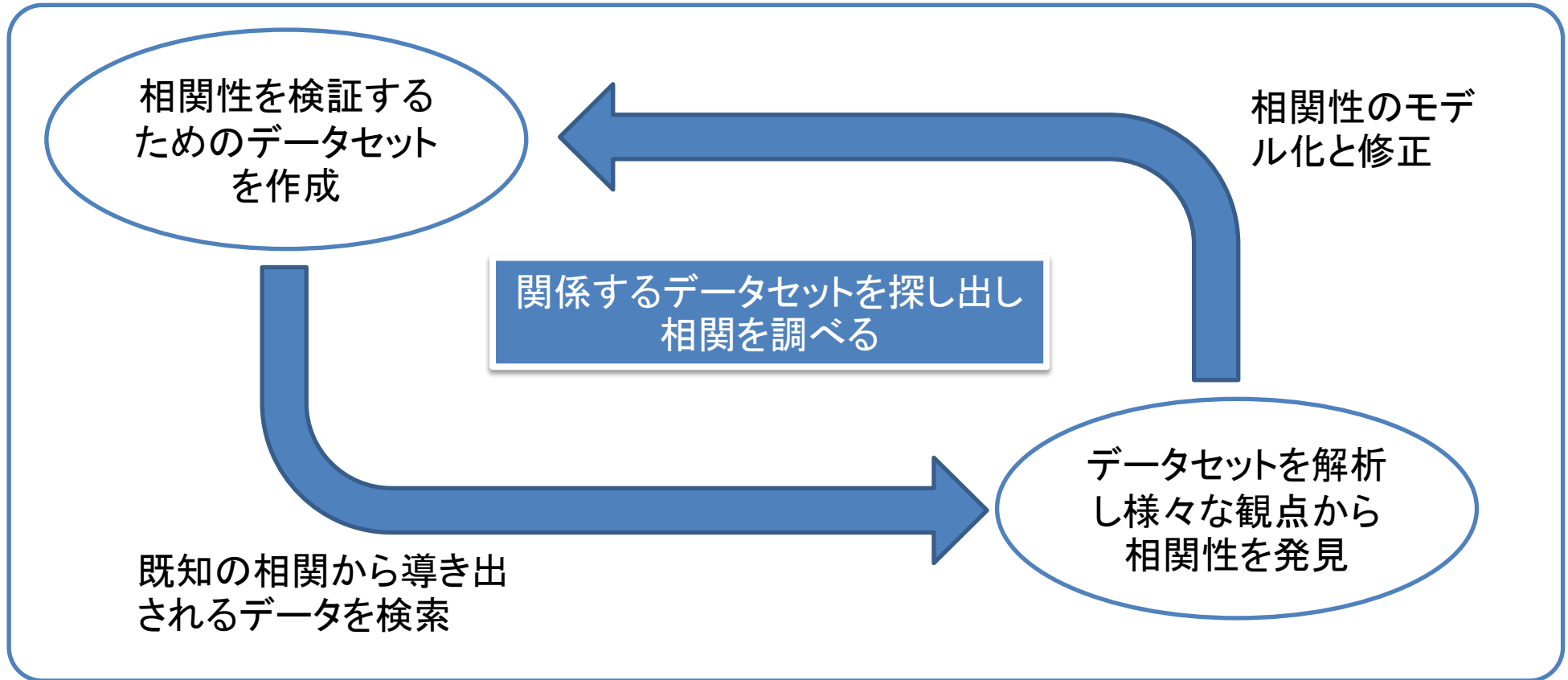


オープンサイエンスデータの 利活用基盤技術

情報通信研究機構

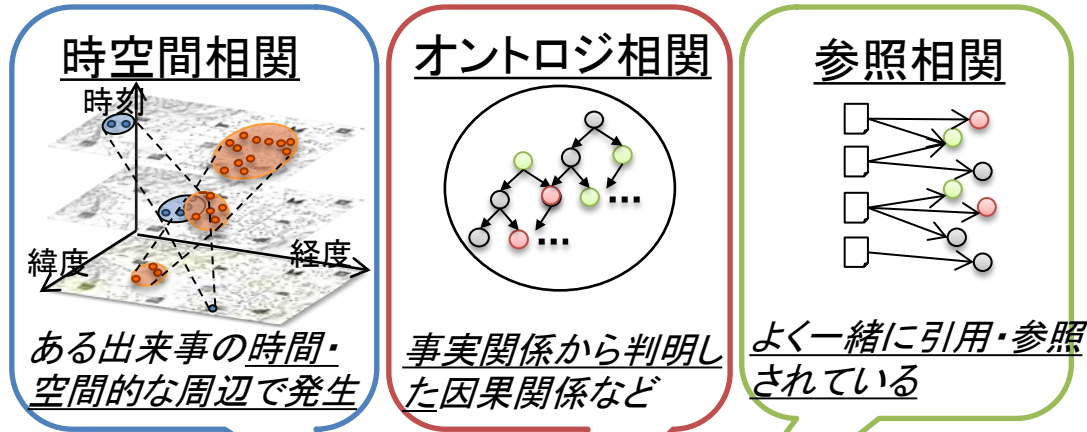
ユニバーサル
コミュニケーション研究所
是津 耕司

