

Japan Open Science Summit 2018

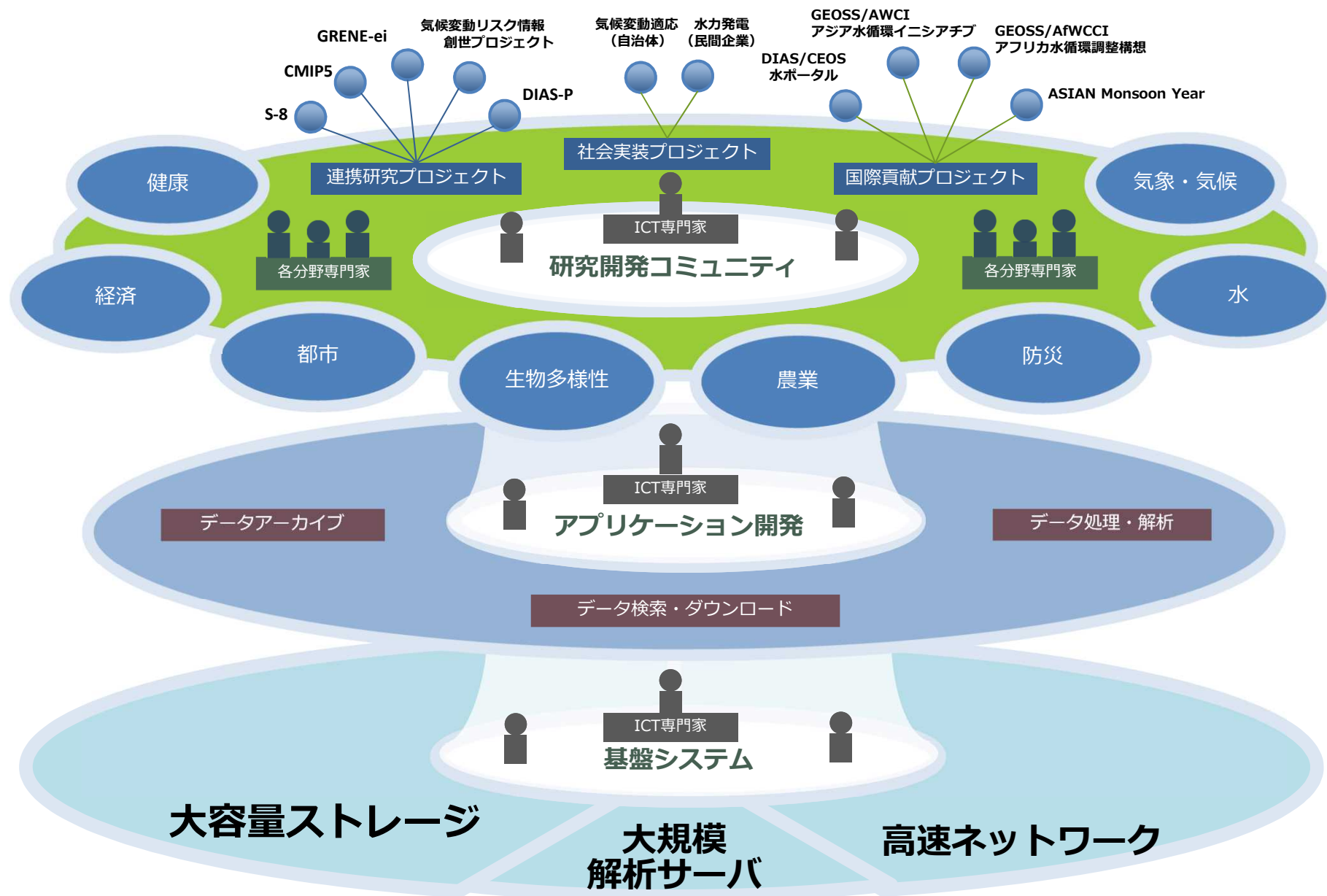
研究データ管理を考える
～DMPの実践に向けた対話～
(研究データ利活用協議会DMP小委員会)

