

RDA東京大会参加報告

能勢 正仁

京都大学大学院理学研究科 地磁気世界資料解析センター/
地磁気世界データセンター・京都

私の背景

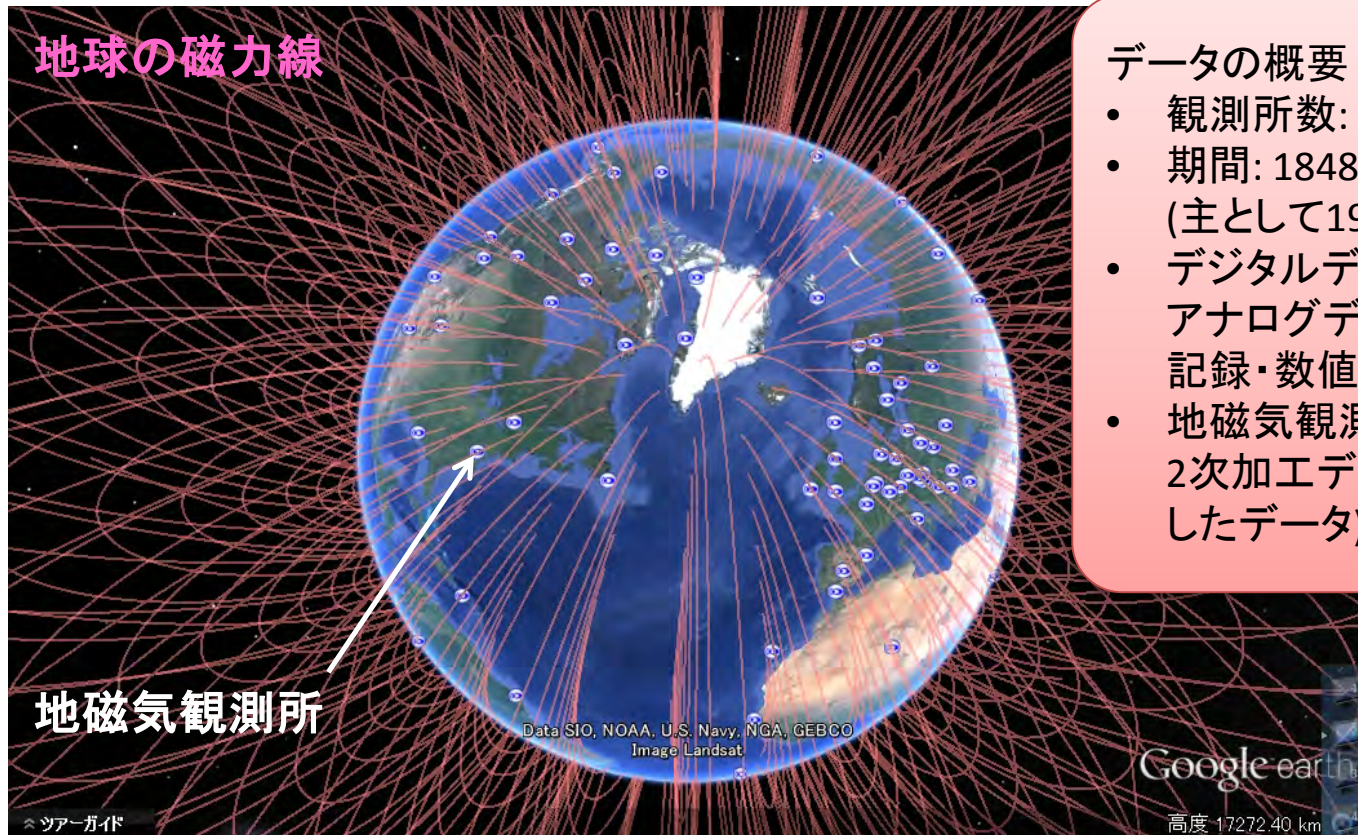
所属：京都大学大学院理学研究科 地磁気世界資料解析センター/

地磁気世界データセンター・京都(World Data Center for Geomagnetism, Kyoto)