統合データシステム
研究開発室
村山 泰啓

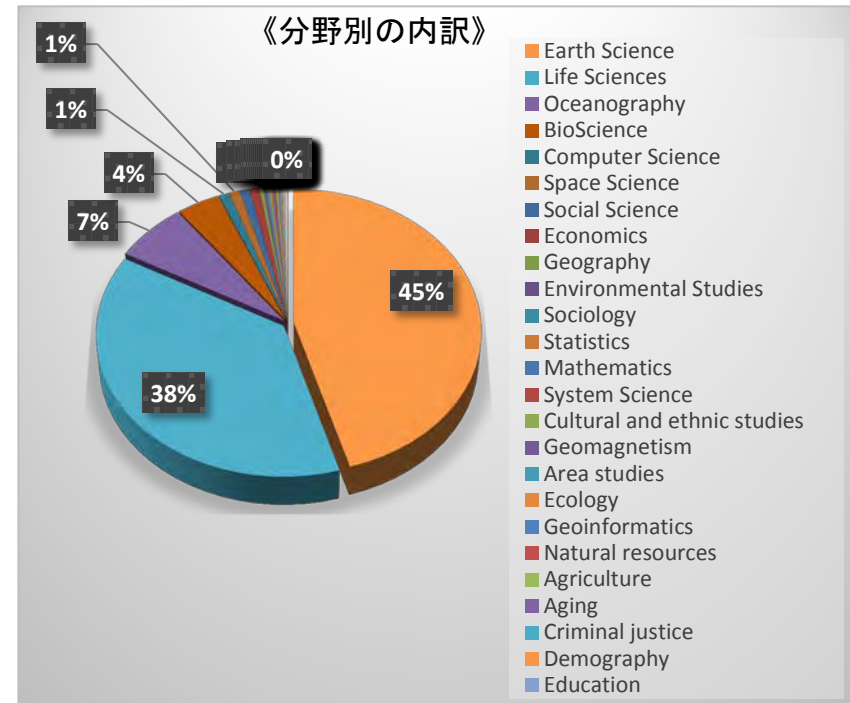
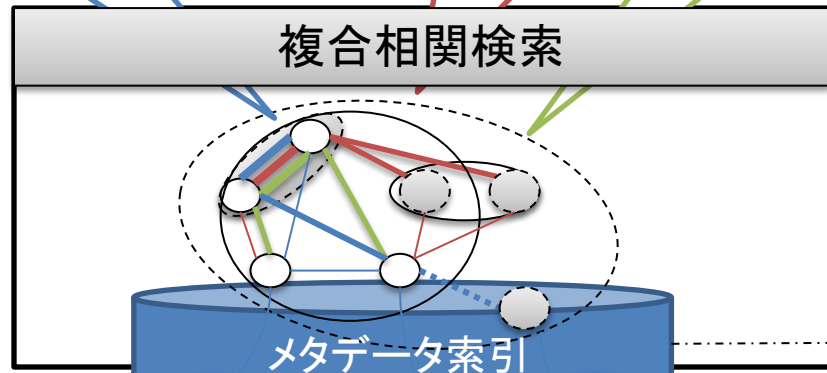


分野横断相関検索(NICT Cross-DB Search)