データベース事業における事例について

絹谷弘子

東京大学 地球観測データ統融合連携研究機構

データ統合・解析システム

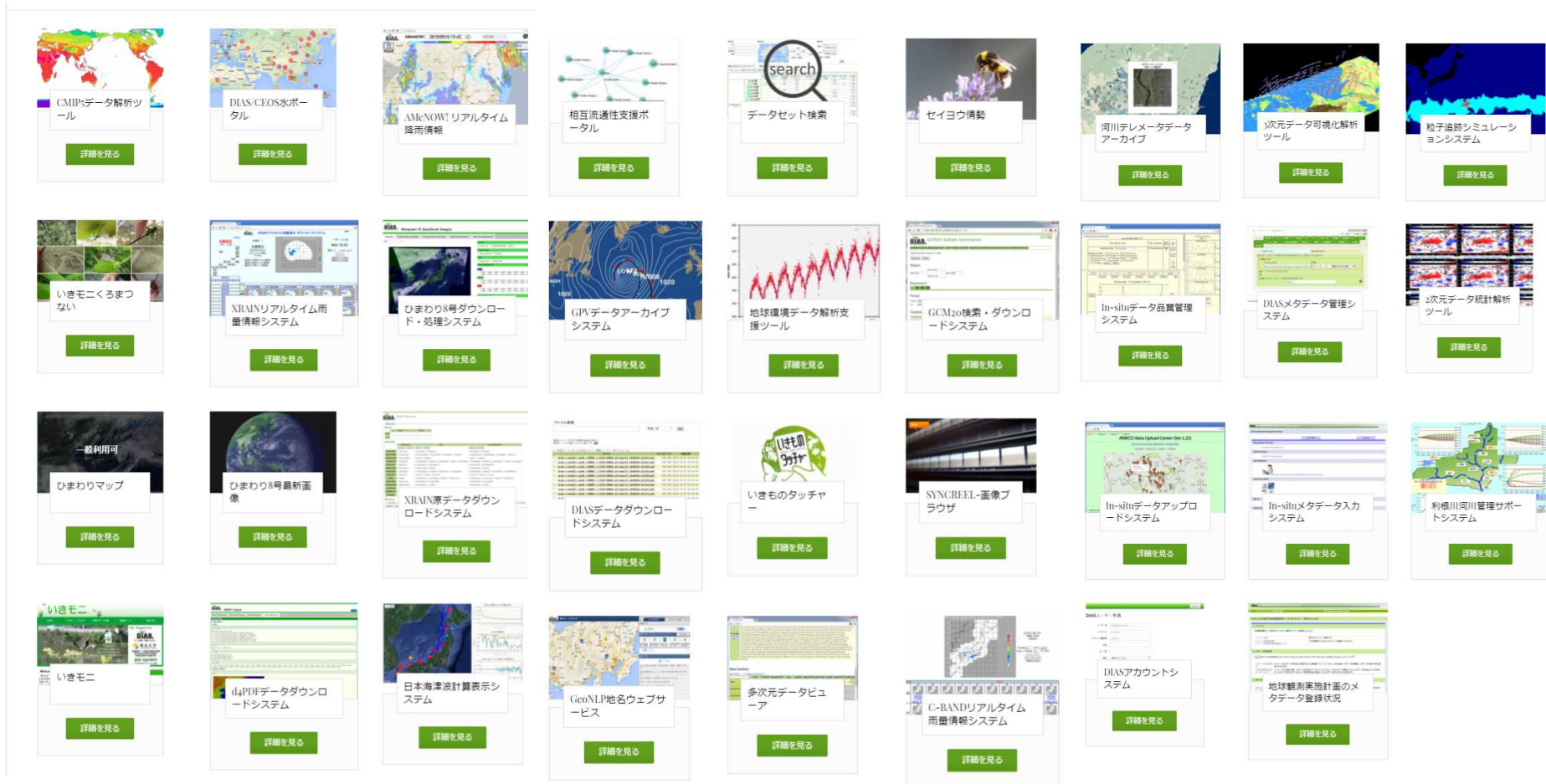


大気、陸域、海洋、人間圏などに関する多様な観測データや気候変動予測結果を格納。

種類	主なデータセット
衛星観測データ	TRMM衛星(降雨情報等)、AMSR/AMSR-E(水蒸気量、水温等)、MODIS(雲、エアロゾル、土地被覆、植生、地表温度等)、SSM/I(風速、土壌の含水率等)、SeaWiFS(クロロフィル分布、植生分布等)、だいち(ALOS)衛星(地形データ等)、しずく(GCOM-W1)衛星(降水量、水蒸気量、海洋上の風速や水温、陸域の水分量、積雪深度等)、LANDSAT衛星(地表面状況)、GRACE衛星(質量分布)、CZCS(海表面のクロロフィル濃度)、ひまわり8号衛星(リアルタイムデータ: 植生、雲、水蒸気量、SO ₂ 濃度、オゾン濃度等)
現場観測データ	国交省の河川テレメータ(降雨量、水位、水質、流速等)、Cバンドレーダー(広域リアルタイム雨量)、XバンドMPレーダー(詳細リアルタイム雨量)、アジア・アフリカ流域データほか
再解析データ	気象庁のJRA-25/55/JCDAS、NCEP/NCAR再解析データ、ECMWFのERA-Interim、JAMSTECの全球海洋再解析(4D-VAR)ほか
予測データ	気象庁の短期予測GVP、IPCCの長期予測CMIP3/CMIP5、JAMSTECの大気・海洋結合アンサンブル予測ほか
ダウンスケーリング	スクリプス海洋研究所の日本域ダウンスケーリング(解像度10km、NCEP/NCAR再解析データ及びRSM)、気象庁の日本域ダウンスケーリング(解像度20km及び5km、NHRCM)ほか
その他	日本全国都市粗度パラメータ、全球都市域マップ(ASTERの目視判読データ)、全球バイオマスデータ、大気汚染物質の排出インベントリ、津波アーカイブ、植生指数(NDVI)、水害統計データ、市民参加型データベースほか