✓ 国際科学会議・世界データシステム(ICSU, World Data System)のメンバー

専門：地球電磁気学・超高層物理学

業務：自身の研究、大学院生の教育、データセンターの業務

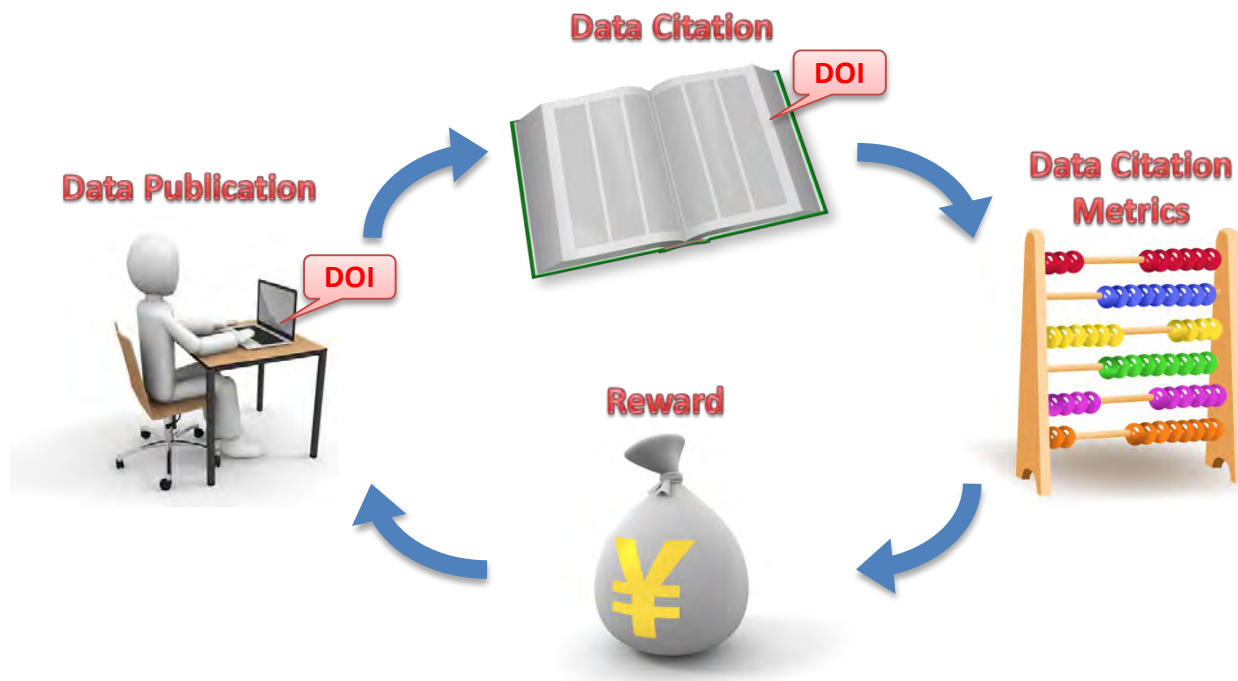


データの概要

- 観測所数: ~150ヶ所
- 期間: 1848年~現在
(主として1957年以降)
- デジタルデータ、
アナログデータ(印画紙
記録・数値表)
- 地磁気観測データ、
2次加工データ(指数化
したデータ)

オープンサイエンスに対する立場

- オープンサイエンスの重要な側面の一つは、データに永続的識別子を付与することにより可能になる「**データ出版・データ引用**」である。
- データの作成、データベースの維持管理、データ処理による2次的価値の付与など、これまで**データ提供者・データセンター**が行ってきた**労力を明らかにできる**。
(研究者にとっての**出版論文数**と同様の意義)
- データ引用により、**データ提供者・データセンター**の寄与を客観的に評価できる。
評価に見合った対価を得ることができる。
(研究者にとっての**被引用論文数**と同様の意義)



RDAとの関わり

RDAへの参加

- 6th Plenary Meetingを含め、2回目
- 最近になって、RDA/WDS Publishing Data Bibliometrics WG、Repository Audit and Certification DSA–WDS Partnership WG、Data Rescue IGの3つのグループに参加し始めた。

関心のあるRDAにおける議論

- 永続的識別子(PID, Persistent Identifier)
- データ出版・データ引用・計量(metrics)
- World Data System(WDS)に関係する議論
- 過去のアナログデータの保管

