多様な観測データ利用活性化を実現する解析環境の一提案

笠原 禎也, 馬渕 嵩大, 宮崎 大夢, 大池 悠太, 松田 昇也, 高田 良宏, 松平 拓也, 後藤 由貴 (金沢大学)

金沢大におけるプラズマ波動データ蓄積状況

あけぼの (1989.2~)

- デジタルデータ: 約2Tbyte (うち約1/3が電波計測データ)
- アナログデータ: DATテープ約28,000本 (デジタル化後のデータ総量は 約30Tbyte)

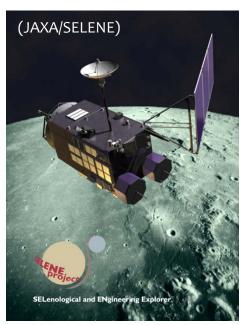
かぐや (2007.12 ~ 2009.9)

■ データ伝送量

全量:10GB/day

(うち LRS/WFC:1GB/day)





一昔前ならば「膨大」なデータではあるが・・・

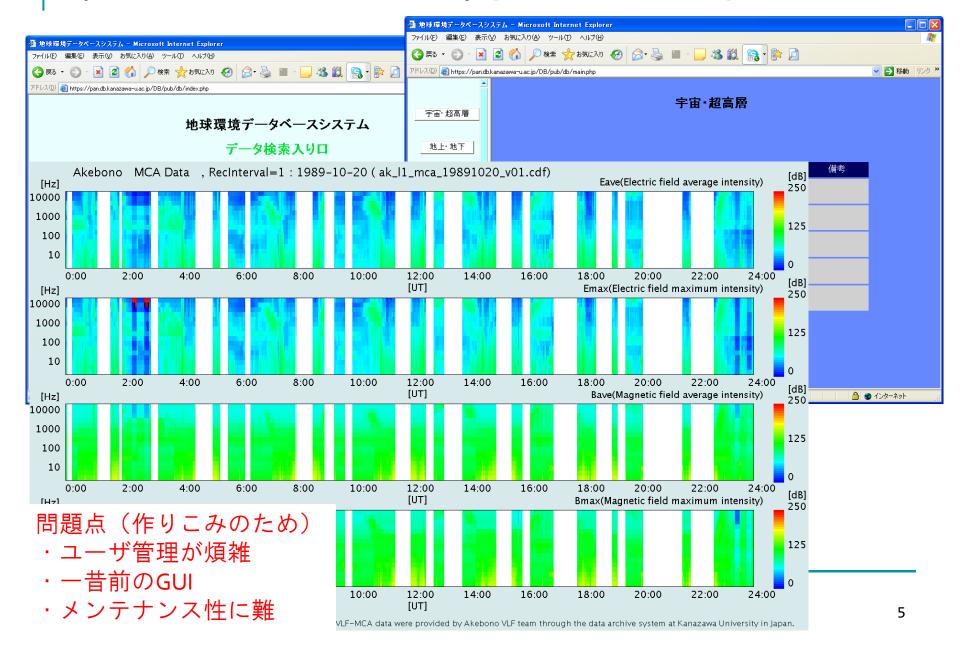
何が必要?これからすべきことは?

- A) 大容量データのDB化と公開システムの整備
 - ■研究者に負担がかからないデータ生成・蓄積機構の構築
 - 利用しやすいデータフォーマット・解析ツールの提供
 - 適切なデータ公開/利用ポリシーに基づくデータ閲覧機構
 - 適切なデータのバージョン管理と更新