- 様々な分野の情報資産から相関のあるデータセットを“横串し”検索



- 25サイト40分野113万件のデータを索引付け (2013年10月時点)
 - 各種情報資産から検索用のメタデータを抽出し収集 (harvester)
書誌情報、地理範囲、時間範囲、説明文、データ属性(属性名、データ型、値範囲等)、データサイテーション(引用情報)等



Cross-DB Searchの検索結果例

例) クエリ: 北半球の "ice sheet"



時空間的かつ概念的に相関の高いデータセットを優先的に検索結果に入れる

追加されたデータセットの例

Arkhipov, SM (2000): Chemistry of ice core FRD79. doi:10.1594/PANGAEA.56202.
In: Arkhipov, Serguei M; Kotlyakov, Vladimir; Punning, Ya-M K; Zogorodnov, V; Nikolayev, V I; Zagorodnov, V S; Macheret, Yu Ya; Vaikmaye, R; Barkov, N I; Korsun, SA; Korotkevich, V; Morev, V A; Evseyev, A V; Vostokova, T A; Andreev, Andrei A; Klementyev, Oleg L; Korotkevitch, YS; Stiévenard, Michel; Sinkevich, SA; Samoylov, O Yu; Gordienko, FG; Korsun, A; Tiugu, KR; Arkhipov, SM (2008): Deep drilling of glaciers: Russian projects in the Arctic (1975-1995). Institute of Geography, Russian Academy of Sciences, Moscow, doi:10.1594/PANGAEA.707363

Wasilowska, Agnieszka; Rzepecki, Marek (2010): Hydrochemistry and pigments measured on water bottle samples at Admiralty Bay in 2007. *Department of Antarctic Biology, Polish Academy of Sciences*, doi:10.1594/PANGAEA.740077
 Impact of climate induced glacier melt on marine coastal systems, Antarctica (IMCOAST) ☺

Klages, JP et al. (2013): Clay mineral content of sediment core PS75/234-1. doi:10.1594/PANGAEA.779835.
In Supplement to: Klages, Johann Philipp; Kuhn, Gerhard; Hillenbrand, Claus-Dieter; Graham, Alastair GC; Smith, James A; Larter, Robert D; Gohl, Karsten (2013): First geomorphological record and glacial history of an interglacial ice stream lobe on the West Antarctic continental shelf. Quaternary Science Reviews, 61, 47-61, doi:10.1016/j.quascirev.2012.11.007
 Paleoenvironmental Reconstructions from Marine Sediments @ AWI (AWI_Paleo) ☺

A. 相関の高いデータセットを追加

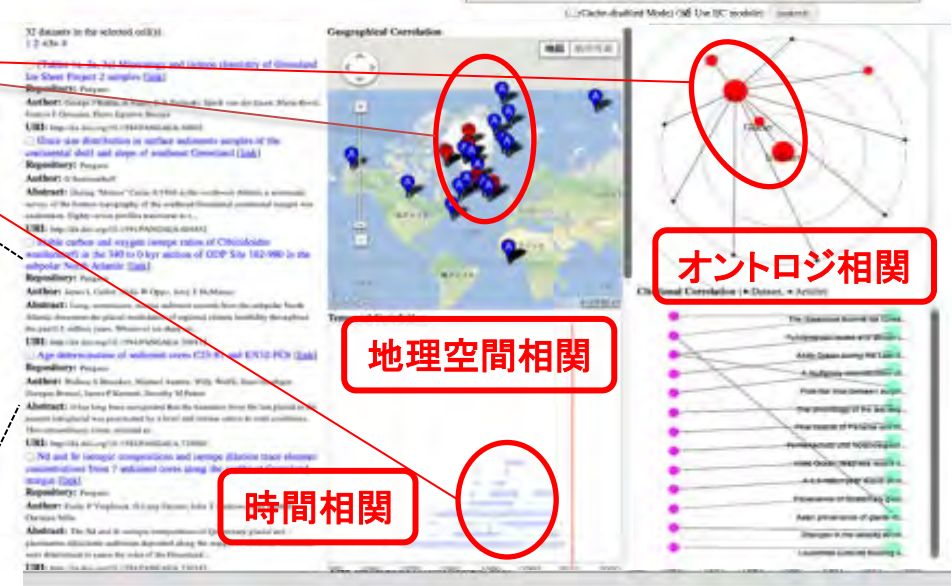
B. 相関の低いデータセットを排除

Cross-DB Searchの検索結果 R'
 $(R' = R + A - B)$

クエリにヒットしたデータセット R
 (ベースライン)