RDA 7th Plenary Meetingの報告事項

- 3月2日 09:00-10:30、RDA Recommendations & Outputs Session
 - RDA/WDS Publishing Data **Bibliometrics** Recommendations
- 3月2日 11:00-12:30、Working Meeting Session 4
 - Joint meeting: **IG Data Rescue**, IG Data Fabric, IG Preservation e-Infrastructure, IG Domain Repositories, IG Libraries for Research Data

RDA/WDS Publishing Data Bibliometrics Recommendations

- RDA/WDS Publishing Data Bibliometrics Working Groupによる勧告
 - Chairs: Kerstin Lehnert, Todd Carpenter, John Kratz, Sarah Callaghan
 - 目的: データ引用の計量とそれに関わるサービスについて最適なものを明確にする。
- 18ヶ月間の活動のまとめ。
- 勧告は、草稿が作成された段階で、最終稿はもう少し時間がかかる。
- おおまかな内容
 1. 基本事項の整理
 2. 関係者へのアンケート調査と現状分析
 3. 計量実施へ向けた他の動きの紹介
 4. 勧告内容

草稿は次のサイトから閲覧できる。

<https://docs.google.com/document/d/10tzs97ZemZc35PRqehwROce3Mb9kBAjuaRU4jaxXd0k>



1. 基本事項の整理

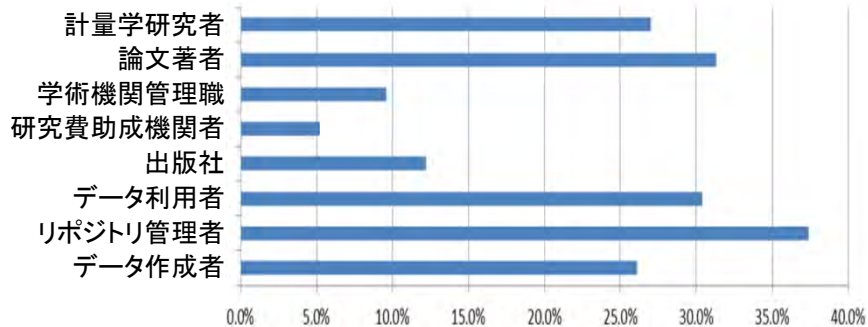
- データ引用計量へのアプローチは、利害関係者毎に異なる。

利害関係者	要求	理由
データ提供者	自分の作成したデータが誰にどのように利用されているか知りたい	1. 次の研究費を獲得するため 2. 新しい共同研究の機会を見つけるため 3. 既存のデータを元にして、新たな研究につながる新しいデータセットを作成するため
研究費助成機関	データの重要性や有用性を明確に偏見なく評価したい	1. 研究費が適切に使われているか調査するため 2. どのデータセットに次の研究費を助成するかを判断するため
データリポジトリ管理者	どのデータセットが研究コミュニティに最も重要かを知りたい	1. 最も重要で利用されているデータセットに財源を投入するため 2. リポジトリが有用なサービスを提供していることを研究費助成機関へ示すため 3. 効率的なリポジトリ運用のためにデータ収集停止の判断に用いるため
	データ提供者にデータの提供を薦めたい	1. データを長期間利用できるように、付随資料と共にアーカイブするため
データ利用者	研究上の問題を解決するのに最適なデータセットを素早く手軽に見つけたい	1. データセットを探したり評価したりする手間を省くため

