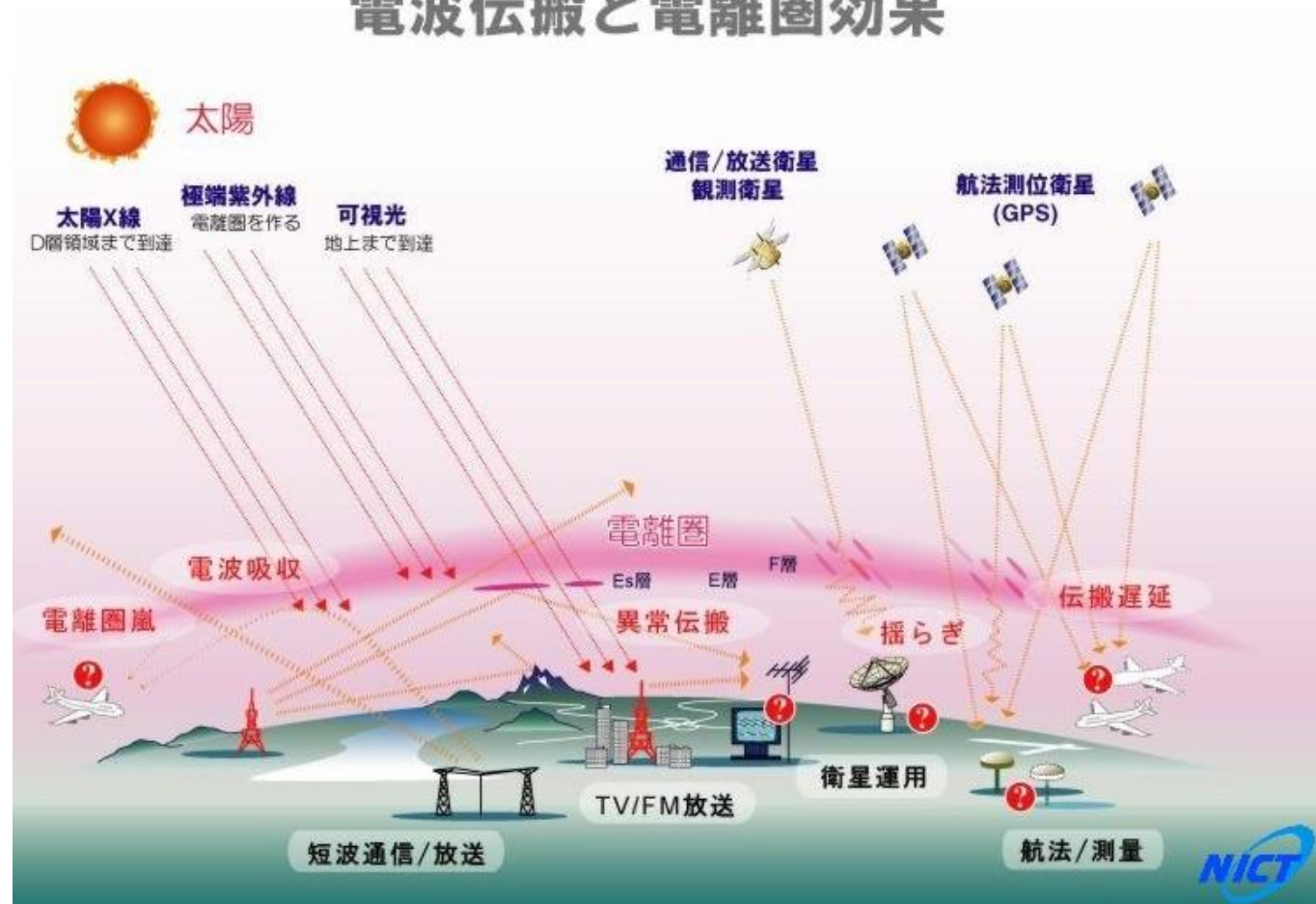
- 大分工業高等専門学校 電気電子情報工学専攻1年(学部3年相当)
- 高等専門学校は、
 - 中学校卒業後、5年間工学系専門科目を中心に学習する。
 - **研究データの取り扱いの早期教育ができる可能性あり。**
 - これに2年加えた専攻科は、学士を取得可能。
 - 大学3年に編入する人も多い。
- 大分高専は、
 - 1学年4クラス(都市、機械、電気電子、**情報**)で構成される。
 - 他高専に比べてなぜか教員の定員が少ない。

1st Topic

超高層物理学における経験モデルやシミュレーションを、どのようにメタデータ表現するか？

超高層物理学とは？

電波伝搬と電離圏効果



<http://wdc.nict.go.jp/IONO/>より引用

大分高専 小山研究室



Koyama(ex-IUGONET)