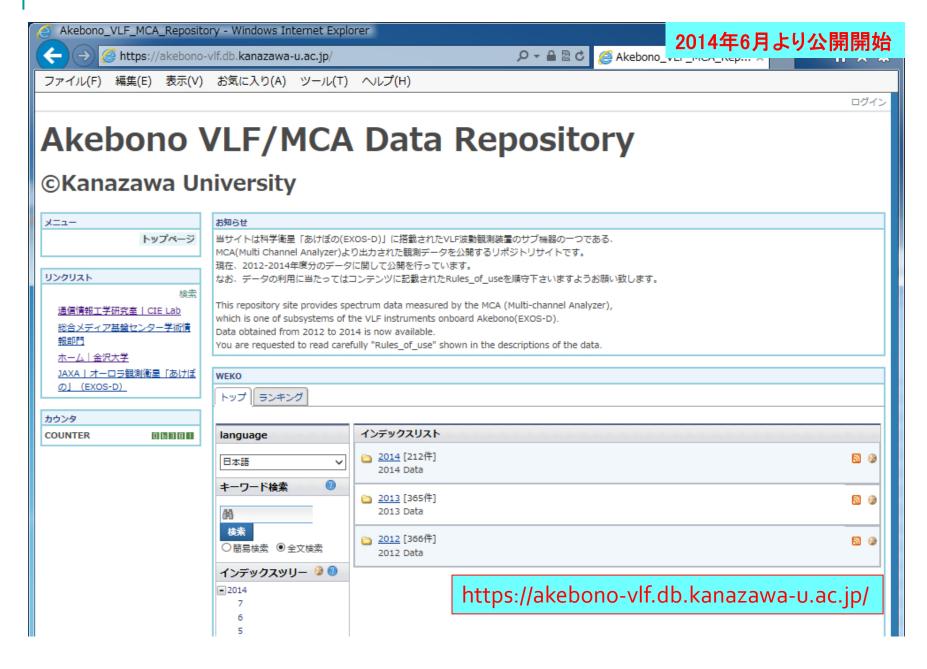
あけぼのVLFデータの現況

- MCA (f<17.8kHz) Spectrum(E x 1, B x 1), CDF化済 (Δt=8s, freq. 16pts.), データ公開済
- ► ELF (f < 80Hz)</p>
 Spectrum(Ex1, Bx3), CDF化済 (Δt=8s, Δf=2.5Hz), 公開準備中
- ▶ PFX (for k-vector analysis, Δf = 50Hz, 100Hz<f<12.75kHz) Waveform (Ex2, Bx3), CDF化未了(一部試作済・評価中) 並行してデータ検索 & 解析ツールの整備が必要</p>
- WBA (analogue waveform, f<15kHz) Waveform (Ex1 or Bx1), アナログ⇒デジタル変換処理中(~2015年度末に完了予定) (~30TB) 較正処理の自動化に難あり(かなり難しい) 網羅的・系統的解析が難しいため、解析可能なイベントリストの公開を優先?

あけぼのVLFデータ公開システム(旧版: Ver.1)



新・あけぼのVLFデータ公開システム(Ver.3)



新システムの特徴

- NII開発のWEKOベースのデータリポジトリを採用
 - □ 今後のバージョンアップ、メンテナンス性の向上
- メタデータ ⇒ CDFのG属性を元に生成
 - 内部ではWEKO基本属性+ユーザ定義属性 (WEKO基本属性…タイトル,キーワード, 言語などの属性)
 - Dublin Coreやjunii2対応のメタデータ形式 等の出力が可能
- WEKOのミドルウェアであるNetCommons2 によるアカウント認証機能を利用可
- SWORDプロトコルによる一括データ登録
 - Sword Client for WEKOを活用
- Web解析用ソフトとの連携
 - 🗅 Awstats, Piwikを用いたアクセス解析 etc.
- データのサムネイルの表示



今後の改善点

- 各種メタデータDBとの連携に向けた改善
 - □ コンテンツ(ファイル)指定がQuery形式

https://akebono-vlf.db.kanazawa-u.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=1671&item_no=1&page_id=13&block_id=9

⇒ コンテンツへのパーマリンク機能を追加実装

https://akebono-vlf.db.kanazawa-u.ac.jp/mca/20140701/

仮実装状態で運用中のため、今後、機能改善・ 拡張を計画中 コンテンツ識別用の item_id 等が ランダムに割り振られてしまう!