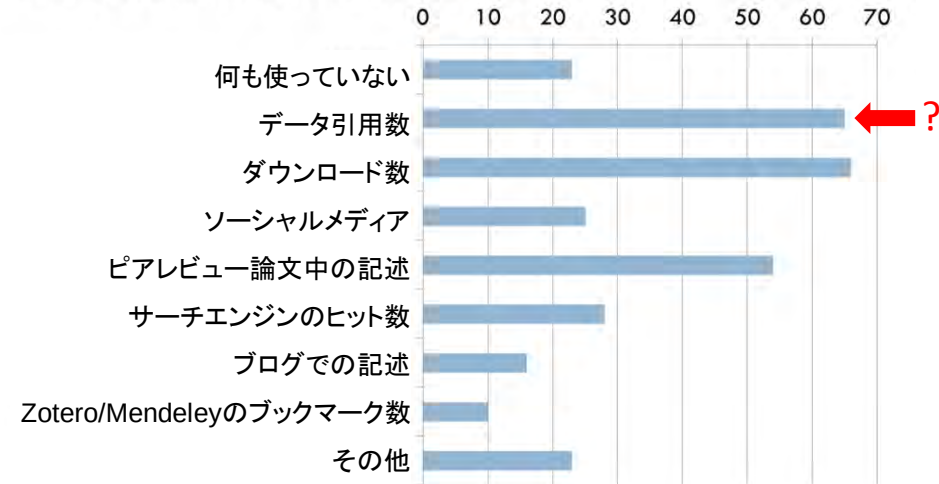
2. 関係者へのアンケート調査と現状分析

- 関係メーリングリストを通じて、ウェブアンケートを実施。115名から回答。

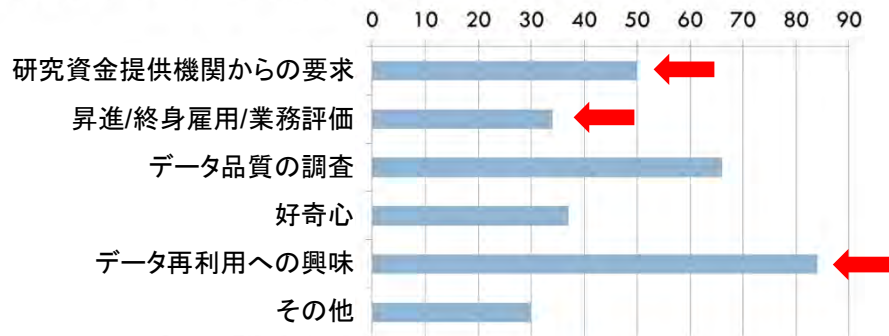
◆ あなたはどのような属性ですか？ (重複回答あり)



◆ データの重要度を評価するために現在何を使っていますか？



◆ なぜデータの重要度の評価に興味があるのですか？



◆ データ引用の計量がもっと広く使われるようになるためには、現在、欠けているものや必要なものは何ですか？

1. 標準
2. データ引用の実例
3. 一貫したPID/DOIの利用
4. 研究文化の変化(データ引用やその計量は役に立つという認識)

4. 勧告内容

- 組織な課題
 - データの利用者が簡単にデータ引用ができるようにすること。・・・リポジトリにデータ引用例を掲載、**学術誌によるデータ引用の手引き**、編集者・査読者によるチェック
 - **研究助成機関がデータの公開とデータの引用を必須にすること。**
 - 現状は、「ある出版社とあるリポジトリの協力」という少数対少数の関係であるが、それを「多数対多数」に拡張するような斡旋サービスが必要。
 - 技術的な課題
 - 学術論文の計量のための既存システムがあるので、それを用いることができる。ただ、データの場合には動的データや粒度という特有の問題を考える必要がある。
 - データ引用が標準になるまでは、**本文中に記載されているデータセットやリポジトリを抜き出すようなテキストマイニングの手法**が必要。
 - 方法論の課題
 - データ引用計量についての研究。(データをオープンにすればデータ引用も増えるか、といった調査)
 - 費用面の課題
 - **リポジトリに多大な費用が掛かる。DOI付与にも多少の費用が掛かる。**
 - OpenなData Citation Serviceといったものが国際的に政府助成金を通して可能か？
DataCiteやCrossRefのような非営利組織がそういったサービスを提供できないか？
- 現状分析や問題点の指摘、今後の方向性や示したもので、継続した議論が必要。
その解決方法などは後継WGで議論される。