Yamauchi



Miyanari

- ・流星電波観測
 - ・観測データ公開
 - ・メタデータ生成

IUGONET

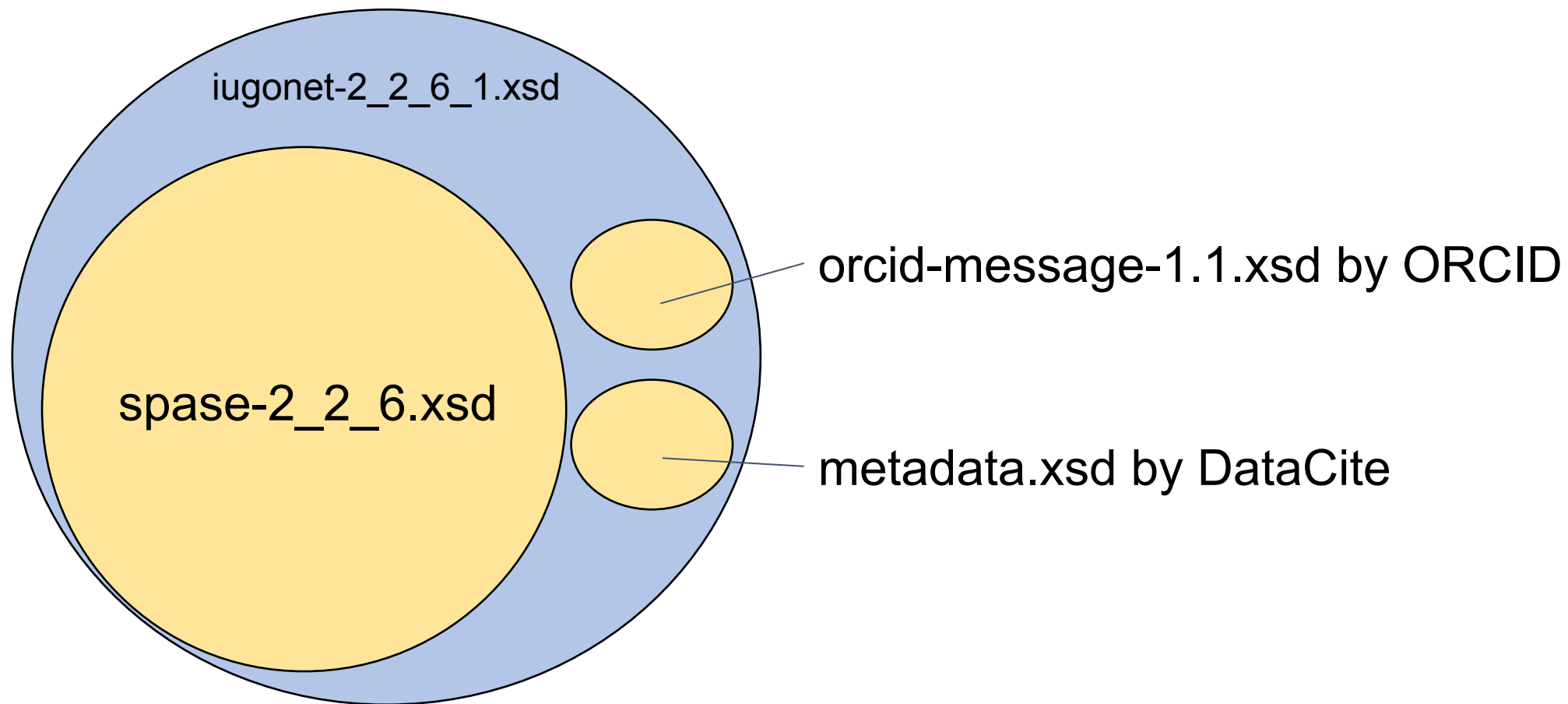


- ・メタデータ・フォーマット策定
- ・メタデータDBの運用
- ・解析ソフトウェアの開発

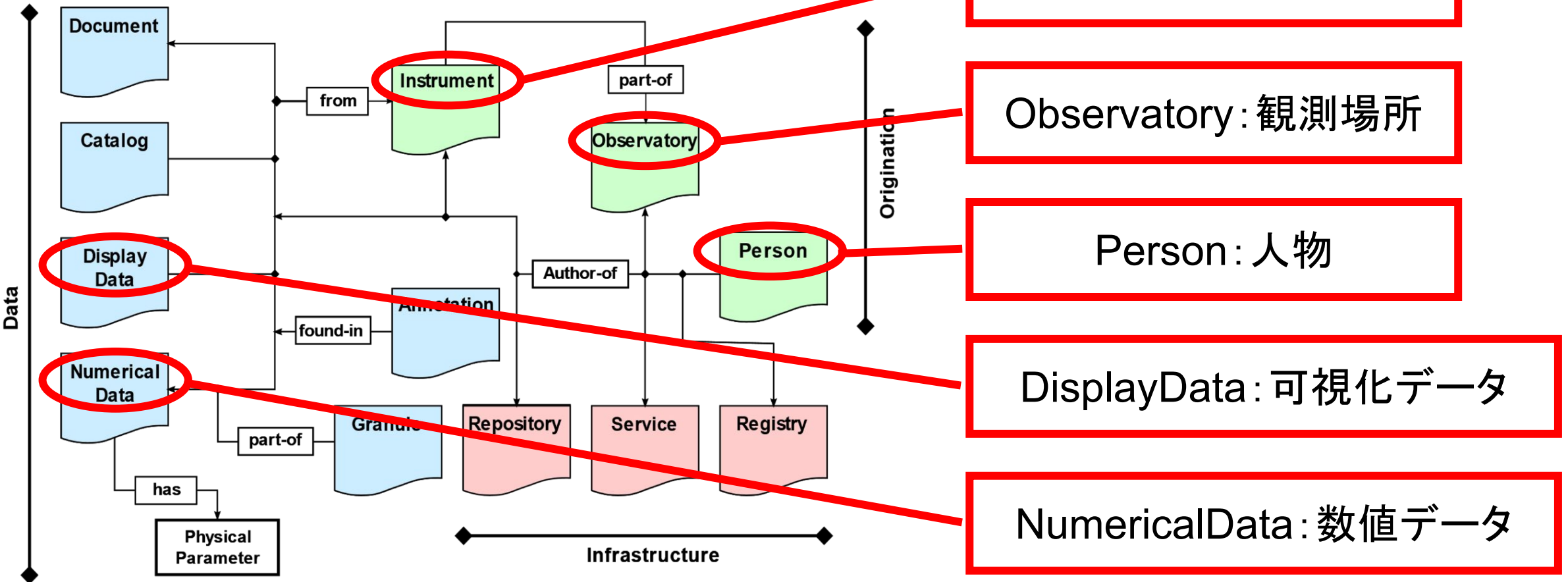
IUGONETメタデータ・スキーマ

- 米国を中心とした太陽地球系物理学の研究者らによるSPASEグループが策定した**SPASEメタデータ・フォーマット**をベースに、
 - 非デジタル保存データへの対応,
 - 太陽観測データに特有の座標系の追加,
 - 観測領域の座標等の情報の追加,
- の拡張を施し、さらには、
- ORCIDへの対応、
 - DOIへの対応、
- を行った。

SPASEとIUGONETメタデータ・スキーマの関係



SPASE Ontology



SPASEオントロジーに従って記述される観測機器, 観測場所, 観測者, 数値データ, 可視化データなどメタデータ表現における重要なリソースタイプが定義されており, メタデータ作成者はこれらを記述し, リンク関係を構築できる

IUGONETメタデータで表現できること

実験科学

(帰納的・データ駆動型)

IUGONETメタデータは観測データの説明用。

理論科学

(演繹的・モデル駆動型)

理論科学や計算科学を支える経験モデルやシミュレーションはメタデータでどのように表現するのか？

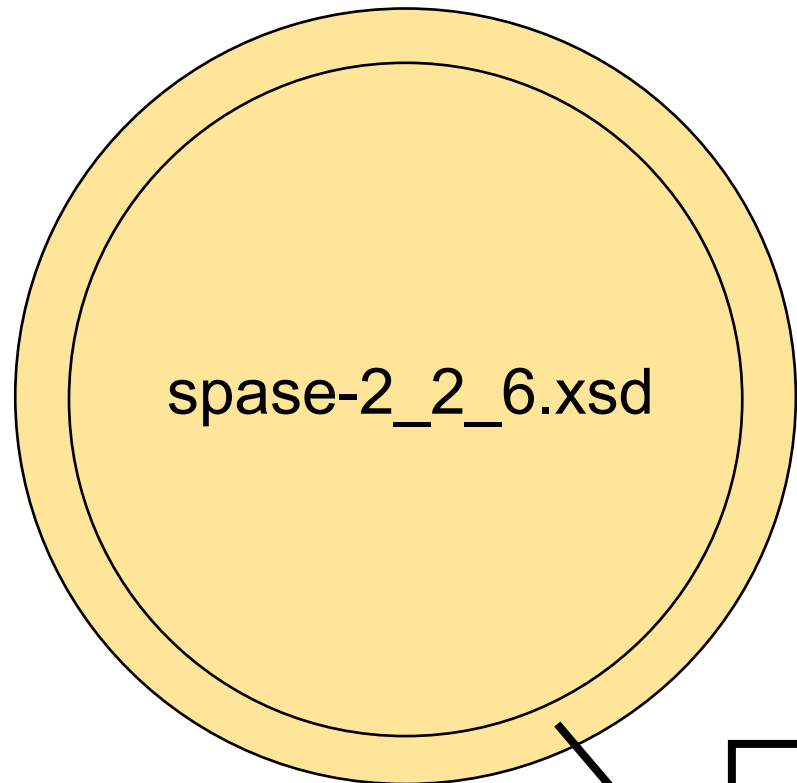
計算科学

(演繹的・モデル駆動型)

データ中心科学

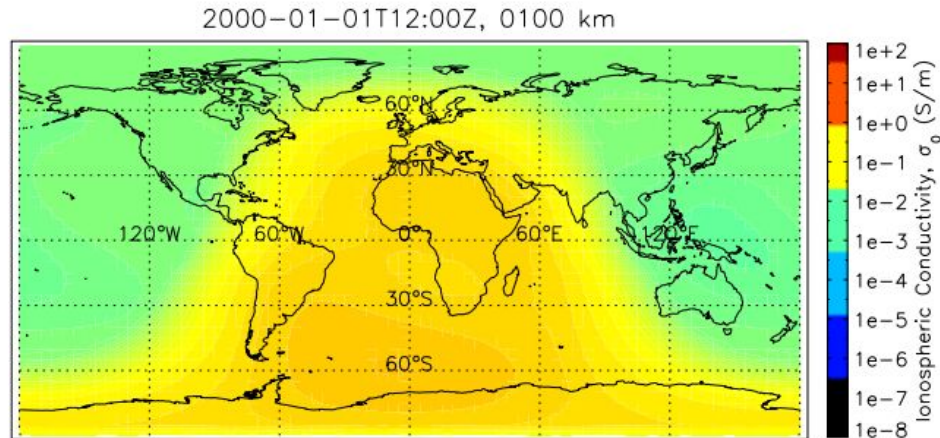
(帰納的・データ駆動)

SPASEメタデータ・スキーマのシミュレーションへの拡張



- 2017-11-30にSPASE Simulation Extensionsがリリースされた。
- IUGONETは未対応。
- 我々のグループが、SPASE Simulation Extensionsを用い、**電離圏電気伝導度モデル**のメタデータの試験的な作成事例を通して、不足要素がないか考える。

(事例) 電離圏電気伝導度モデル



- 緯度、経度、高度、日時を入力すると、3つの電離圏電気伝導度(平衡伝導度, ペダーセン伝導度, ホール電導度)を計算する経験モデル。
- 内部的には、IGRF、MSIS、IRIという3つの経験モデルをマッシュアップしている。

