

学術データのアーカイブとメタデータ整備：名古屋大学の取り組みから

東海国立大学機構

名古屋大学 宇宙地球環境研究所

統合データサイエンスセンター：三好、能勢、梅田、増田、堀、新堀

電磁気圏研究部：大塚

国際連携センター：塩川

名古屋大学/岐阜大学附属図書館

端場、直江、石田、鈴木（現：神戸大学）

名古屋大学情報戦略室

松原、青木

名古屋大学情報基盤センター

高橋、山田、林

名古屋大学宇宙地球環境研究所（共同利用・共同研究拠点）でのデータアーカイブ・公開

[科学衛星や地上観測のデータアーカイブ・公開]

- ・科学衛星：ERGサイエンスセンター、ひのでサイエンスセンター（JAXA/国立天文台と共同）
- ・研究所各プロジェクトのデータアーカイブ・公開
（年間ファイルダウンロード：数百万、公開データ：数PB）



Hinode Science Center at Nagoya

[WDC]

- ・宇宙線に関する世界データセンター

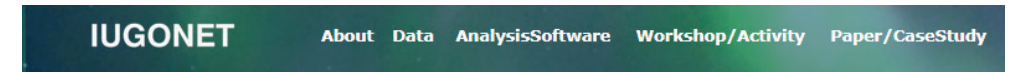


[データ出版・引用の推進]

- ・研究所独自のプレフィックス付きDOIを科学データに付与し、データアーカイブとリンクしたLanding Pageを整備
 - ・米国地球物理学連合の学術誌等では、使用データのデータリポジトリやDOIを明記しないと投稿に進めない。
 - ・データの整備や公開、維持にあたっている取り組みがきちんと評価されるように

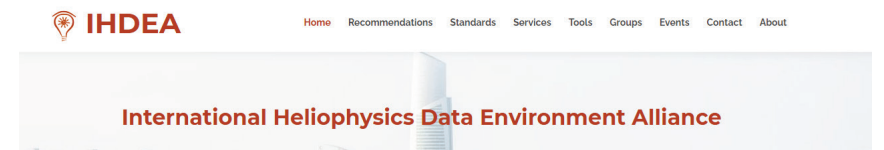
[コミュニティプロジェクトやコンソーシアムへの参加]

- ・IUGONET(超高層分野の大学間連携)
- ・IHEDA (International Heliophysics Data Environment Alliance)へ JAXAとともに理事機関として参加



[共同利用]

- ・データベース共同利用、計算機共同利用を実施



大学・研究機関における学術データの取り扱い

今後の大学・研究機関の学術データに求められること

- ・適切なメタデータを付与した上での安定したアーカイブ（メタデータ、長期アーカイブ）
- ・大学学術資産として、学術コミュニティ、一般社会への利用の拡大（データ検索性、アクセス）

大学リポジトリ、DOI付与、データアーカイブへの期待

現状

大学には様々な研究分野があり、多様なデータ、データファイルサイズの規模感も多様

- ・ユースケースに即した形での課題抽出の必要性
- ・大規模アーカイブ装置について、利用目的とコストパフォーマンスの観点からの評価の必要性



課題の抽出や、モデルスキーム構築のために、実際の研究現場におけるパイロット的な実験が必要

名古屋大学における取組

目的・必要性

目的：

宇宙地球環境データを例にして、データの標準化、メタデータの整備、大規模データアーカイブの運用等、**デジタルユニバーシティ化に向けた基礎技術を全学のパイロット事業として実施。**
宇宙地球環境研究所の共同利用（データベース共同利用）のスキームへ反映。