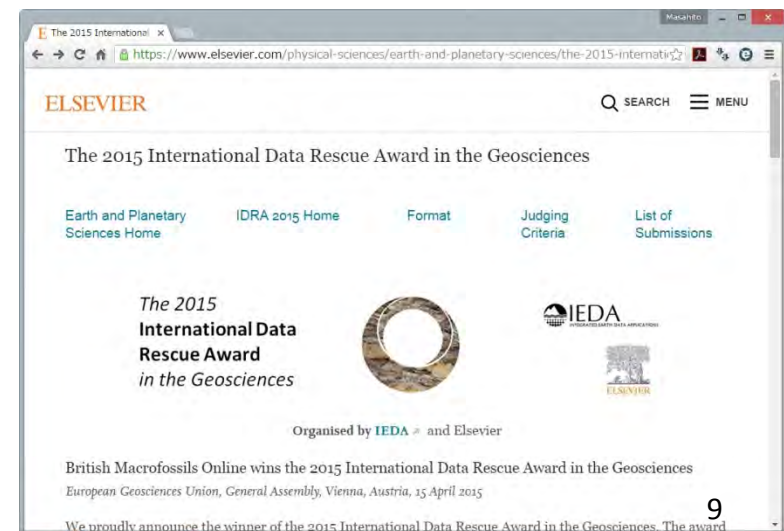
Data Rescue Interest Group

目的

1. 世界で行われているData Rescueの活動をリストアップし、カタログを作成する。
2. データのデジタル化や関連した作業についてのアドバイスシステムを構築する。
3. 他のRDAグループと意見交換し、古いデジタルデータへ新たな価値を付加する。

RDA 7th Plenary Meetingでの会合

- Case Study of successful data-recovery project
 - 50年以上の社会科学データのアーカイブ。パンチカードに記録されたデータをスキャンする。関連資料やコードブックはPDFに。
 - Cornwell Study of occupational retirementとしてICPSRのデータベースに登録。公開。
 - 新しいデータの管理(\$140,000) ≈ 古いデータの救助(\$120,000+時間)
- 3rd international Elsevier Award for "Data Rescue" in the Geosciences (late 2016)
 - 地球科学分野で、古いデータの再利用のために知識やツールや標準を共有するコミュニティ構築を促進するため。
 - 評価基準は、アクセスのしやすさ、使いやすさ、永続性。
 - 2013年16候補、2015年33候補、2016年4月に公募要項。



まとめ

- オープンサイエンスの重要な側面の一つは、データに永続的識別子を付与することにより可能になる「データ出版・データ引用」である。
- RDAでPublishing Data Bibliometrics WGが形成され、そこで言及されている。
- 国際的なオープンサイエンスの議論・動向に対応し、2015年3月には内閣府・総合科学技術・イノベーション会議から最終報告書「我が国におけるオープンサイエンス推進のあり方について」が公開された。
 - オープン研究データに関する5原則 (8-9ページ)
 - b. 容易にアクセスできること
「研究データの引用としてDOI(Digital Object Identifier)のような世界標準仕様に沿ったデータの同定と所在地(Landing Page)を特定する用意が必要である。」
 - オープンサイエンス推進の基本的考え方 (15-16ページ)
 - 4. 公的研究資金を用いた研究を実施する機関の責務
「保存すべき研究成果及び研究資源の全てに永続性のあるデジタル識別子(Persistent Object Identifier)を付与し、管理する仕組みを確立する必要がある。」
 - オープンサイエンスを推進する際の留意点 (20ページ)
 - 5. 研究者及び科学コミュニティに対するインセンティブ
「これまでデータを作成し、他の研究者に提供・利用できるようにする活動は、論文投稿に比べて、十分に評価されてこなかった分野も多い。このため、政策誘導として、研究者及び科学コミュニティに対するインセンティブを高め、オープン化に対する努力を評価することが重要である。」

まとめ

• 今後の方針

- DOI付与の継続・・・「JaLC研究データへのDOI登録実験プロジェクト」の成果。地磁気世界資料解析センターが保持している地磁気データへの付与。
- 出版社への働きかけ・・・超高層物理学分野の学術誌(編集者)との議論。
- 国内研究費助成機関への働きかけ
- RDAとの関わり・・・Plenaryミーティングは、成果発表の場。その間のテレビ会議、メーリングリストでの議論が重要。

