



DIASプロジェクトの取組み

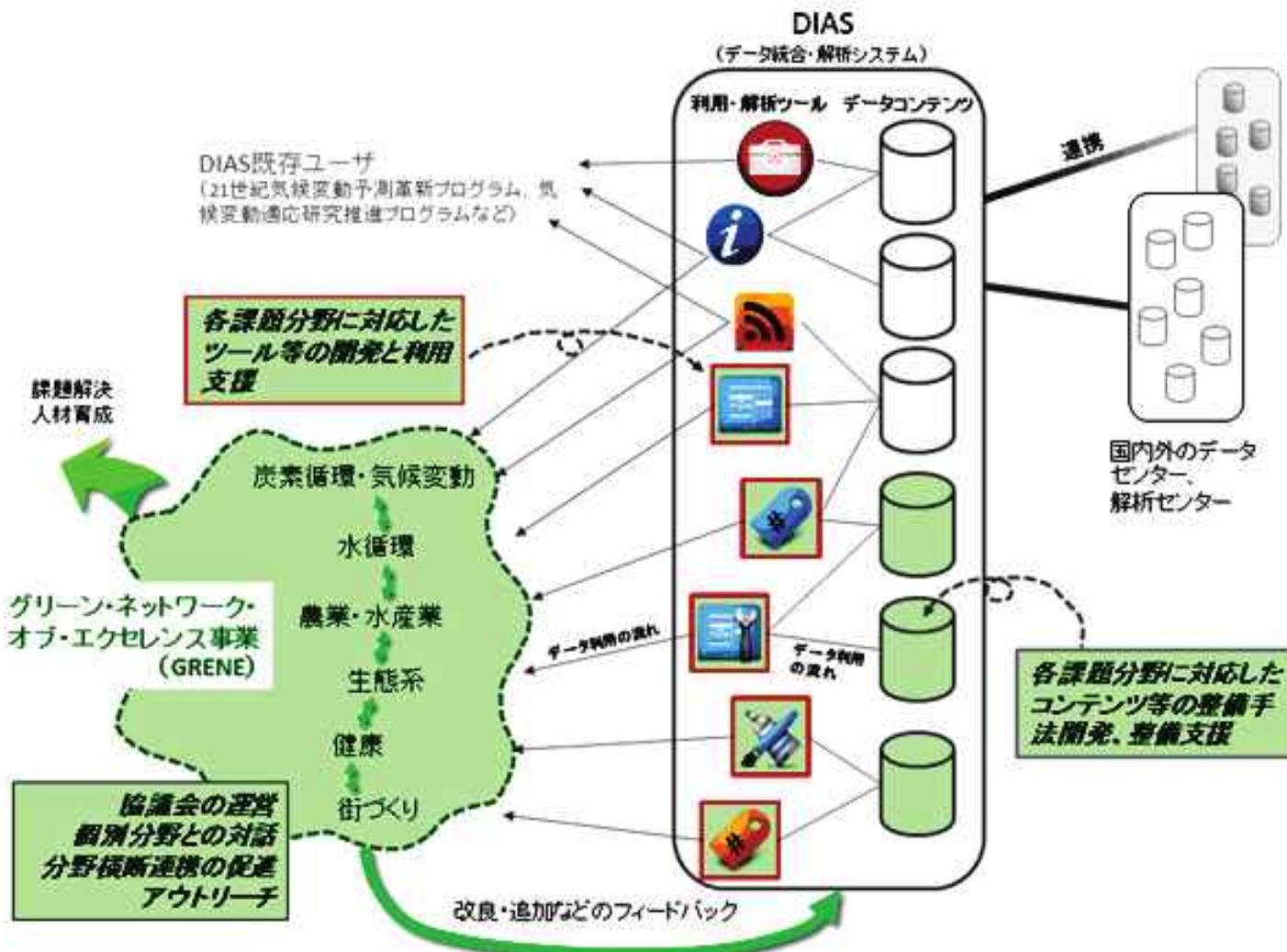
国立情報学研究所
北本 朝展（きたもと あさのぶ）

DIAS-P 地球環境情報統融合プログラム



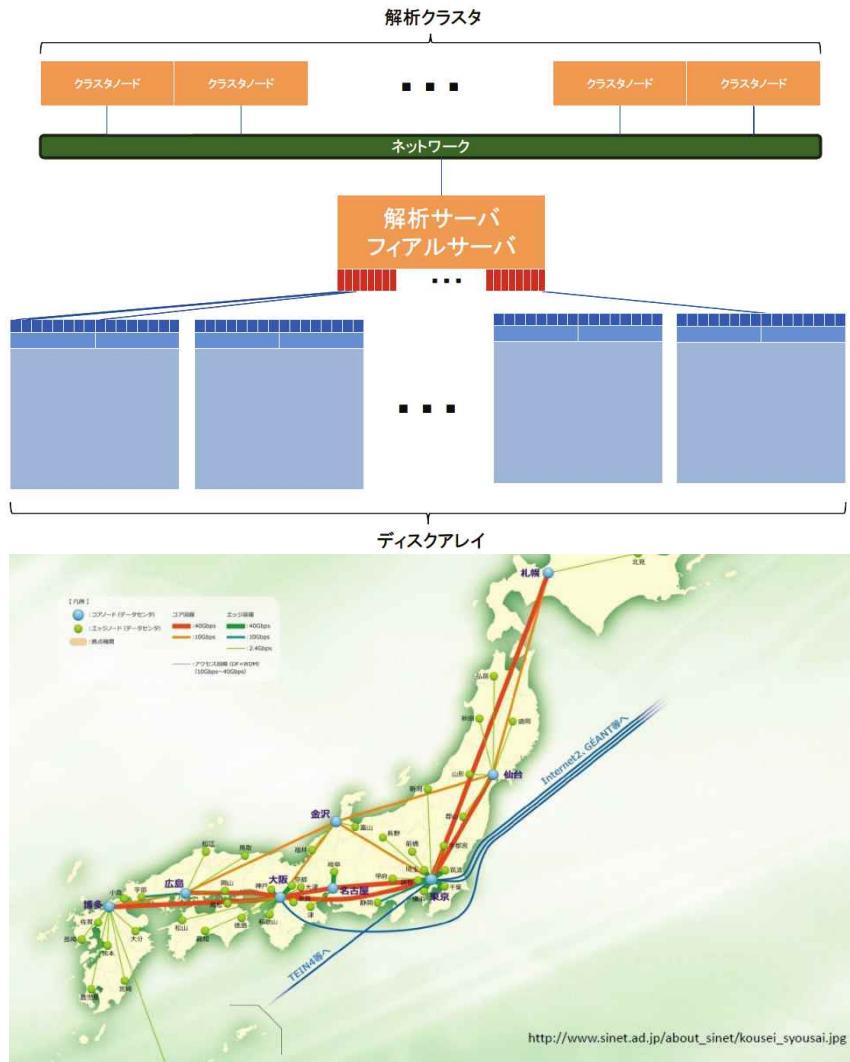
- 平成23年度(2012)～平成27年度(2016)
- 文部科学省気候変動適応戦略イニシアチブ事業のひとつ
- データ統合解析システム(DIAS)の高度化・拡張・利用促進

DIAS-PとGREENE-ei



DIASの情報基盤

システム接続図(2013年度増設)



- ・超大容量ストレージ (ディスク18.8PB + テープライブラリ6.2TB = 合計25PB)
- ・解析クラスタを持つ超大容量解析空間 (16コア×120ノード以上)
- ・分野を超えるデータの相互流通機能
 - ・高速ネットワークSINETによるリアルタイム処理 (最大40Gbps回線)
(・スパコンとの連携)

データの検索・ダウンロード

- ❖ DIASのデータセット全体のマトリックスによる俯瞰
- ❖ キーワード検索/場所検索/期間検索
- ❖ 外部システムとの連携(同時検索)
- ❖ データファイルのダウンロードにつなげる

マトリックスによるデータセット俯瞰

軸の切替

分類に対応するデータセット数

リンクをたどってデータセット一覧を取得(カテゴリによるデータセット検索)