直接クエリにヒットしなくても、クエリに関する数多くのcitationを集めているデータを検索結果に追加

データサイテーション関係の可視化

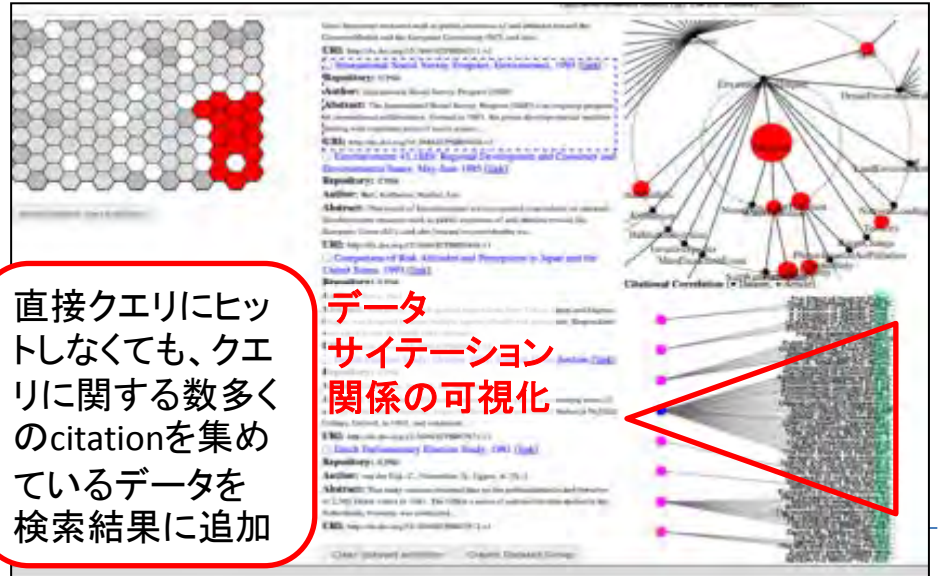


オントロジ相関

地理空間相関

時間相関

クエリ: pollution

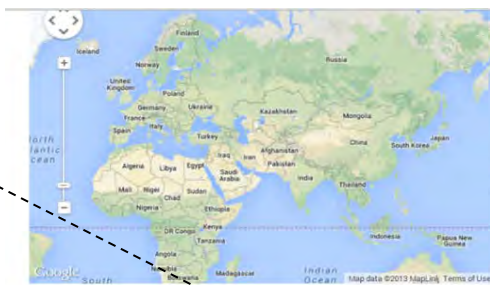


E.g.) Query: PM2.5 effect

29 datasets in selected cell(s).
<1> 2 3

1. [Japan_Airpollution_Soramame](#) [link]
Repository: CPSeoS
Abstract: WD, datetime, PM2.5, WS, SP, NO2, TEMP, NO, HUM, station_code, CH4, SPM, latitude, CO, SO2, city_name, station_name, THC, station_type, longitude, NMHC, NOX, OX
URI: #Japan_Airpollution_Soramame

2. [National Air Pollution Surveillance \(NAPS\) Network in Canada](#) [link]
Repository: GCMD
Author: Government of Canada, Environment Canada, Environmental Technology Center, Analysis & Air Quality Division
Abstract: The National Air Pollution Surveillance (NAPS) Network was established in 1969 as a joint program of the federal and provincial governments to monitor and assess the quality of the ambient air in Cana....
URI: http://gcmd.gsfc.nasa.gov/KeywordSearch/MetadataTransformation.do?Portal=GCMD_legacy&MetadataType=0&ibnode=mdfb4&EntryId=CANEMRCCRS...