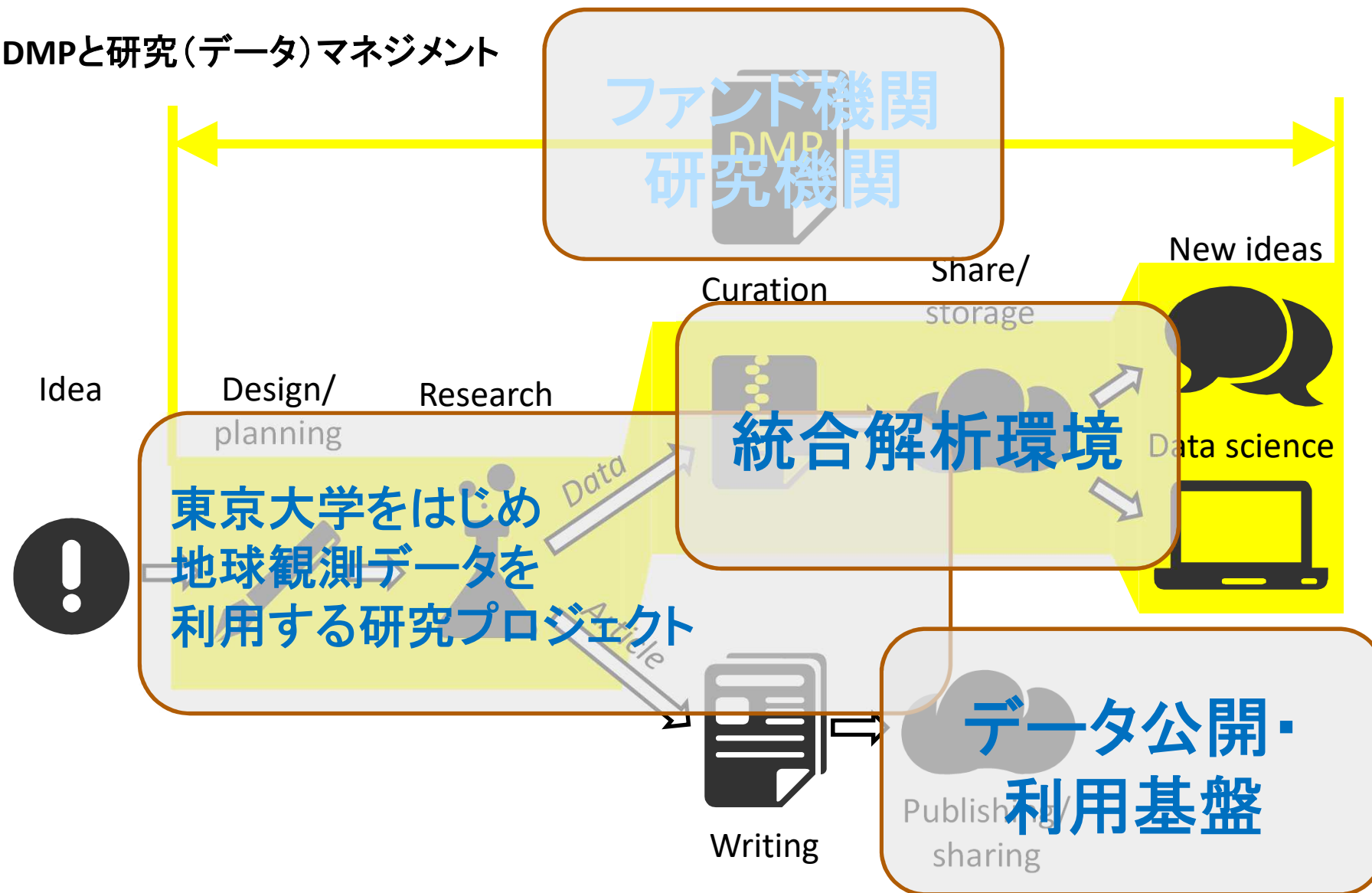
DIAS アプリケーション

www.diasjp.netより



DIASの位置づけ

DMPと研究(データ)マネジメント



リポジトリ

データ管理計画の立案には、データセットの最終的かつ長期的な格納場所、すなわちリポジトリを

選定することも必要です。リポジトリを選定する際、研究者は以下の点について考慮する必要があります。

- ・類似するデータセットはどこに保存されているのか？
- ・検討しているリポジトリのアクセスと利用に関するポリシーはどのようなものか？
- ・当該データをどの程度の期間、保存するのか？ どの程度の期間、保存しなければならないのか？
- ・リポジトリの管理者は？ 機関か？ 営利目的の提供者か？
- ・リポジトリの利用にかかる費用は？ その支払い方法は？
- ・複製、永続性、監視、災害復旧、事業の継続性に関するポリシーはあるか？