> 意味のあるパラメータ指定で コンテンツにアクセス可能

- □メタデータ登録内容の妥当性・互換性検証
- □コンテンツ種の追加・登録
- □ 認証機能付きコンテンツとの共存方式の検討

何が必要?これからすべきことは?

- A) 大容量データのDB化と公開システムの整備
 - 研究者に負担がかからないデータ生成・蓄積機構の構築
 - 利用しやすいデータフォーマット・解析ツールの提供
 - 適切なデータ公開/利用ポリシーに基づくデータ閲覧機構
 - 適切なデータのバージョン管理と更新
- B) 所望データの条件に合致したデータ検索機能の充実
 - → 刻々と変化する観測条件・観測モードから所望データをどう抽出するか?
 - 適切かつ汎用的なメタデータの定義とメタデータDB(検索機能含む)の整備
 - より詳細な観測ステータスDBの構築・整備

衛星軌道条件・観測モード(機器HK) 情報の検索システムの整備・充実

あけぼの衛星(1989~)の場合

* 軌道情報

約2,700万件

*機器HK情報

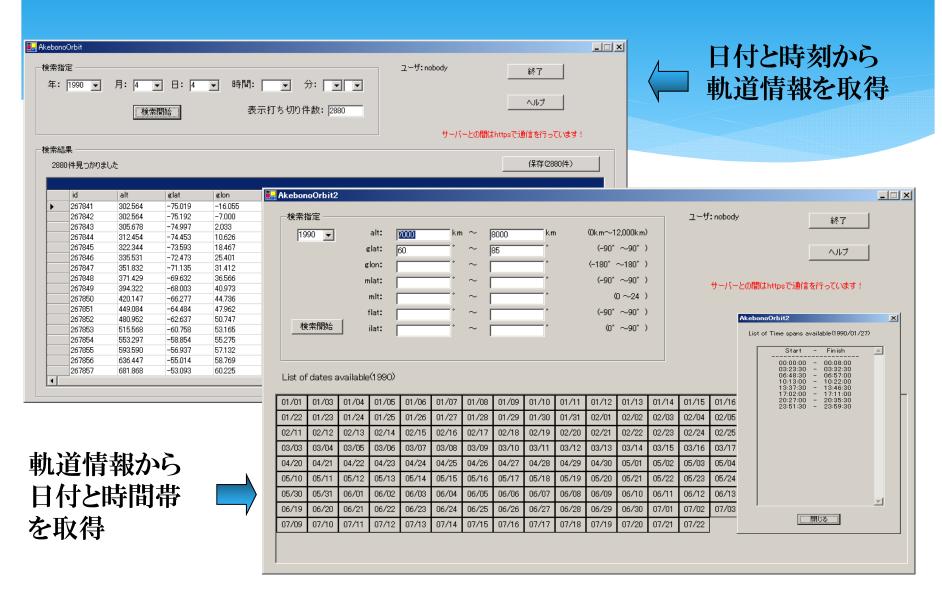
約16億件

(ただし、観測モードの種類は限定的)



KAGUYA, MMO etc. の近年の科学衛星 観測器のインテリジェント化に伴い、観測モードが多様化・複雑化 (1サブ機器だけで数十~数百種類の観測モードを実装)

軌道条件検索画面(Akebono)