検索結果

外部システムとの連携

データのダウンロード

メタデータのダウンロード

データ俯瞰・検索システムでのドキュメント閲覧ページ

ログイン画面
メールアドレスがあれば誰でもアカウント作成可能

ダウンロード一覧

データポリシーの表示 → ファイルダウンロード

データのダウンロード

メタデータのダウンロード

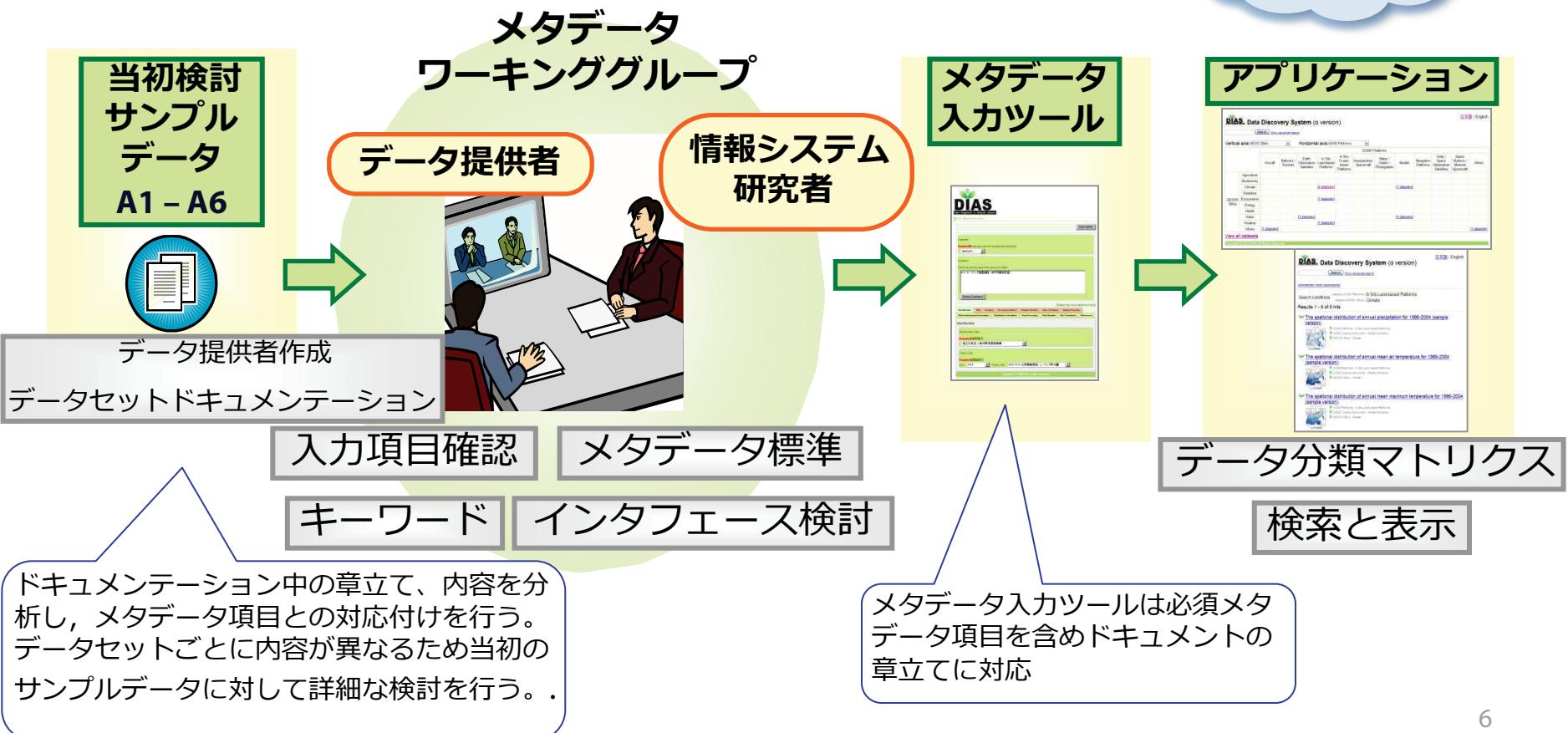
ダウンロード可能なファイル一覧

データポリシーの表示 → ファイルダウンロード

メタデータ形式の検討