必要性：

研究活動におけるデータ管理、整備をとりまく環境は急速に変化。
大学におけるユースケース例を抽出し、研究活動を円滑に進めるためのスキームを確立。

取り組みの概要

宇宙地球環境研究所が収集している宇宙地球環境データを例にして、

- 1) データの標準化、DOIの付与、メタデータの整備等のデータマネジメントの実証実験
- 2) 学術データアーカイブへのコールドストレージ利用の実証実験

事業内容



2. 学術データアーカイブとしての
大規模コールドストレージ
活用実証実験

学術データアーカイブ実証実験

実施内容

- ・学術データのアーカイブ先としての光ディスクの活用

獲得知見

- ・光ディスクへのデータ保管スキーム、問題点、CP評価

学術データ整備実証実験

実施内容

- ・IUGONETと連携したメタデータ整備事業 (IUGONETメタデータスキーム)
- ・名大ISEEにおける研究データへのDOI付与

獲得知見

- ・学術データのアーカイブに必要なメタデータの把握
- ・名古屋大学におけるメタデータ登録の作業スキーム

事業内容：コールドストレージへのデータアーカイブ実験

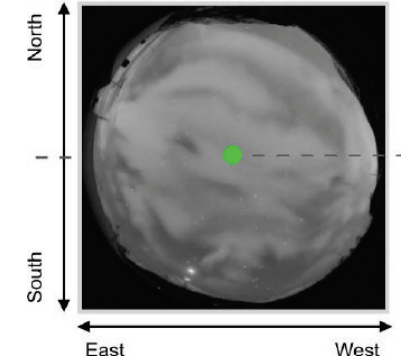
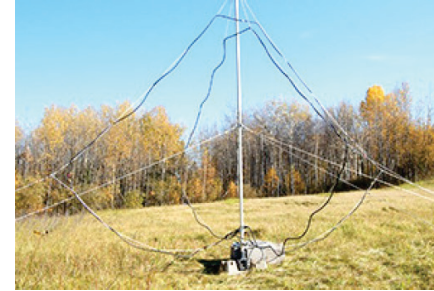
学術データアーカイブ実証実験

■ 宇宙地球環境研究所

- 宇宙電波の観測データ、オーロライメージ観測データを対象に実施
1Pbyte以上のデータを蓄積

- 記録容量：5Tbyte
- 堅牢性が高い、安全性

学術データの長期保管向き



■ 情報基盤センター コールドストレージシステム

- キャビネット内の光ディスクを
ドライブに搬出入するためのロボットシステム
- フロントエンド経由で学内LANに接続
- 本実験で2Pbyte使用

利便性が高いアーカイブシステム



- ・アーカイブ作業に関する課題の抽出
- ・将来的な全学学術データアーカイブシステムとしての適用可能性の検証

事業内容：コールドストレージへのデータアーカイブ実験

学術データの長期アーカイブに対する光ディスクのメリット/デメリット

ハードディスクアレイ	光ディスクアーカイブ
2～3年が寿命（回転系の摩耗、磁気の劣化）	100年以上の寿命、水や熱にも強い
5年程度でシステム更新が必要 2021年4月において1PBあたり1000-2000万円	光ディスクの追加・ドライブの更新のみ 本申請では、1PBあたり400万円 （維持費：年間 20万円）
高速なI/O、データ公開サーバに直結	ディスクの入れ替え、低速なI/O

[抽出された課題]

- ・大学において、学術データの保管・管理が求められる状況において、相対的に安価な光ディスクのメリットは大きい。
- ・ユースケースとして、観測生データのような利用頻度は少ないが恒久アーカイブが必要とされるデータの保管には適するが、利用頻度が高いデータの保管には適していない面がある。
大学におけるユースケースの事例を集めて、適切な活用例のガイドラインを作る。

事業内容：メタデータ整備/DOI登録

[メタデータ登録：IUGONETと連携]