doi:10.17591/55838dbd6c0ad

The screenshot shows a web browser window with the URL http://www2.nict.go.jp/isd/doi-landingpage/wds/10.17591_55838dbd6c0ad. The page title is "Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska". The main text describes the data source: "Horizontal wind velocity in the altitude range of approx. 60-90 km is observed with Poker Flat MF (medium frequency) radar, using the radar wave at 2.43 MHz. The radar receives weak radio echo signals returned from the weakly ionized atmosphere (ionospheric D-region) at the target altitudes, to deduce horizontal air motions (Murayama, Y., K. Igarashi, D. D. Rice, B. J. Watkins, R. L. Collins, K. Mizutani, Y. Saito, and S. Kainuma, Medium Frequency Radars in Japan and Alaska for Upper Atmosphere Observations, IEICE Trans., E83-B, pp.1996-2003, 2000). Poker Flat MF radar has been constructed as part of Japan-US joint research program of Arctic middle & upper atmosphere ("Alaska Project") in collaboration between National Institute of Information and Communications Technology, Japan (formerly Communications Research Laboratory), and Geophysical Institute, University of Alaska Fairbanks." Below the text is a "Data Citation" section with the citation: "Citation: Alaska Project of NICT (CRL)-GI/UAF, Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska, doi:10.17591/55838dbd6c0ad". A "General Characteristics" section lists: "Parameters: Mesospheric horizontal wind velocity", "Processing level: Latitude: 65.1, Longitude: -147.5", "Temporal resolution: 30 minutes", "Start date: 1998-10-16T01:45:00", and "Stop date: -PT1D". At the bottom, there is a map of Alaska with a red dot indicating the location of Poker Flat.