- ResourceID [spase://IUGONET/SimulationModel/ION_COND(TBD)]
- ResourceHeader
 - ResourceName [iug_ionospheric_cond]
 - ReleaseDate [2014-02-01T00:00:00.000]
 - Acknowledgement[]
 - Contact
 - PersonID [spase://IUGONET/Person/Yukinobu.Koyama]
 - Role [PrincipalInvestigator]
 - Description [To calculate ionospheric conductivity]
- SimulationType [Hybrid]
- CodeLanguage [IDL]
- InputProperties
 - Property
 - Name [Date], [Height], [Latitude], [Longitude]
- OutputParameters
 - Parameter
 - Name [平衡伝導度]
 - Parameter
 - Name [ペダーセン伝導度]
 - Parameter
 - Name [ホール伝導度]
- ModelURL [?]
- SimulatedRegion [earth]
- SpatialDescription
 - Dimension [3]

試作版：電離圏電気伝導度モデルのメタデータ

SPASEの用意するシミュレーションのメタデータにはモデルのライセンスを記述する場所がなかった。

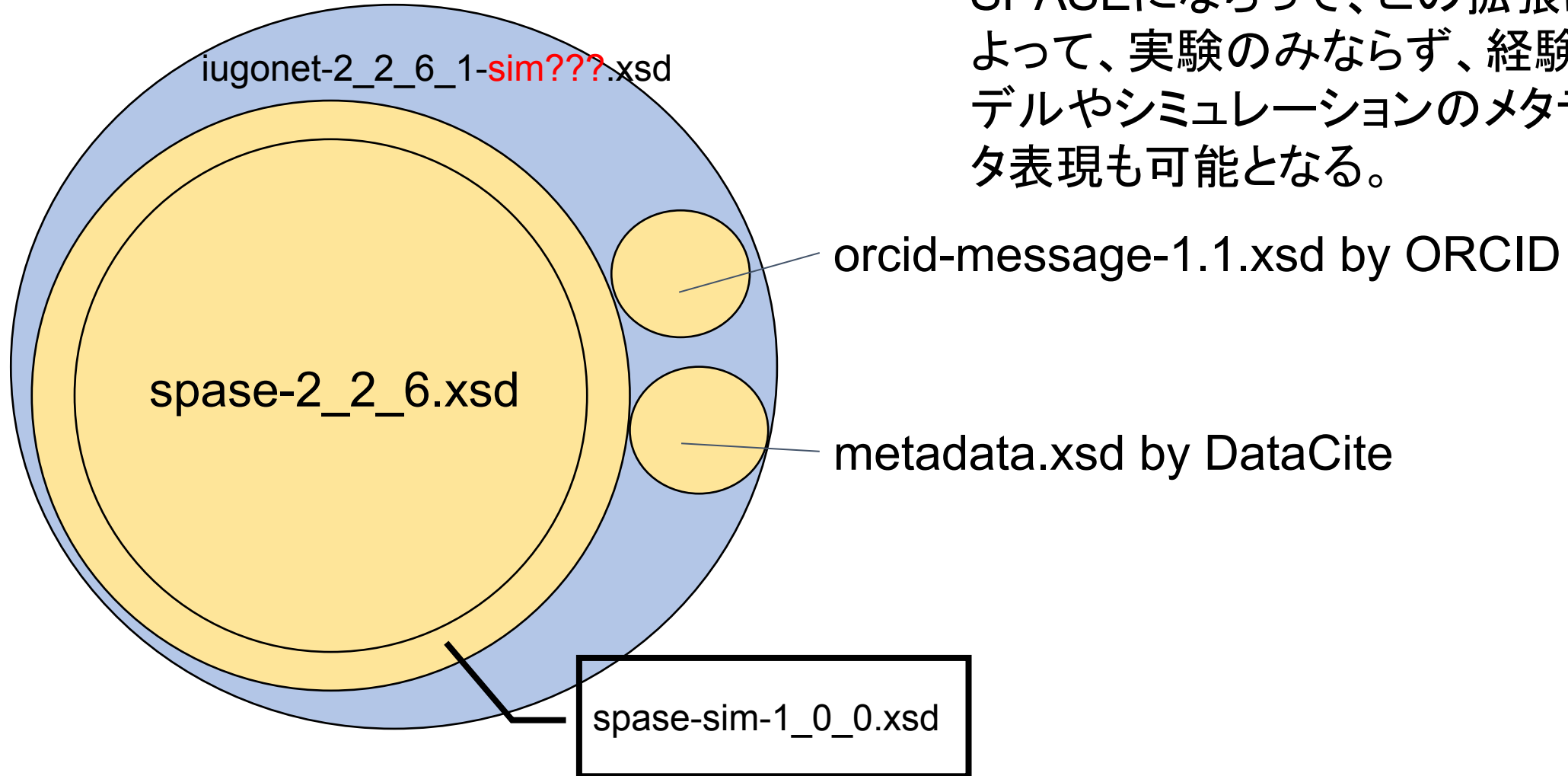
電離圏電気伝導度モデルのライセンスはStandard CPC license,
<http://cpc.cs.qub.ac.uk/licence/licence.html>.

- IUGONETはライセンス記述を取り入れている
- 「SPASE Simulation Extensions」と「IUGONETの独自拡張」のマッシュアップ(?)

```
<xsd:complexType name="AccessInformation">
  <xsd:complexContent>
    <xsd:extension base="spase:AccessInformation">
      <xsd:sequence>
        <xsd:element name="Description" type="xsd:string" maxOccurs="1" minOccurs="0" />
        <xsd:element name="DOI" type="datacite:doiType" minOccurs="0" maxOccurs="unbounded" />
        <xsd:element name="license" type="xsd:anyURI" minOccurs="0" maxOccurs="unbounded" />
      </xsd:sequence>
    </xsd:extension>
  </xsd:complexContent>
</xsd:complexType>
```

IUGONETのシミュレーション対応への提案

SPASEにならって、この拡張によって、実験のみならず、経験モデルやシミュレーションのメタデータ表現も可能となる。



2nd Topic

どのようにして小規模グループによって得られた観測データを共有するか？

(共同研究者の宮成が主のテーマ)

小規模グループにおける観測データ共有

問題点

- 高専は機関リポジトリ等のデータ公開基盤を持っていない。
 - 福井県地域共同リポジトリのように、大学、高専、図書館などのデータを保持する地域のリポジトリもない。
 - データサーバーの運用工数が捻出できない。
 - メタデータ作成などのキュレーションの工数が捻出できない。
-
- 共同利用機関や、繋がりのあるより大きな機関にデータを受け入れてもらうほどとも思わないデータである。

小規模グループにおける観測データ共有

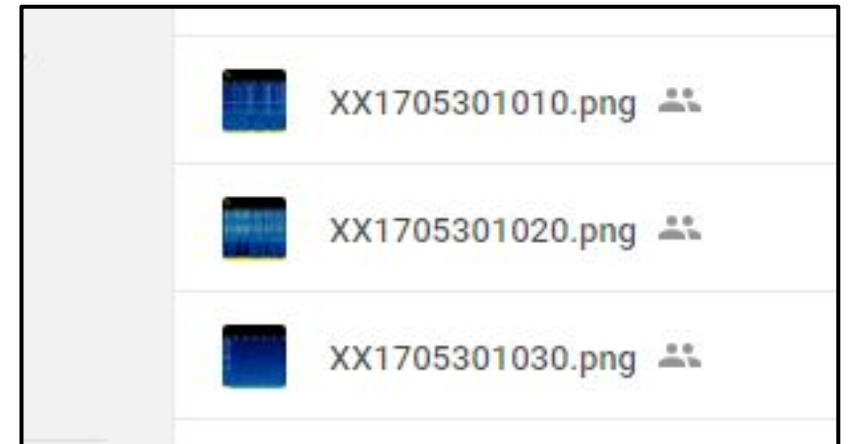
- 方法
 - Google DriveをData Repositoryとして利用し, 自ら公開する.
 - 時事刻々と生成される観測データのメタデータはGoogle Apps Scriptを用いて自動生成する.