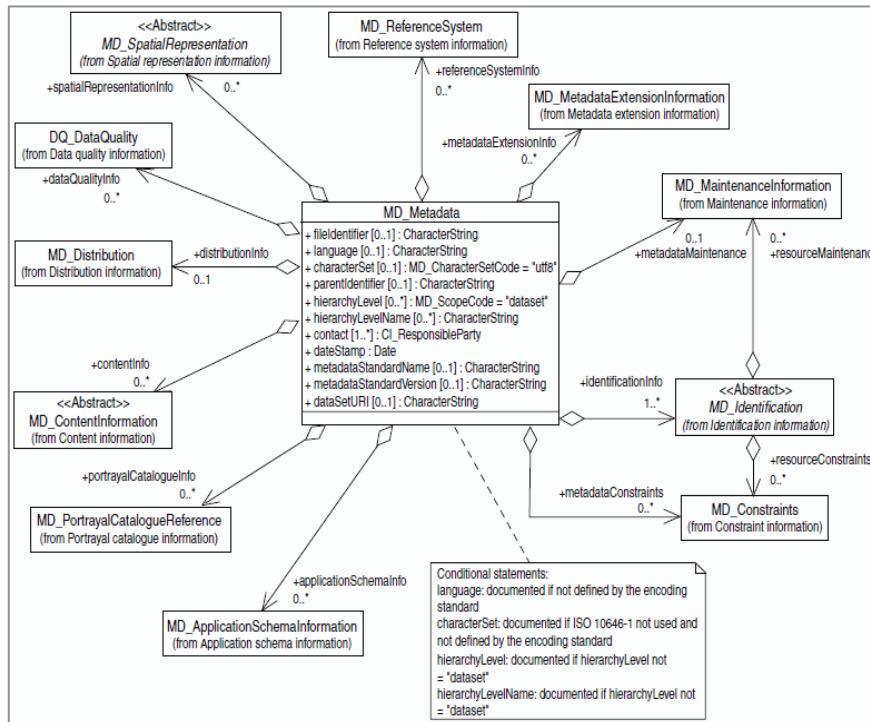
- データ提供者が作成しているデータセット紹介のドキュメントを分析
- ドキュメントに含まれているメタデータの収集
- ISO19115（地理空間メタデータ標準）の項目との対応づけ
- 検索に役立つキーワードについての検討

試行錯誤



メタデータ標準の参照

- 地球環境データの多くは、地理・時間の属性を持つ。
- ISO/TC211 が定めた地理空間メタデータ標準であるISO19115 / ISO19139 を参考に、DIASメタデータ形式を定める。