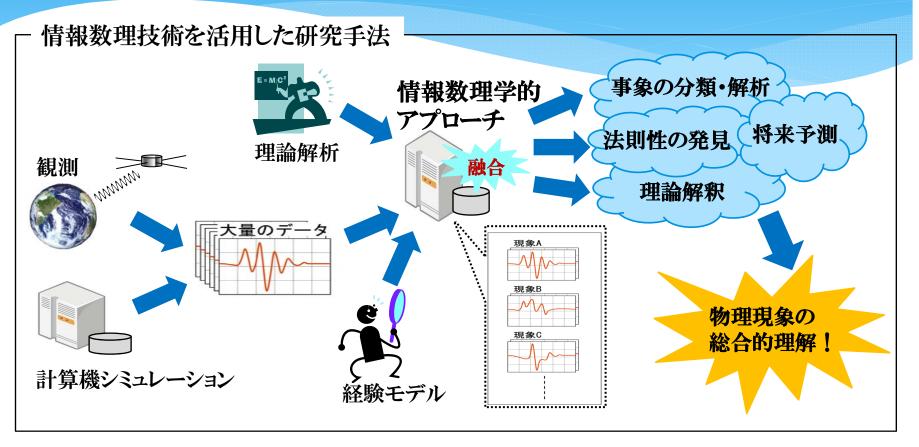


現行版は他システムとの連携が考慮されていない&一般ユーザには難解⇒要改善

何が必要?これからすべきことは?

- A) 大容量データのDB化と公開システムの整備
 - 研究者に負担がかからないデータ生成・蓄積機構の構築
 - 利用しやすいデータフォーマット・解析ツールの提供
 - 適切なデータ公開/利用ポリシーに基づくデータ閲覧機構
 - 適切なデータのバージョン管理と更新
- B) 所望データの条件に合致したデータ検索機能の充実
 - → 刻々と変化する観測条件・観測モードから所望データをどう抽出するか?
 - 適切かつ汎用的なメタデータの定義とメタデータDB(検索機能含む)の整備
 - より詳細な観測ステータスDBの構築・整備
- C) 特徴データ・類似データの検索機能
 - 解析者にとって必要なデータと無用なデータの識別
- D) 解析情報・解析データの共有機構の整備・・・共同研究者間の情報流通