Use Google Drive as Data Repository

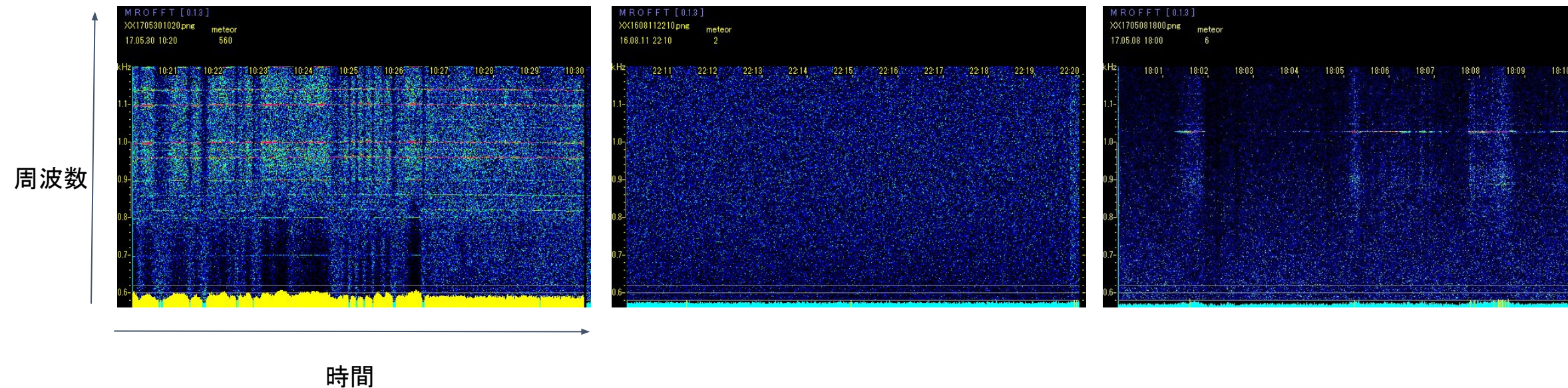


マグネチックループアンテナを用いた
流星電波観測

Synchronize observation
data with Google Drive



流星電波観測例(画像形式のスペクトル)

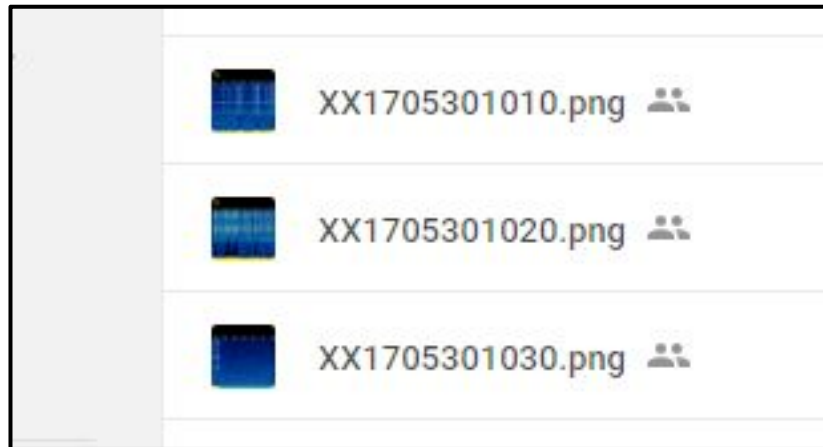


直接波観測

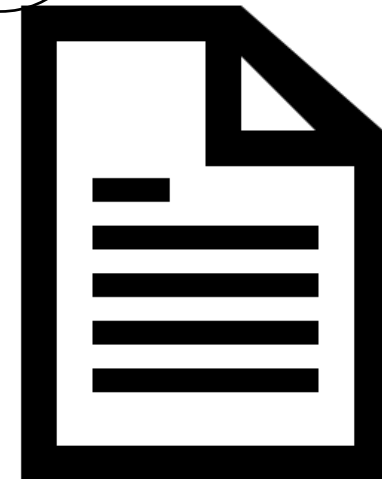
信号無し

流星による
電離圏反射波観測

To create and open to the public metadata by using Google Apps Script



Automatically retrieve the filename, creation date, URL from observational data, and create metadata by using Google Apps Script.
画像データは10分に一度生成される。メタデータ作成は、1日1回定時にトリガーをかける。



.XML

(例) 自動生成されたGranule型メタデータ

```
<Spase>  
<Granule>  
  <ResourceID>space://ONCT/Granule/Meteor</ResourceID>  
  <ReleaseDate>2017-05-30T10:3000</ReleaseDate>  
  <ParentID>space://ONCT/DisplayData/Meteor (TBD) </ParentID>  
  <StartDate>2017-05-30T10:3000</StartDate>  
  <StopDate>2017-05-30T10:3000</StopDate>  
  <Source>  
    <SourceType>TBD</SourceType>  
    <URL>URL</URL>  
  </Source>  
</Granule>  
</Spase>
```

メタデータ自動生成までの一連の流れ

- メタデータを自動生成できるようになった。
- これによってデータ共有ワークフローを簡略化できた



流星電波観測

観測データ
アップロード



メタデータ
自動生成



メタデータ

大分高専の小さなラボによるデータ共有における課題

- **観測データの質のチェックは後回し** → 観測データがハードディスクに死蔵するのを避けるため。Galaxy Zooのように、オンライン上でCitizenの力を借りて、データの質のチェックを行えば良いと考える。
- **IUGONET形式のメタデータをファイル形式で生成したのみで、まだ検索可能ではない。** → IUGONETにメタデータをHarvestingしてもらってもいいが義務感が大きい。OpenSearchの様なAPIを用意して横断検索してもらうのが楽！？

まとめ

科研費の細目の、理工系 - 数物系化学 - 地球惑星科学 - 超高層物理学におけるメタデータを介したデータ共有について事例を挙げて考えた。

- 日本の超高層物理学分野であるIUGONETにおいて、経験モデルおよびシミュレーションのためのメタデータは、SPASEのSimulation Extensionsを使えば良い。
- 小規模な研究室における観測データの共有ワークフローの事例を示した。

補足

| | FileName | | | | | |
|---|------------------|---------------------|----------|------------------|---|---|
| | A | B | C | D | E | F |
| 1 | FileName | LastUpdate | DataSize | Date | | |
| 2 | XX1607211336.png | 2017/10/05 16:52:19 | 3371 | 2016/07/21/13:36 | | |
| 3 | XX1610131350.png | 2017/10/05 16:52:17 | 94487 | 2016/10/13/13:50 | | |
| 4 | XX1705301030.png | 2017/10/05 16:52:14 | 122225 | 2017/05/30/10:30 | | |
| 5 | XX1705301020.png | 2017/10/05 16:52:11 | 129532 | 2017/05/30/10:20 | | |
| 6 | XX1705301010.png | 2017/10/05 16:52:08 | 126321 | 2017/05/30/10:10 | | |
| 7 | | | | | | |
| 8 | | | | | | |
| 9 | | | | | | |

新規

マイドライブ > HROFFT > Backup

- マイドライブ
- パソコン
- 共有アイテム
- 最近使用したアイテム
- Google フォト
- スター付き
- ゴミ箱

12.5 MB/15 GB を使用中

容量をアップグレード

| 名前 ↑ | オーナー | 最終更新 |
|--|------|------------|
|  XX1607211336.png | 自分 | 2017/10/06 |
|  XX1610131350.png | 自分 | 2017/10/06 |
|  XX1705301010.png | 自分 | 2017/10/06 |
|  XX1705301020.png | 自分 | 2017/10/06 |
|  XX1705301030.png | 自分 | 2017/10/06 |

組織の識別子

現段階ではSPASEオントロジー内で固有の組織IDが割り振られており、超高層物理分野においては問題ない

他分野との連携を考えた場合も、メタデータにおいては名前空間があるので識別はおそらく可能

でも...！！

名前空間に依らない国際識別子を利用してもいいのでは？

メリット:

名前空間を考慮する必要がなく判別が容易

関連するシステム構築の際、設計が容易になる可能性

ISNI : International Standard Name Identifier

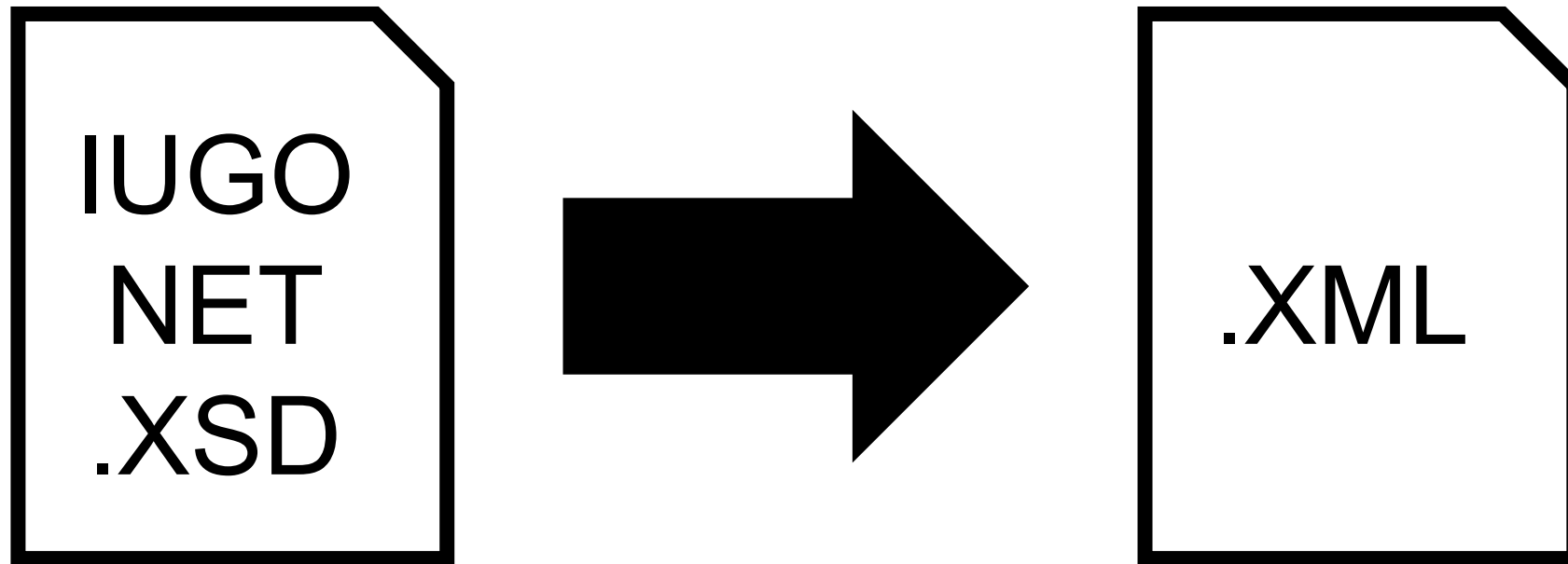
知的創作物やコンテンツに関わる個人及び組織に付与される国際的かつ分野横断的な識別子^{*1}

しかし, ISNIが付与された機関は少ない?

*1 E17773 – 識別子の架け橋: 国際標準名称識別子ISNI - <http://current.ndl.go.jp/e1773>

IUGONETメタデータ・スキーマ

- IUGONET XSD(XML Schema Definition)に則り, XML形式のメタデータを作成する.



IUGONETが表現するデータ

実験科学

(帰納的・データ駆動型)

数千年前: 観測をもとに自然現象を解明する科学
経験記述的な科学手法, 天動説

理論科学

(演繹的・モデル駆動型)

数百年前: 解明した現象から導かれる理論の科学
数式を用いた解析的な科学手法

計算科学

(演繹的・モデル駆動型)

数十年前: シミュレーションで予測・再現する科学
HPCを用いた自然現象の計算, 可視化

データ中心科学

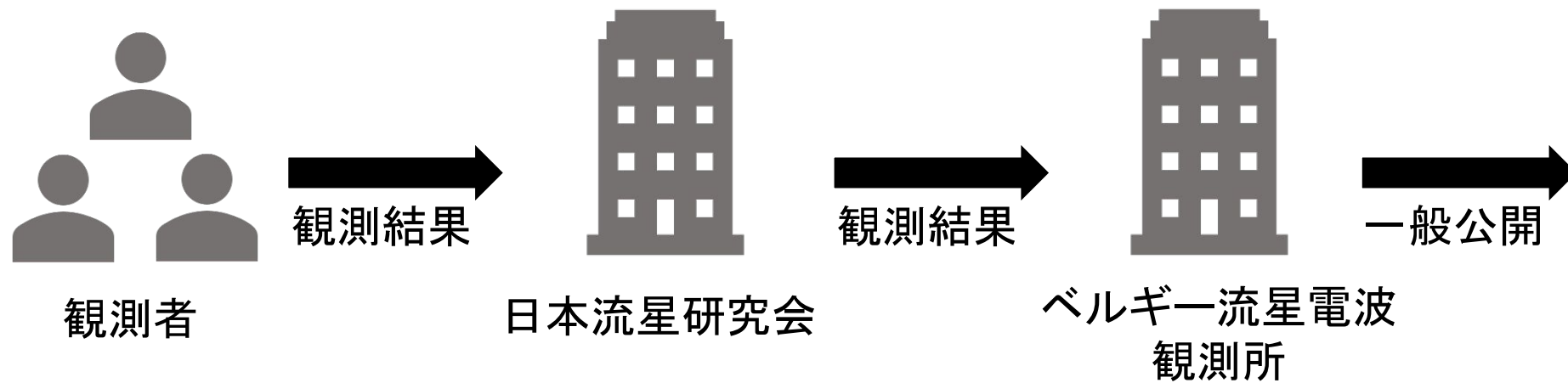
(帰納的・データ駆動)