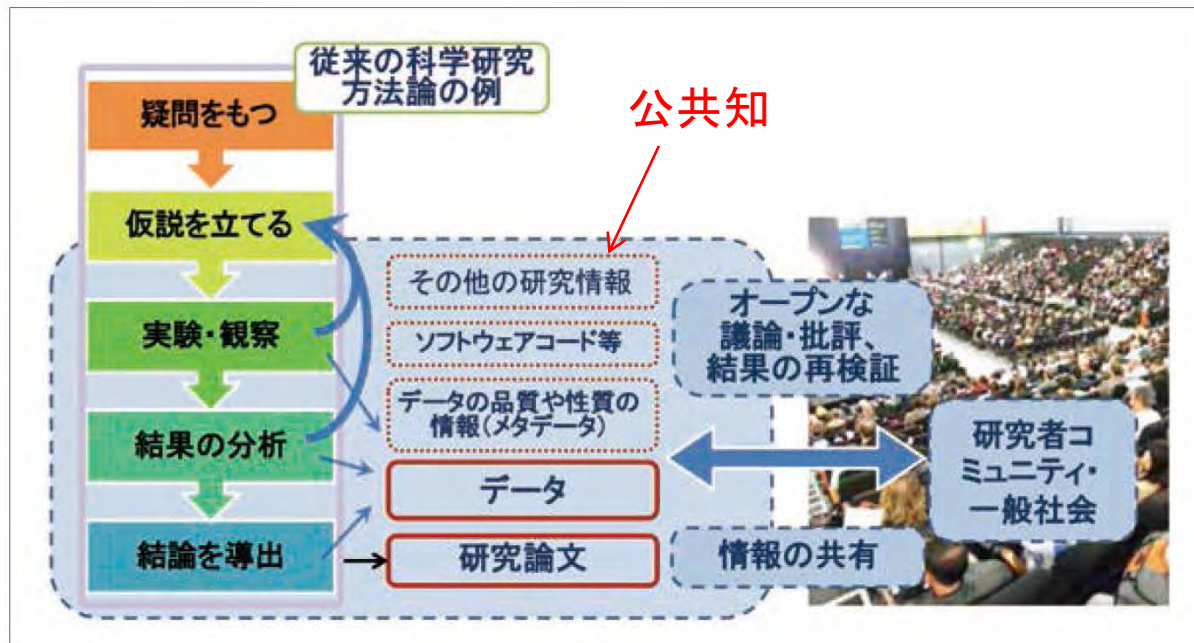
doi:10.17593/14515-74000

The screenshot shows a web browser window with the URL www2.nict.go.jp/isd/doi-landingpage/wds/10.17593_14515-74000. The page title is "Dst Index". The main text describes the index: "The Dst index was proposed by Sugiura [1964] to measure the magnitude of the current which produces the axially symmetric disturbance field. It is derived from geomagnetic field variations in the H component measured at 4 stations at low latitude." Below the text is a "Data Citation" section with the citation: "Citation: World Data Center for Geomagnetism, Kyoto, M. Nose, T. Iyemori, M. Sugiura, T. Kamei (2015), Geomagnetic Dst index, doi:10.17593/14515-74000". A "General Characteristics" section lists: "Parameters: Geomagnetic Index", "Processing level: Final", "Temporal resolution: 1 hour", "Start date: 1957-01-01T00:00:00", and "Stop date: 1 hour, before Present". An "Observatory Location" section lists four observatories: "Kakioka" (Latitude: 36.23°, Longitude: 140.19°), "Honolulu" (Latitude: 21.32°, Longitude: 202.00°), "San Juan" (Latitude: 18.11°, Longitude: 293.85°), and "Hermanus" (Latitude: -34.43°, Longitude: 19.23°). At the bottom, a note states: "(Note: Latitudes and longitudes are expressed in decimal degrees. Eastern longitudes and Northern latitudes are positive.)".

Backup Slides

オープンサイエンスとは？

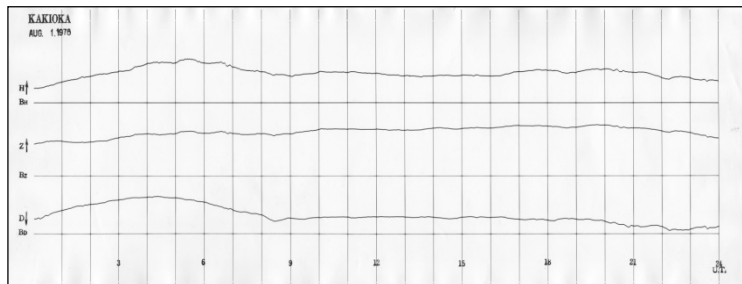
- オープンサイエンスとは、
 - 公的研究資金を用いた研究成果(論文、**観測データ**)に対し、
 - 専門分野・産業界・社会一般から広く容易なアクセスを可能にし、
 - 効果的に科学技術研究を推進する、とともに
 - 追試験・再検証により研究結果の再現・検証が可能となるようなサイエンスの進め方のこと。



地磁気世界資料解析センターで行っているData Rescue

(1) マグネトグラム

(印画紙に記録された地磁気変動データ)



(2) スキャナーでデジタル画像化(600 dpi)

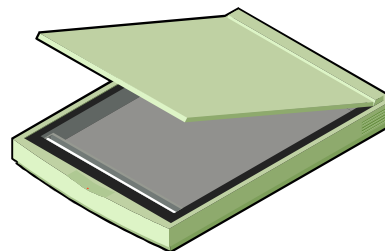
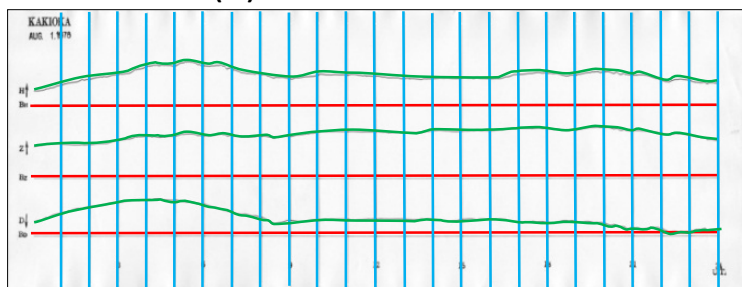


Image files

画像上の線をプログラムでトレース、半自動化
(柿岡地磁気観測所・増子氏開発)