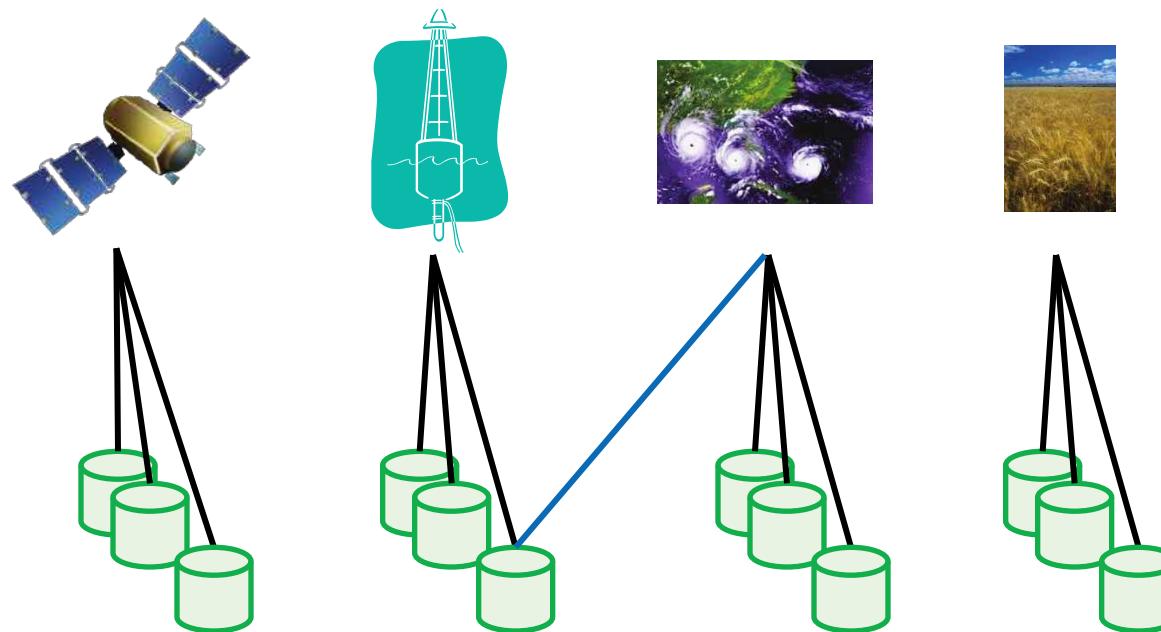
ISO19115

The screenshot shows the XML Schema Definition (XSD) for the ISO 19115 metadata entity. The schema includes:

- Annotations for target namespace (`<xsi:schema targetNamespace="http://www.isotc211.org/2005/gmd" elementFormDefault="qualified">`) and documentation (`<xsd:documentation>This file was generated from ISO TC/211 UML class diagrams == 01-26-2005 12:40:00 =====</xsd:documentation>`).
- Imports for various ISO 19115 components like spatialRepresentation.xsd, metadataExtension.xsd, content.xsd, metadataApplication.xsd, applicationSchema.xsd, portrayalCatalogue.xsd, dataQuality.xsd, and freeText.xsd.
- A complex type definition for `MD_Metadata_Type` which includes:
 - An annotation for information about the metadata.
 - A sequence of elements: fileIdentifier, language, and characterSet.

ISO19139

プロジェクトとデータセットの関係



- ❖ **プロジェクトメタデータ**： 名称, 連絡先, キーワード, ホームページアドレス, データ利用規約, データ引用文, データ免責事項
- ❖ (1) 長期にわたるデータセットの場合、(2) 関与するプロジェクトが複数存在する場合、にも対応できる。

データへのアクセス権限管理

- ❖ ドキュメント/メタデータには全員がアクセスできる。
- ❖ データへのアクセス権限には以下がある。
 1. 完全にフリーアクセス
 2. 同意により利用できる
 3. 承認を必要とする
 4. その他/特殊扱い
- ❖ アクセス権限は、データ提供者が定めるポリシーに従い、個別に設定する。

メタデータ入力キャンプ

- ・ 入力方法・入力内容などが不明なため、メタデータの登録が滞っているデータ提供者に対し、疑問に答えて入力を促進する場を設定。
- ・ 2014年7月、11月、2015年2月に実施。



検索機能

ボタンをクリックして、範囲選択

What? 全て: 降水量
タイトル:
連絡情報:
概要:

Where? 

When? 開始日 1950 1 1 検索条件に含める
終了日 2010 12 31 検索条件に含める
 重複 範囲内
 重複 期間内

キーワード検索 場所指定による検索 期間指定による検索

データ俯瞰・検索システム(β) A Search and Discovery System for DIAS Datasets English

ホーム 使い方 このサイトについて

検索条件 キーワード: 降水量
51 件中 1 - 10 件目

1986-2004年の年降水量の空間分布(サンプル版)
GCMDプラットフォーム: 地上プラットフォーム
GEOSS社会利益領域: 気候
GCMDサイエンスキーワード: 大気, 地表

西スマトラレーダー・雨量計統合降水量データ第1.1版
GCMDプラットフォーム: 地上プラットフォーム
GEOSS社会利益領域: 災害, 水, 気象
GCMDサイエンスキーワード: 大気, 地球水循環

JAMSTEC Data Catalog レーダー・雨量計統合降水量データ
GCMDサイエンスキーワード: 大気

MAHASRI Pathfinder アジア域降水量格子点データ
GCMDプラットフォーム: 地上プラットフォーム
GEOSS社会利益領域: 気候, 水, 気象
GCMDサイエンスキーワード: 大気, 地球水循環

植物指標と気温, 降水量, 光合成有効放射
GCMDプラットフォーム: 地球観測衛星, 地球
GEOSS社会利益領域: 生態系
GCMDサイエンスキーワード: 生物圏, 地球