Dataset Conditions

(Any fields): PM2.5

Title:

Description:

Author:

Project:

Parameter:

Supplemental Datasets:

Check existing dataset groups

Correlation Constraints

Spatial: (0.5) Temporal: (0.5)

Wide ignore spatial correlation ignore temporal correlation

Narrow ignore spatial correlation ignore temporal correlation

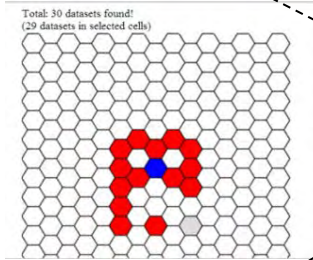
Loose ignore spatial correlation ignore temporal correlation

Tight ignore spatial correlation ignore temporal correlation

Broad ignore spatial correlation ignore temporal correlation

ignore ontology

(地理)空間相関



29 datasets in selected cell(s).
<1> 2 3

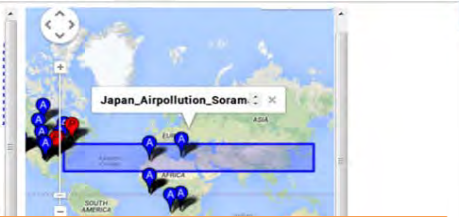
1. [Japan_Airpollution_Soramame](#) [link]
Repository: CPSeoS
Abstract: WD, datetime, PM2.5, WS, SP, NO2, TEMP, NO, HUM, station_code, CH4, SPM, latitude, CO, SO2, city_name, station_name, THC, station_type, longitude, NMHC, NOX, OX
URI: #Japan_Airpollution_Soramame

2. [National Air Pollution Surveillance \(NAPS\) Network in Canada](#) [link]
Repository: GCMD
Author: Government of Canada, Environment Canada, Environmental Technology Center, Analysis & Air Quality Division
Abstract: The National Air Pollution Surveillance (NAPS) Network was established in 1969 as a joint program of the federal and provincial governments to monitor and assess the quality of the ambient air in Cana....
URI: http://gcmd.gsfc.nasa.gov/KeywordSearch/MetadataTransformation.do?Portal=GCMD_legacy&MetadataType=0&ibnode=mdfb4&EntryId=CANEMRCCRS...

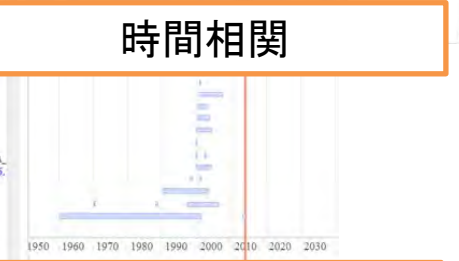
3. [NARSTO EPA_S_S_LOS_ANGELES_PM2.5 Particulate Nitrate](#) [link]
Repository: GCMD
Author: Unknown
Abstract: The overall objective of the LA Superfund SCPCS is to conduct monitoring and research that contributes to a better understanding of the measurement, sources, size distribution, chemical composition, and health effects of particulate matter in the Los Angeles basin.
URI: http://gcmd.gsfc.nasa.gov/KeywordSearch/MetadataTransformation.do?Portal=GCMD_legacy&MetadataType=0&ibnode=mdfb4&EntryId=US_S_S_LOS_ANGELES_PM2.5 Particulate Nitrate

4. [US satellite-derived map of PM2.5 averaged over 2001-2006](#) [link]
Repository: GCMD
Author: Aaron S. Donkelaar
Abstract: In many developing countries, the absence of surface-based air pollution monitors makes it difficult, and in some cases impossible, to get even a rough estimate of the abundance of a subcategory of air pollution.
URI: http://gcmd.gsfc.nasa.gov/KeywordSearch/MetadataTransformation.do?Portal=GCMD_legacy&MetadataType=0&ibnode=mdfb4&EntryId=US_S_S_DerivedParticulateName2001-06

5. [NARSTO EPA_S_S_LOS_ANGELES TEOM PM2.5](#) [link]
Repository: GCMD
Author: Unknown
Abstract: The overall objective of the LA Superfund SCPCS is to conduct monitoring and research that contributes to a better understanding of the measurement, sources, size distribution, chemical composition, and health effects of particulate matter in the Los Angeles basin.
URI: http://gcmd.gsfc.nasa.gov/KeywordSearch/MetadataTransformation.do?Portal=GCMD_legacy&MetadataType=0&ibnode=mdfb4&EntryId=US_S_S_LOS_ANGELES TEOM PM2.5