現在 : 大量のデータを解析することにより浮か
上がる理論を追求, ビッグデータ解析

び

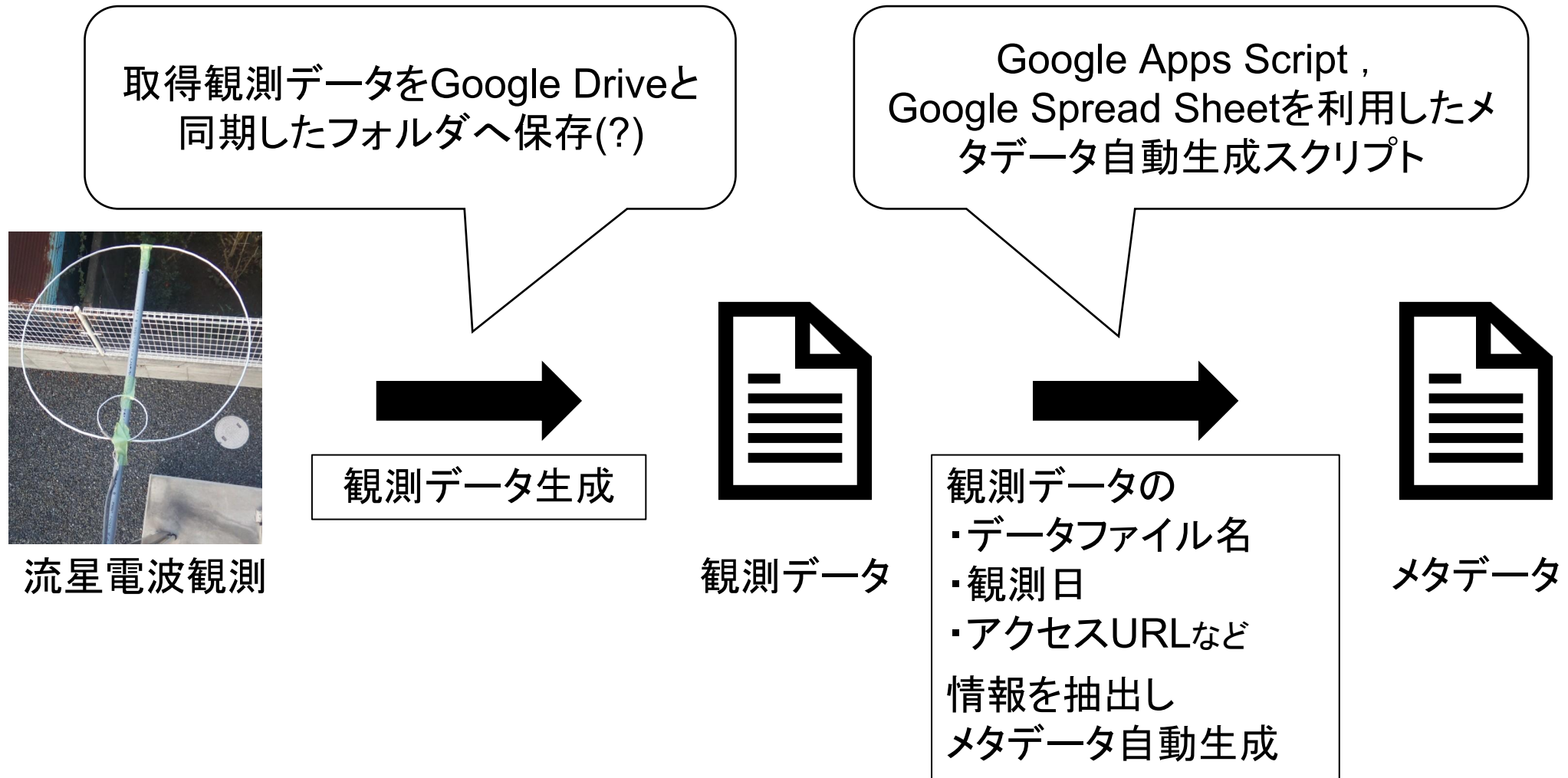
小規模グループにおける観測データ共有

現在主流な流星電波観測における観測データ共有ワークフロー



大分高専 小山研究室のメタデータ自動生成への取り組み

宮成が確立した観測データのメタデータ自動生成の一連の流れ



1. 超高層物理学におけるデータ共有



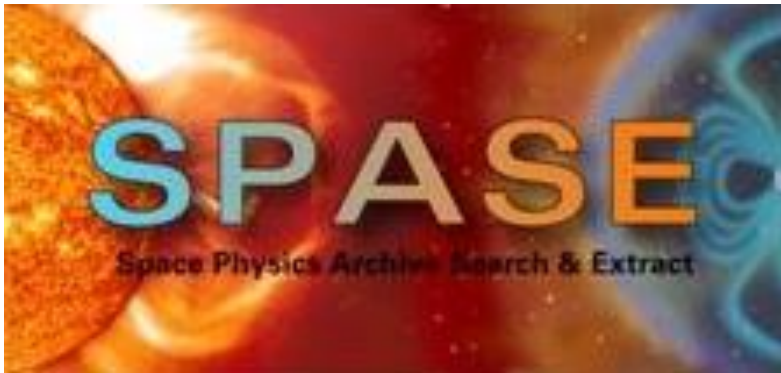
日本: IUGONET

Inter-university Upper atmosphere Geophysical
Observatory NETwork

- ・大学間連携プロジェクト

米国: SPASE

- ・太陽地球系物理学の研究者らで構成されるグループ



- ・観測データのためのメタデータフォーマットを策定

- ・観測データのメタデータ収集

- ・データベース構築

- ・データ解析ソフトウェアの開発

なぜメタデータなのか



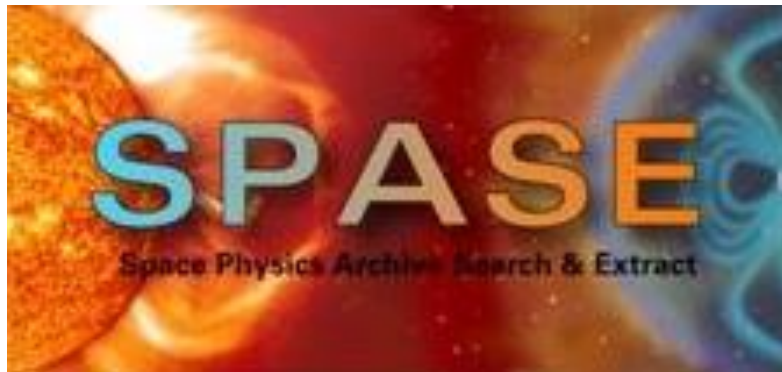
日本: IUGONET

- ・大学間連携プロジェクト

米国: SPASE

- ・太陽地球系物理学の研究者らで構成されるグループ

- ・観測データのためのメタデータフォーマットを策定
- ・観測データのメタデータ収集
- ・データベース構築
- ・データ解析ソフトウェアの開発



京大・理・地磁気センター

名大・STE研

観測データDB
データ形式: IAGA

観測データDB

NASA

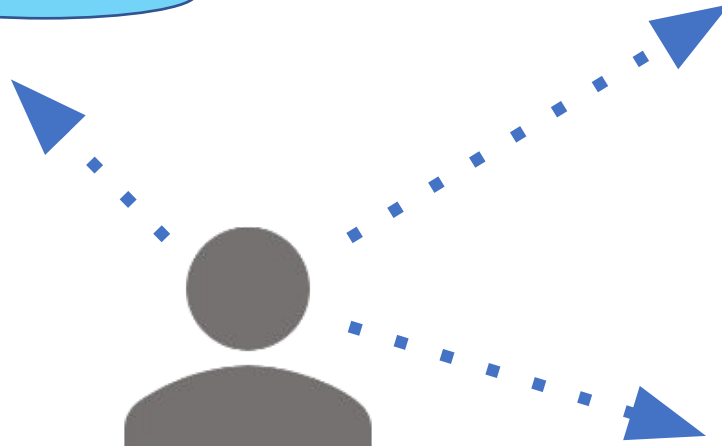
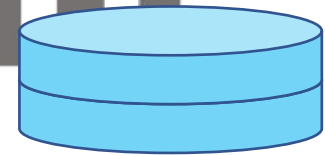
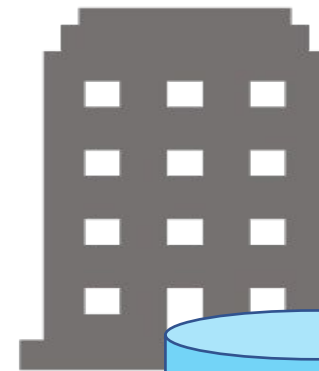
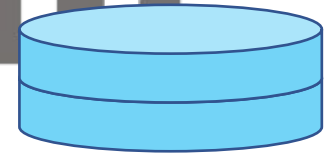
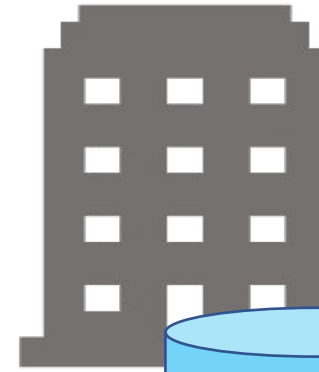
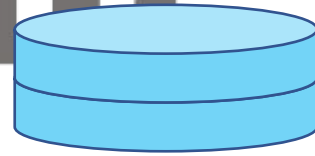
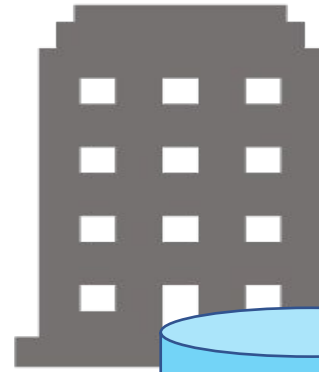
観測データDB
データ形式: FITS

データ利用者

欲しいデータはどこ?