入力された検索条件にマッチするデータセット一覧を取得

DIAS データ俯瞰・検索システム(β) A Search and Discovery System for DIAS Datasets English

西スマトラレーダー・雨量計統合降水量データ第1.1版
Data File Download データをダウンロードする

タイトル
名前: 西スマトラレーダー・雨量計統合降水量データ第1.1版
略称: WeSRI 1.1
メタデータID: WeSRI20140122104857-DIAS20120518114232-j-a

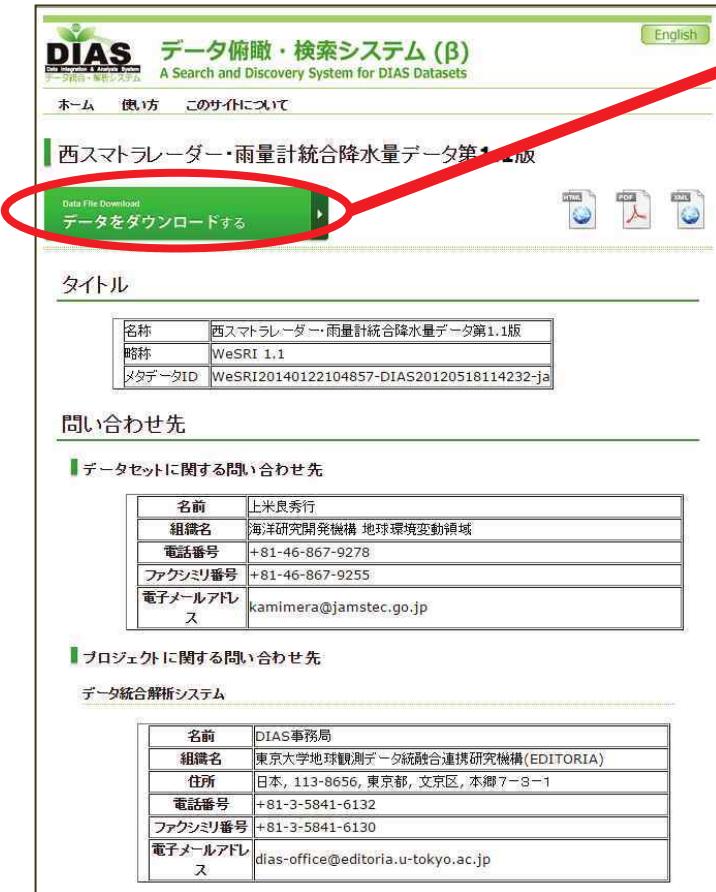
問い合わせ先
データセットに関する問い合わせ先
名前: 上木良秀行
組織名: 海洋研究開発機構 地球環境変動領域
電話番号: +81-46-867-9278
ファクシミリ番号: +81-46-867-9255
電子メールアドレス: kamimera@jamstec.go.jp

プロジェクトに関する問い合わせ先
データ統合解析システム
名前: DIAS事務局
組織名: 東京大学地球観測データ統合連携研究機構(EDITORIA)
住所: 日本, 113-0656, 東京都, 文京区, 本郷7-3-1
電話番号: +81-3-5841-6132
ファクシミリ番号: +81-3-5841-6130
電子メールアドレス: dias-office@editoria.u-tokyo.ac.jp

ドキュメントの閲覧

データダウンロード

- ❖ ユーザ認証
- ❖ データセットごとのファイル一覧管理
- ❖ データポリシーの管理



DIAS データ俯瞰・検索システム (β)
A Search and Discovery System for DIAS Datasets

ホーム 使い方 このサイトについて

西スマトラレーダー・雨量計統合降水量データ第1.1版

Data File Download
データをダウンロードする

タイトル

名称	西スマトラレーダー・雨量計統合降水量データ第1.1版
略称	WeSRI 1.1
メタデータID	WeSRI20140122104857-DIAS20120518114232-ja