時間相関



概念相関



複合相関検索

クエリにヒットするデータを起点に、周辺の地域・時間・概念のデータを芋づる式に発見

- Air pollution (particles) data
- Aerosol data
- Disease observation data
- Virus observation data

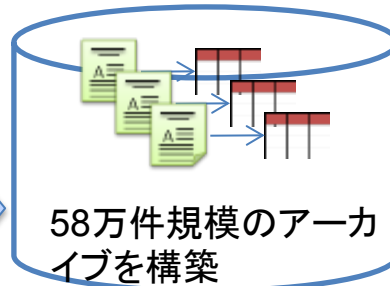
Data citationの編集

オンラインジャーナルやWeb記事からのdata citationを協調編集するData Citation Wikiシステムを開発



編集

Data citation アーカイブ



58万件規模のアーカイブを構築

分析

参照関連マイニング

参照元テキストと参照先データの典型的な組み合わせを発見する方式を実装

テキスト属性

タイトル、トピック、
キーワード、など

データ属性

時間・場所、観測対象、
作成者、特異値など

参照関連
ルール群



Data citationの収集

Pangaea (地球環境データ)やICPSR(社会科学データ)を中心にdata citationをharvesting

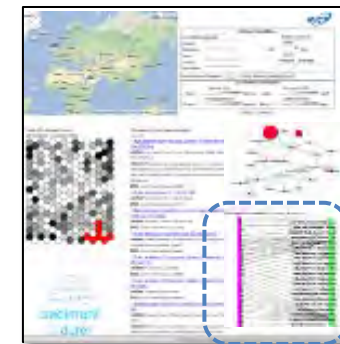


収集

WDS分野横断相関検索への応用

周辺のデータ(時空間相関)や概念的につながりのあるデータ(オントロジ相関)に加え、一緒に使われることが多いデータ(参照相関)を芋づる式に検索

検索

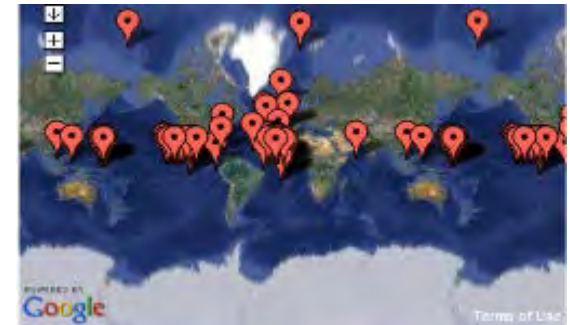
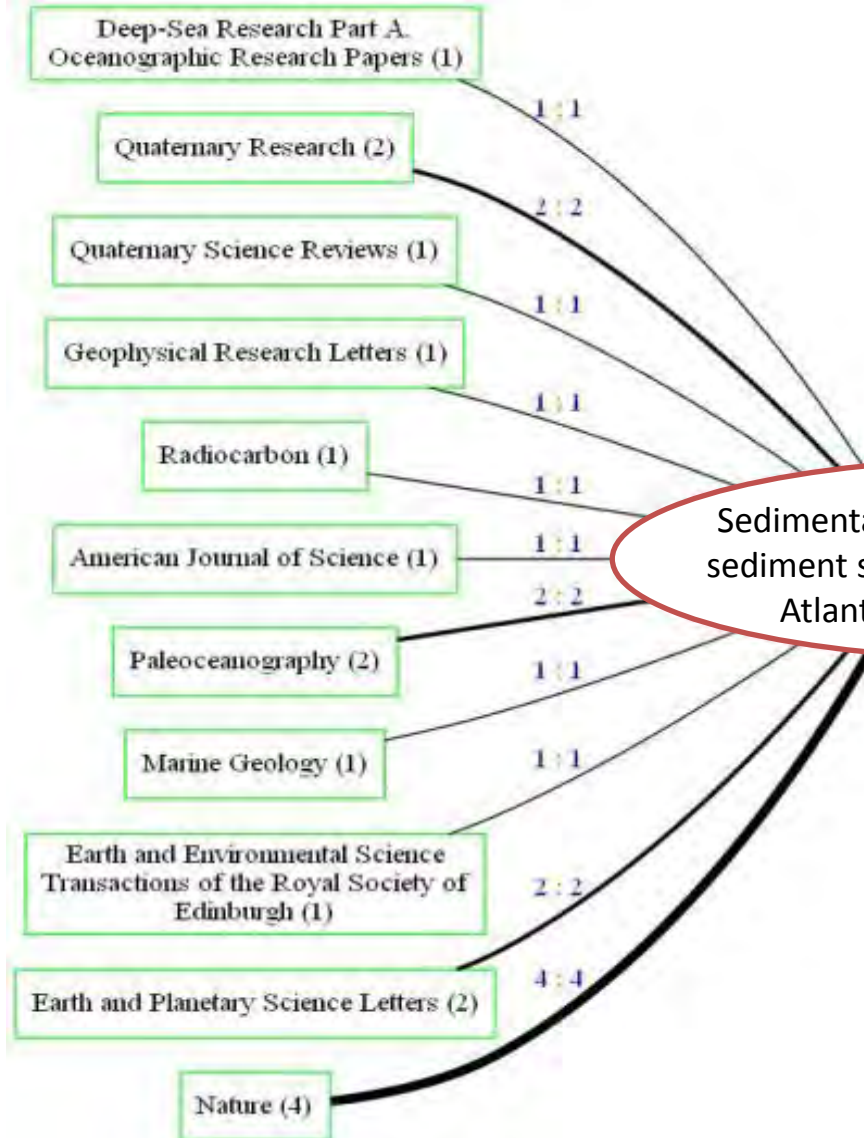