観測データは分散管理されており、
一つにまとめるのは困難

分散管理ゆえ、目的とするデータ
への到達が困難



メタデータ: データについての情報を記述したデータ

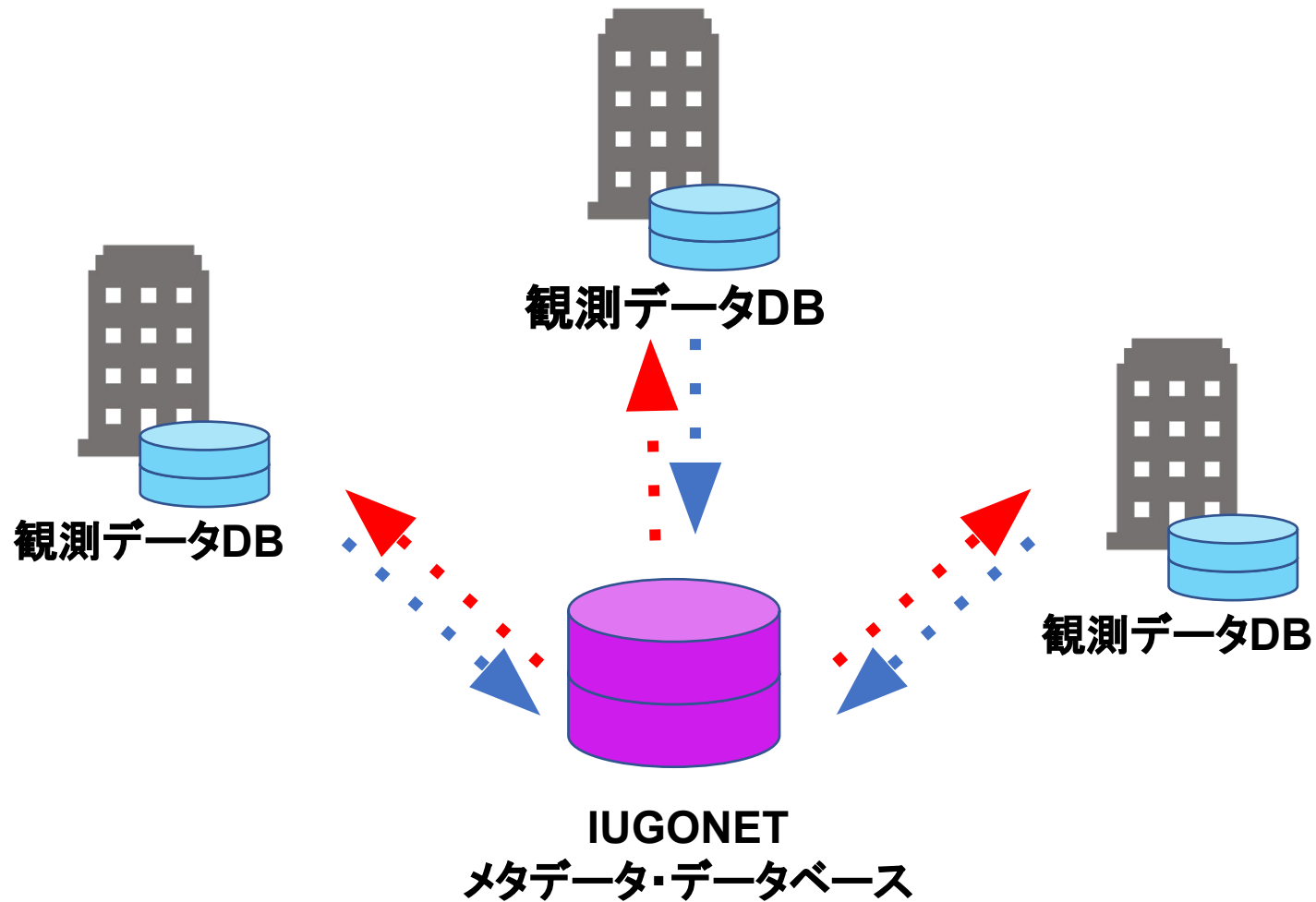
**観測データの
メタデータ**

**データの所在(URL)
観測機器
観測場所
利用規定
など**

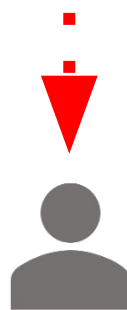
メタデータDBの構築によって

主要項目ごとにアーカイブが可能
メタデータ要素を基にした検索が可能

データへの到達がより容易となる



- ▶: メタデータ取得
- ▶: メタデータ登録

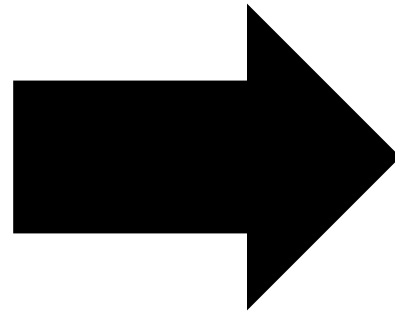
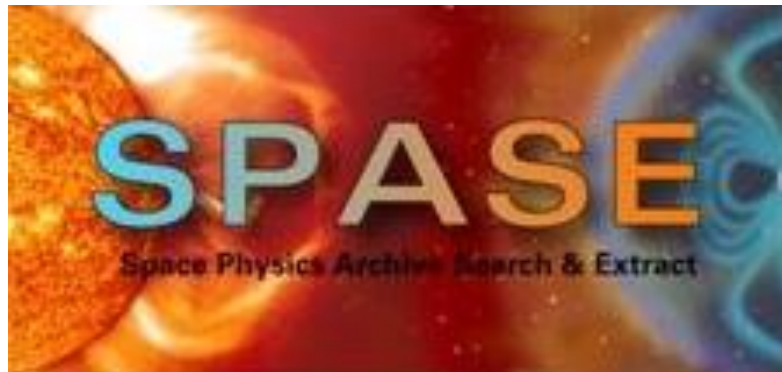


メタデータに基づく検索機能を使い、観測データの保管場所(URL)を把握できる。

2. IUGONETメタデータ・スキーマ

メタデータ・スキーマ...メタデータの構造(作り方)を定義

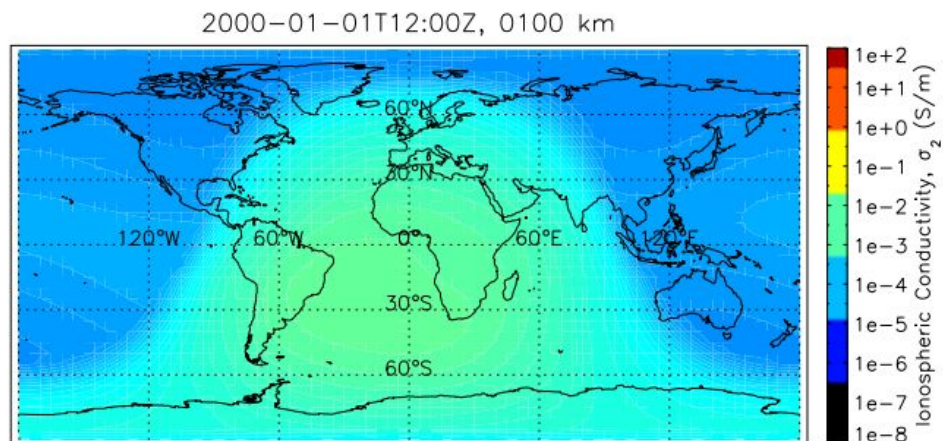
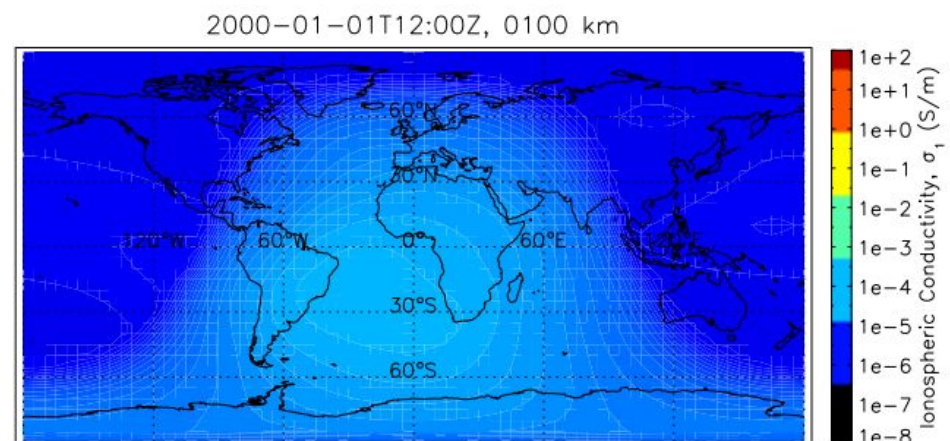
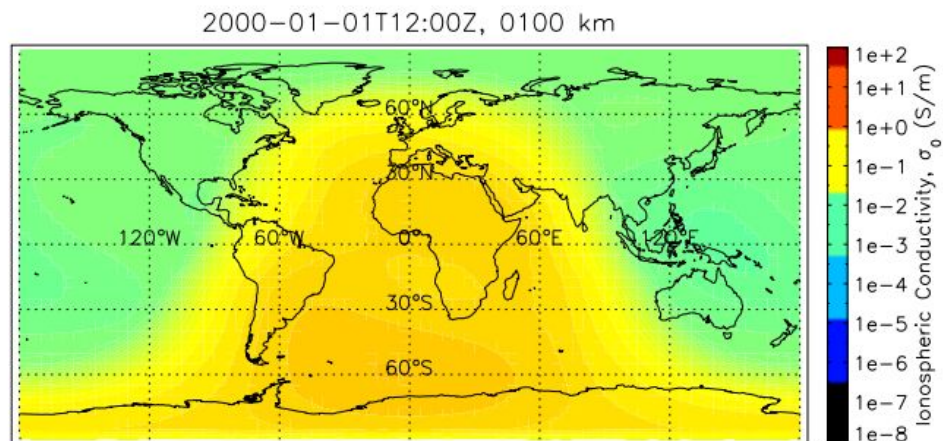
超高層物理分野と親和性が高い太陽地球系物理学を対象とした
SPASEメタデータ・スキーマを拡張して公開している