問い合わせ先

データセットに関する問い合わせ先

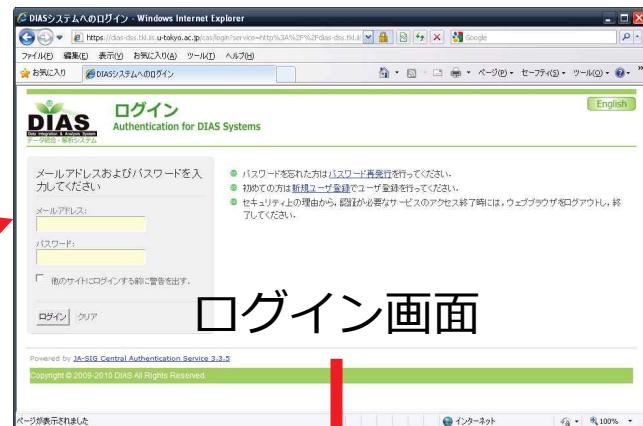
名前	上米良秀行
組織名	海洋研究開発機構 地球環境変動領域
電話番号	+81-46-867-9278
fax番号	+81-46-867-9255
電子メールアドレス	kamimera@jamstec.go.jp

プロジェクトに関する問い合わせ先

データ統合解析システム

名前	DIAS事務局
組織名	東京大学地球観測データ統合連携研究機構(EDITORIAL)
住所	日本、113-8656、東京都、文京区、本郷7-3-1
電話番号	+81-3-5841-6132
fax番号	+81-3-5841-6130
電子メールアドレス	dias-office@editoria.u-tokyo.ac.jp

データ俯瞰・検索システムでの
ドキュメント閲覧ページ



ログイン画面



DIAS ファイルダウンロード

管理者一覧 データセット一覧 ファイル一覧 ダウンロード制限 許可申請一覧 ログアウト

ダウンロード一覧

データセット名
説明

西スマトラレーダー・雨量計統合降水量データ第1.1版
西スマトラ島に展開された気象レーダーと地上の雨量計による観測値を組み合わせて、レーダー単独よりも降水量の見積りが正確で、雨量計単独よりも空間分布が詳しい、格子型の降水量データを作成した。あらかじめレーダーの観測値と地上の雨量計の観測値をつきあわせて経験的に関係式を作成しておき、その関係式をもとにレーダーの観測値を降水量に換算するという処理方式を採用した。レーダーピームが地形に遮られて山かけとなるときのデータの欠損を示するために、地形データに基づいた山かけ評価の手法を開発して適用した。短い波長の電波は降雨減衰を受けやすいが、その影響は考慮していない。

データセットドキュメントへのリンク
http://dias-dmrg.tkl.iis.u-tokyo.ac.jp/doc/cmp/82206_H21-07_0014_2010-09-08.html

ファイル検索

単語はスペース区切りで複数を指定出来ます。
検索モードによる違いは以下の通りです。[検索](#)

1 / 1ページ、4 / 4 レコード、開始レコード 1 , 終了レコード 4

タイトル	ファイルサイズ	作成日時
download all (WeSRI)	1,545,019	2013-01-08 12:40:37
maps > mon > mon_rr > data > v1.1.tar.gz	1,431,984	2013-01-08 12:40:37
maps > mon > mon_rr > doc > wesri_v1.1_en.pdf	52,493	2013-01-08 12:40:37
maps > mon > mon_rr > doc > wesri_v1.1_ja.pdf	131,532	2013-01-08 12:40:37

一括ダウンロード

選択ファイル 0個 0バイト

ダウンロード 全チェック 全取消

<< 前へ ! 次へ >>

ダウンロード可能なファイル一覧

DIAS-PでのDOI付与計画

- 既に**データセット単位**でユニークな番号を付与済み⇒これをDOIに移行する。
- もともと**DataCiteに近いメタデータ形式**のため、JaLC形式への変換はおおむね可能。
- データセットの定義は**データ作成者の定義**に従うため、統一基準は定めていない。
- データセットの識別子=約200、データファイル=約4000万。**粒度のギャップ**。