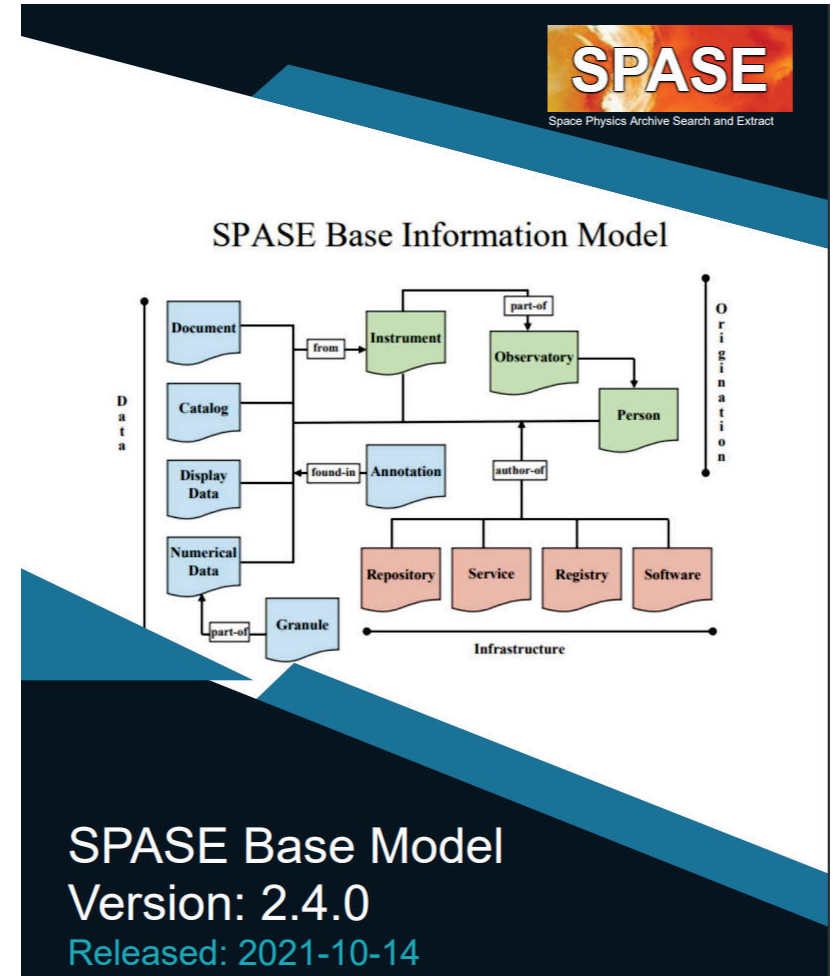
- これまでのメタデータは、SPASE2.2.6をベースにして、DOI, ORCIDなどを扱える独自拡張を含めたメタデータスキーマに従って作成してきた。
- 現在は、SPASE2.4.0であり、モデル値のメタデータも記述できるので、メタデータスキーマのアップデートとそれに応じたメタデータへの書き換えを実施。
- 昨年度、1000件以上のデータを登録。

[DOI登録]

- ISEEのDOIプレフィックス(10.34515)を取得し独自プレフィックス付きのDOI付与を可能に。
- あらせ衛星データ、地上観測データ、フレアカタログ、C14年輪データ等にDOIを付与
- 今後、研究所および共同利用データへのDOI付与を継続
- 科学データ出版・引用を推進

[抽出された課題]

分野（本事業では宇宙科学分野）のメタデータと、今後学術機関としての整備が必要なメタデータとの対応（マッピング）をどのように進めていくか。名古屋大学では、今年度、スキーマやマッピングの調査・検討予定。



SPASE Base Model

Version: 2.4.0

Released: 2021-10-14

まとめ

- 名古屋大学では部局連携で、宇宙地球環境研究所のデータを例に
 - 学術データアーカイブ（光ディスク）の実証実験
 - 学術データ整備（メタデータ、DOI）の実証実験を実施中。
- 実験による知見をもとに、それぞれの実証実験から課題を抽出
- 特に 研究分野独自のメタデータと、学術機関リポジトリのメタデータとのマッピングについては、研究機関や分野を超えた共通の課題であると思われるので、より広い範囲での知見の共有や議論の機会を得たい。