超高層物理向けに拡張



(事例) 電離圏電気伝導度モデル



日時を指定すると、
全地球上の電離圏電気伝導度
(平衡伝導度, ペダーセン伝導
度, ホール電導度)を出力

電離圏電気伝導度モデル

NumericalOutput

- SimulatedInstrument -
- MeasurementType -
- SpatialDescription -
- SpectralRange -
- SimulatedRegion -
- InputResourceID -
- Parameter -
- SimulationProduct -
- Property -
- Extension -

AccessInformation :
InputResourceID :
MeasurementType :
Parameter
 Description :
 Name :
 ParameterKey :
 Particle :
 Units :
ResourceHeader :
ResourceID :
SimulatedRegion :
SpatialDescription :

電離圏電気伝導度モデル

DisplayOutput

- SimulatedInstrument -
- MeasurementType -
- SpatialDescription -
- SpectralRange -
- DisplayCadence -
- SimulatedRegion -
- InputResourceID -
- Parameter -
- SimulationProduct -
- Property -
- Extension -

| 記号 | 定数, 変数の説明 |
|----|----------------|
| | 平衡伝導度 |
| | ペダーセン電導度 |
| | ホール伝導度 |
| | 電荷 |
| | 電子密度 |
| | 電子の質量 |
| | イオンの質量 |
| | 電子の衝突周波数 |
| | イオンの衝突周波数 |
| | 電子-中性粒子の衝突周波数 |
| | 電子-イオンの衝突周波数 |
| | イオン-中性粒子の衝突周波数 |
| | 中性粒子温度 |
| | イオン温度 |
| | 電子温度 |
| | 電子の旋回角周波数 |

小規模グループにおける観測データ共有

問題点

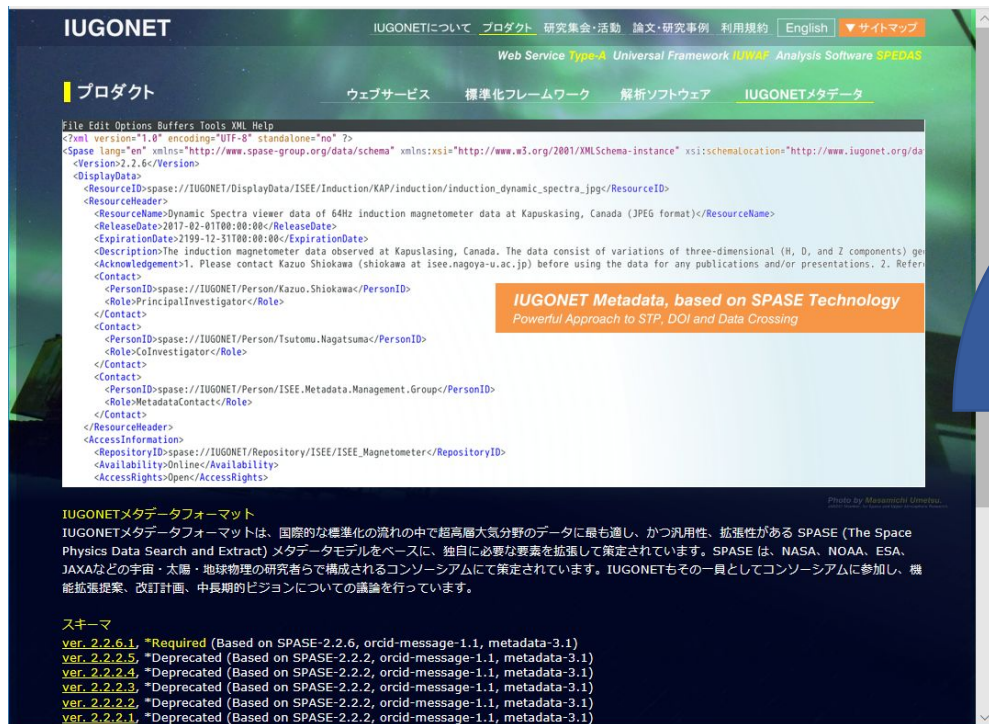
- データ・サーバーの運用工数が捻出できない。
- データのメタデータ作成などのキュレーションの工数が捻出できない。

現在主流な流星電波観測における観測データ共有ワークフロー



IUGONETメタデータ・スキーマ

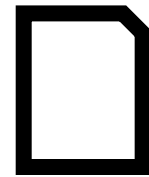
IUGONETのウェブサイト公開されており、
メタデータの作成に利用される。



The screenshot shows the IUGONET website interface. At the top, there are navigation links for 'IUGONETについて', 'プロダクト', '研究集会・活動', '論文・研究事例', '利用規約', and 'English'. Below this, there's a 'プロダクト' (Products) section with a sub-menu for 'IUGONETメタデータ'. The main content area displays XML metadata for a resource. The XML includes fields like <ResourceName>, <ReleaseDate>, <ExpirationDate>, <Description>, <Acknowledgement>, <Contact>, <RepositoryID>, <Availability>, and <AccessRights>. A blue box highlights the text 'IUGONET Metadata, based on SPASE Technology' and 'Powerful Approach to STP, DOI and Data Crossing'. Below the XML, there's a section for 'IUGONETメタデータフォーマット' (IUGONET Metadata Format) and a 'スキーマ' (Schema) section listing various versions of the metadata format, such as 'ver. 2.2.6.1, *Required (Based on SPASE-2.2.6, orcid-message-1.1, metadata-3.1)'.

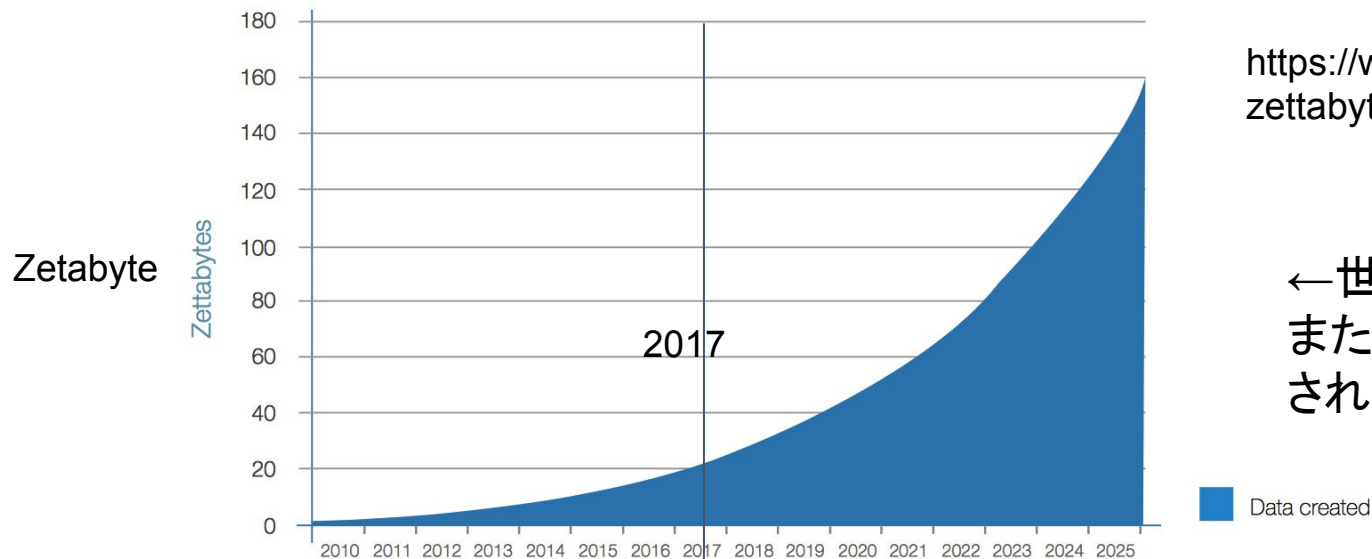
Download
Schema

Create
Metadata



Summary of the 2nd Topic

- “Small laboratory” is just metaphor(暗喩).



<https://www.storagenewsletter.com/2017/04/05/total-ww-data-to-reach-163-zettabytes-by-2025-idc/>

←世界のストレージ容量。科学データもまた、時を経て肥大化すると容易に想像される。

- データの洪水をさばきされるための設備とデータキュレーターを確保できない機関のたどる道は一緒！？