DIAS-PからJaLCへの貢献

国立情報学研究所 北本様の意見とJaLC事務局回答

参考1

順番	JaLC項目	要素・属性	内容(質問・要望など)	回答(問題点・定義の整理)	回答案
1	10 25 35 51 62	sequence属性 (content/creator/contributor/affiliation_name 要素)	<input type="checkbox"/> 入力値について質問・要望 xs:string型で定義されており、備考欄に「同一 XML ファイル内でユニークにする」とあるが、ユニークであれば問題ないか? 数値として、順序関係は評価されないので? あるいは、もし順序の評価だけに使うのであれば、sequence 属性を使わずとも、タグの順番で代用できないか? 独自拡張の必須指定とするならば、定義と用途を明確にして頂きたい。	<input type="checkbox"/> JaLC 独自属性(必須指定) ・JaLC (research_data_v1.21.xsd) の定義 <xssattribute name="sequence" type="xs:string" use="required"/>	contentに関してはユニークな半角数字であれば問題ありません。 creator/contributor/affiliation_nameに関しては、筆頭にしたいものに“1”を、それ以降順番に“2”、“3”、…としてください。 XMLの中の記述方法として“2”的あとに“1”が来ても構いません(“1”を筆頭として登録します)。 現行の仕様は、ジャーナルアーティクルなど他のコンテンツと同様としたためこのようになっております。Sequence 属性を使わずタグの順番で代用する方式については、今後の検討課題とさせて下さい。
2.	11	doi 要素	<input type="checkbox"/> 表記ルールについて質問 DataCite では 10(ディレクトリサーバ番号)で開始される xs:token 型のパターンとして細かく定義されているが、JaLC では xs:string 型の自由記述となっている。 両者のマッピングに問題はないか?	<input type="checkbox"/> JaLC と DataCite で仕様が異なる ・DataCite 3.1 の定義 <xssimpleType name="doiType"> <restriction base="xs:token"> <xspattern value="10%.*\." /> </restriction> </xssimpleType>	マッピングは問題ありません。 JaLCでは、DOIプレフィックスがJaLCに登録されているものと一致しているかの確認を行った上で、DataCiteへテボジットしています。DOIのサフィックスは自由記述ですが、DataCiteで扱える文字は下記となっております。 0-9 a-z A-Z - (dash) . (dot) . (underscore) + (plus) : (colon) / (slash)

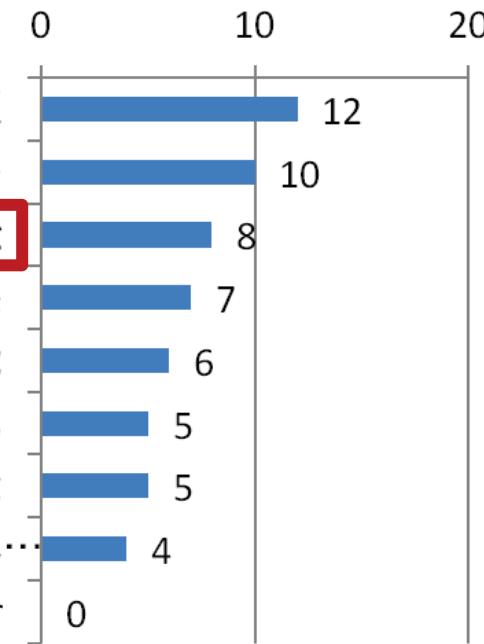
国際的な相互運用性の確保と、**日本語を扱うための拡張**を中心に、要望を提出。

JaLCへの期待

- データサイテーションへの研究者の期待は高い。DOI付与に加え、DOIの利活用（引用等）についても議論していきたい。

Q22. この条件を満たすデータ基盤に対しては、自分のデータを預けても良いと思う項目を選んで下さい（複数選択可）

- データの持続的な保存・管理に関する保証
データ提供プロセスに関する技術サポート
データ・サイテーション機能を備えていること
提供先からのアクセスログや利用者に関する情報提供
自分のデータに対して個別の利用規約を設定可能
データの喪失、毀損時の責任の明確化
ジャーナルとのリンク機能を備えていること
提供先が希望するデータ／メタデータ／サービス等の標準／規格に十分適合し…
Other



研究データ基盤に関する意識調査（2015年） 東京大学 小野雅史氏による中間分析結果



地球環境情報統融合プログラム

<http://www.editoria.u-tokyo.ac.jp/projects/dias/>