情報数理技術を活用した新しいデータ解析法の積極的導入



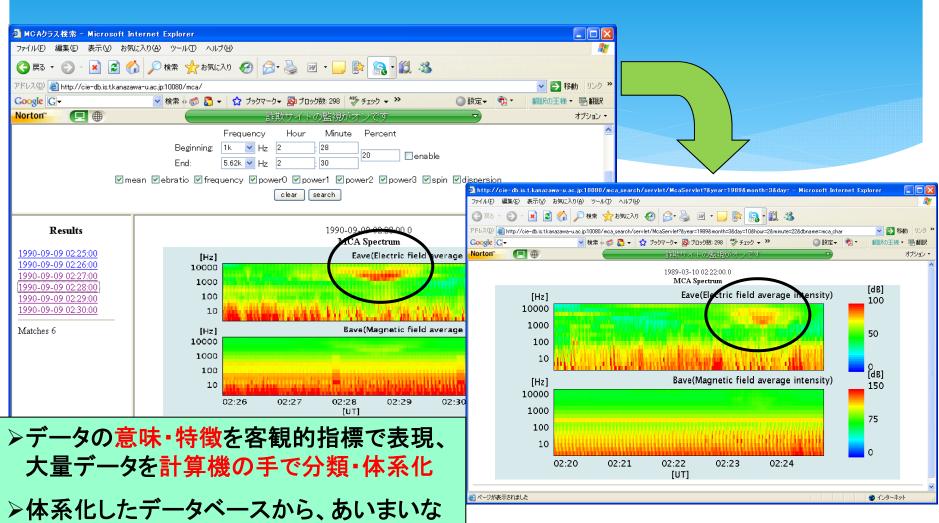
「地球電磁気学・地球惑星圏科学の現状と将来」(2013年1月)より

とはいえ、一朝一夕に実現できるものではない・・

類似データ検索システムの改善・活用

検索語を柔軟に解釈し、特徴的な未知・

発見的データを検索・抽出



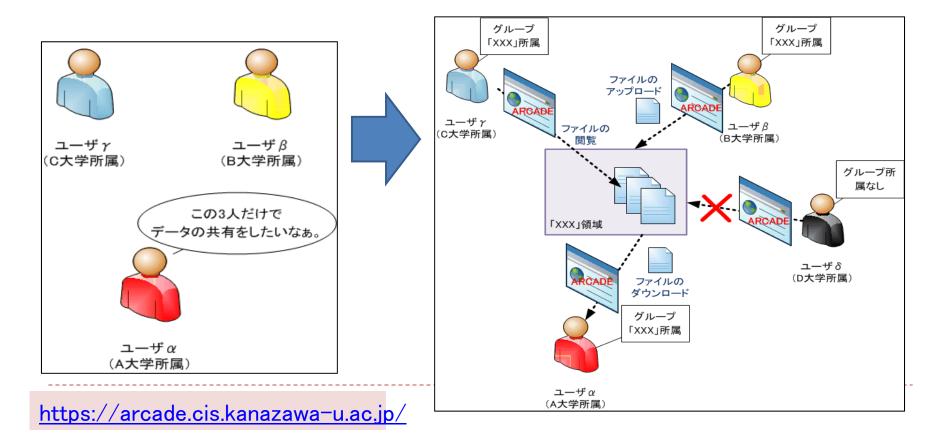
あけぼのVLF/MCA・かぐやWFCで実証試験中

→ 将来的にあけぼのWBAへの適用をめざす!

組織間データ共有システムARCADE設計コンセプト

ARchive system for Cross-reference Across Distributed Environment

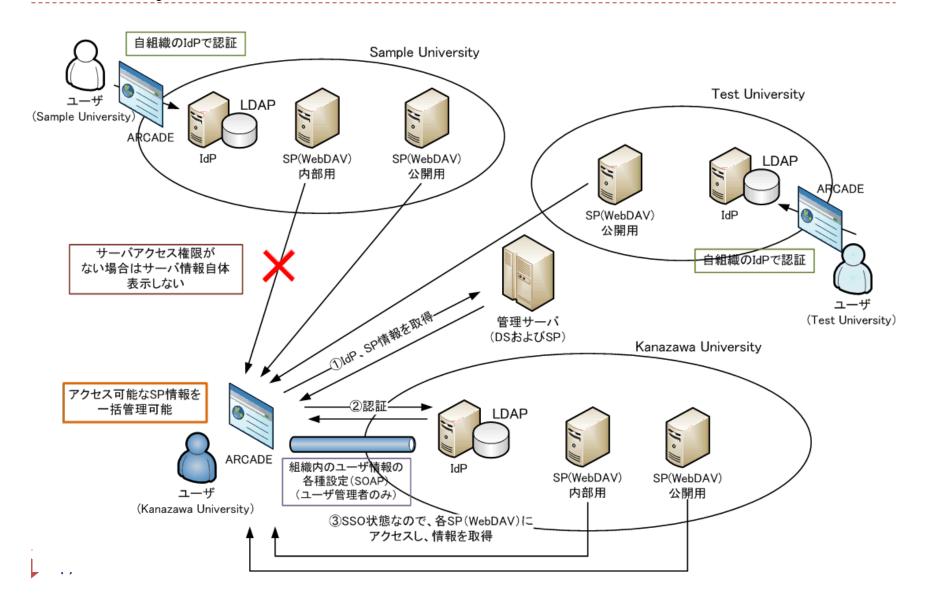
- **異なる研究組織**に属するユーザーが自由に電子ファイルを共有可能 (学術情報なので信頼できるファイルサーバでデータを管理)
- ▶ 使い勝手はDropboxのようにドラッグ&ドロップで使用できる
- ▶ アクセス権限の設定が、個人・研究室レベルの小規模単位で自在にできる



Concept of ARCADE

https://arcade.cis.kanazawa-u.ac.jp/

ARchive system for Cross-reference Across Distributed Environment



まとめと将来展望

- ンメタデータ定義・登録、実データ登録
 - ✓ データのアーカイブ&リポジトリ登録体制はほぼ確立
 - ✓ 様々なメタデータDB, 実データDBとの連携強化
- ▶より高度なデータ検索機能の開発
 - ✓ 観測モード検索機能の拡張・整備
 - ✓ データカタログ・イベントリストの作成とデータベース化
- ▶情報数理技術の積極的な導入
 - ✓ 類似データ検索機能
 - ✓ 解析情報・解析データの共有機構 (連携認証機能の導入を含む)

実解析にそのまま直結できる仕組み作りの実現!