Data Citationに基づく参照相関の例 (2)

テキスト属性

→

データ属性



Data collected from 1963 ~ 1981 all over the world

Sedimentation rates calculated on surface sediment samples from different site of the Atlantic and Pacific Oceans, 1991.



Created by **Wallace Smith Broecker** (1931 -)

This is an authoritative data cited by wide variety of documents...

潜在的対象データの概要

(これまでは原則、共同研究ベースで公開;PIIによって異なる。
⇒できるだけフリーオープンにしていきたい。)

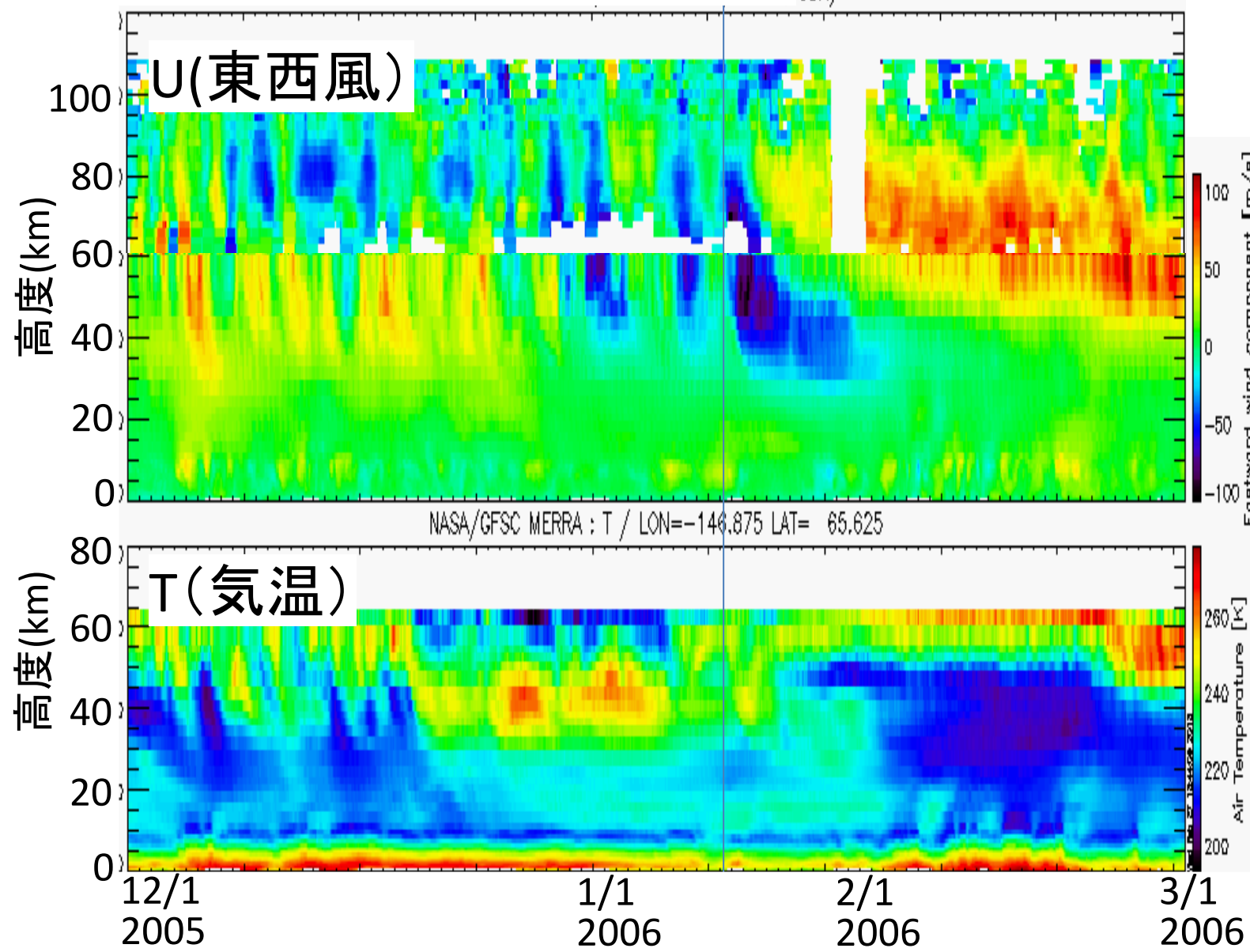
観測地点	観測装置名	観測期間	
アラスカ	イメージング・リオメータ	1995 ~ 1999	超
	ミリ波ラジオメータ	1999, 2003 ~ 2006	成
	レイリーデータ	1997 ~ 2005	成
	多波長ライダー	2003 ~ 2006	対
	ファブリ・ペロー干渉計	2000 ~ 2009	超
	分反射レーダ (FCAWind / METWind / DAE※1)	1998 ~ 現在	中
	FTIR	2007 ~ 2011	対
	大気光イメージャー	2000 ~ 2007	超
	オーロラウェブカメラ	2004 ~ 現在	超
	HFLレーダ	2004 ~ 2007	超