(3) トレースの結果



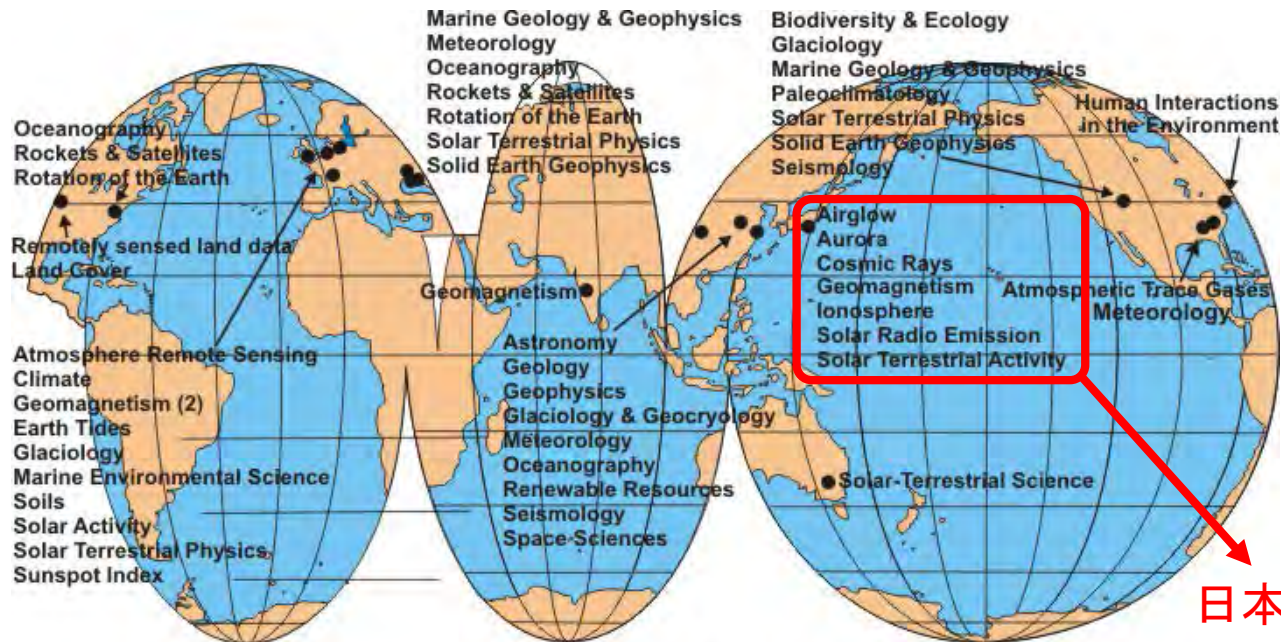
(4) 高時間分解能のデジタルデータ (1分値および7.6秒値)

H	x-axis	y-axis	hour	minute	height
412	242	0	0.764	10.245	
413	241	0	0.892	10.202	
414	241	0	1.019	10.202	
415	241	0	1.146	10.202	
416	241	0	1.274	10.202	
417	241	0	1.401	10.202	
418	241	0	1.529	10.202	
419	240	0	1.656	10.160	
420	240	0	1.783	10.160	
421	240	0	1.911	10.160	

1956年-1975年の20年分の柿岡地磁気観測所のマグネトグラムについて処理

WDC (World Data Center)

- 1957年7月-1958年12月：国際地球観測年(IGY)実施、66か国、約4000観測点
 - 気象・地磁気・極光・電離層・大気光・太陽活動・宇宙線・ロケット・人工衛星・緯度・経度・氷河・海洋・地震・重力・大気放射能
- 1957年-1958年：World Data Centerの創設
 - IGY期間中に汎地球的に取得された地球物理・太陽データを保存するため。
- 2007年：創設50周年、12ヶ国、51ヶ所のデータセンター。
 - 分野も土壌、気候、生物多様性、環境などを含むようになった。



日本国内は7ヶ所

WDCからWDS (World Data System)へ



- 50年が経過した後、International Council for Science (ICSU、国際科学会議)傘下のデータ関連組織の見直しが行われた。
- 2009年より、World Data Systemとして新たに発足。(WDC+FAGS=WDS)
 - 自然科学分野だけでなく、今後は人文・社会科学系のデータセンターも加える。
 - 分野横断型データ利用に対処するため、メタデータやデータ利用システムの整備。
 - データ出版・データ引用への対応。最新のIT技術への対応。
- WDS/International Program Office (WDS/IPO)
 - 情報通信研究機構に設置が決定。2011年度より業務を開始。