稚内	分反射レーダ (FCAWind / DAE※1)	1996 ~ 現在	中
	VHF	2001 ~ 現在	対
山川	分反射レーダ (FCAWind / DAE※1)	1998 ~ 2008	中

観測地点	観測装置名	観測期間	
沖縄 (亜熱帯)	400MHzWPR	2003 ~ 現在	対
	遠距離海洋レーダ (LROR)	2002 ~ 2010	海
	地上気象測器 (MOS)	2003 ~ 現在	対
	沖縄偏波降雨レーダ (COBRA)	2003 ~ 2004	対
	ドップラソータ (DS)	2001	対
	小金井 (都市環 境)	シーロメータ	2003 ~ 現在
水蒸気ラジオメータ		2004 ~ 2008	対
1.3GHz ウィンドプロファイラ		1993 ~ 2006	対
1.3GHz ウィンドプロファイラ (LQ-4)		2011 ~ 現在	対
UV放射計		2003 ~ 2008	対
地上気象測器		2007 ~ 2011	対
全天カメラ		2010 ~ 現在	対
放射収支計		2010 ~ 現在	対
気象観測装置 (鉄塔上段)		2010 ~ 現在	対
気象観測装置 (鉄塔中段)		2010 ~ 現在	対
気象観測装置 (鉄塔下段)		2010 ~ 現在	対
超音波風速計		2010 ~ 現在	対

上記 赤枠分を登録:30万件(→133万件)

IUGONET/UDASでのデータプロット・解析： 超高層大気観測と全地球気象データ

アラスカ上空のMFレーダー+MERRA (65N)



- UDAS上でのMFレーダー、気象データの処理モジュールを開発
- 図は、2006年2月の成層圏突然昇温の例。
- 中間圏観測レーダーでツールを共通化
→世界中のレーダーデータで全球解析。