研究データに関して利用可能なリポジトリには主に2種類あります。学問分野を限定したリポジトリ(例、主題リポジトリ)と汎用リポジトリです。

≒ 研究データを公開しているデータベース

研究データ管理, Carly Strasser 著, 機関リポジトリ推進委員会訳, 2016 年3 月発行より

論文の査読において、**エディタと査読者がデータを閲覧する場合がある。**

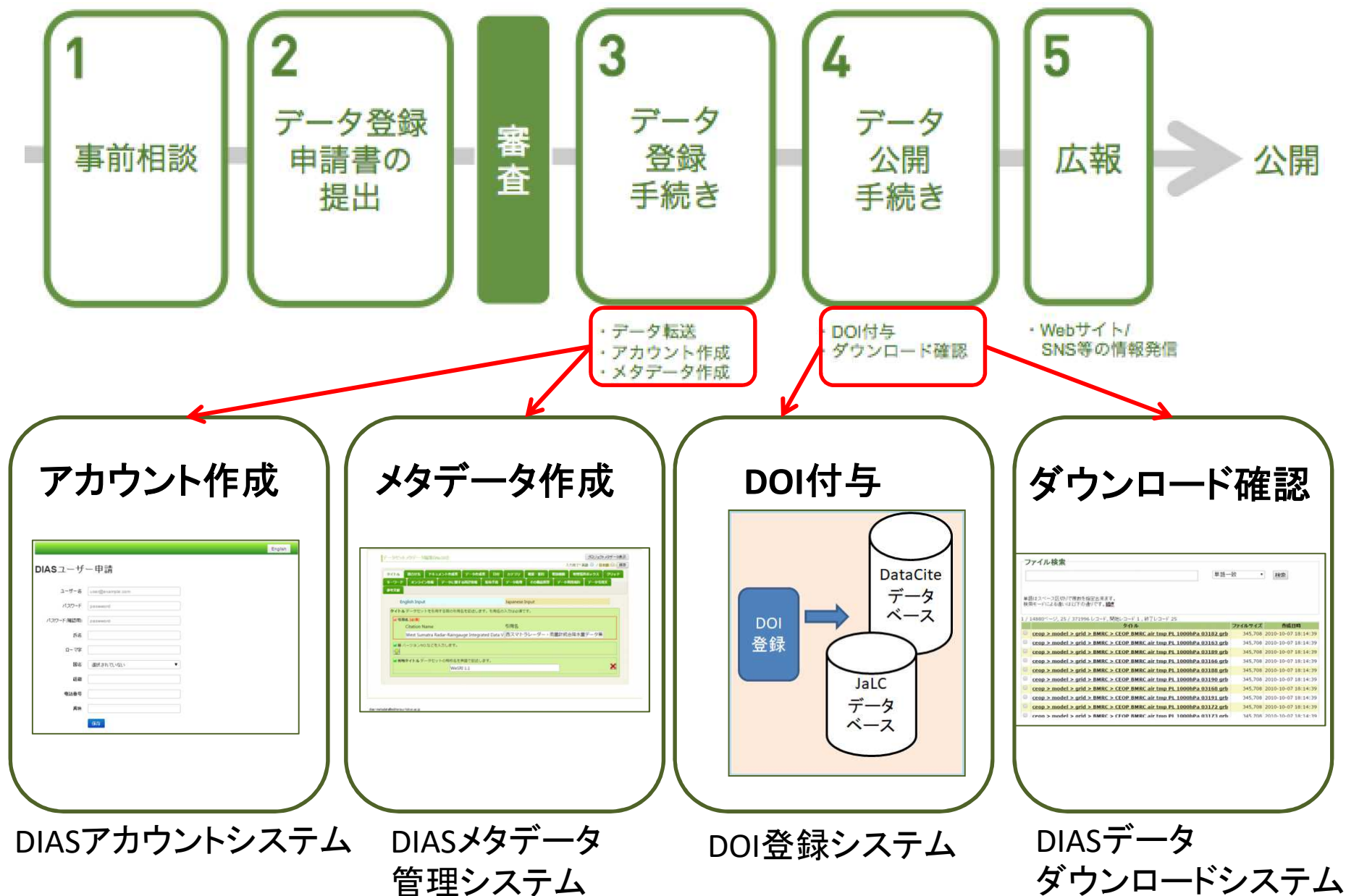
- 「新たなデータセットを開発した」と主張する論文や、人が「使いたい」と思うデータを解析した論文の場合、高い確率でそうなる。
- データが公開されるまで査読が完了しない場合や、論文が受理されても印刷に進まないことがある。
- エディタの判断による。エディタはジャーナルのデータポリシーを著者に守らせる立場にあるため。
- 一般に査読者が誰かは公表されないため、**論文投稿と同時に(あるいはそれに先立ち)、データを誰でも見られる状態にしておく必要がある。**

このため、**多くの場合、データをリポジトリに登録する。**

地球科学の場合、著名なレポジトリには・・・

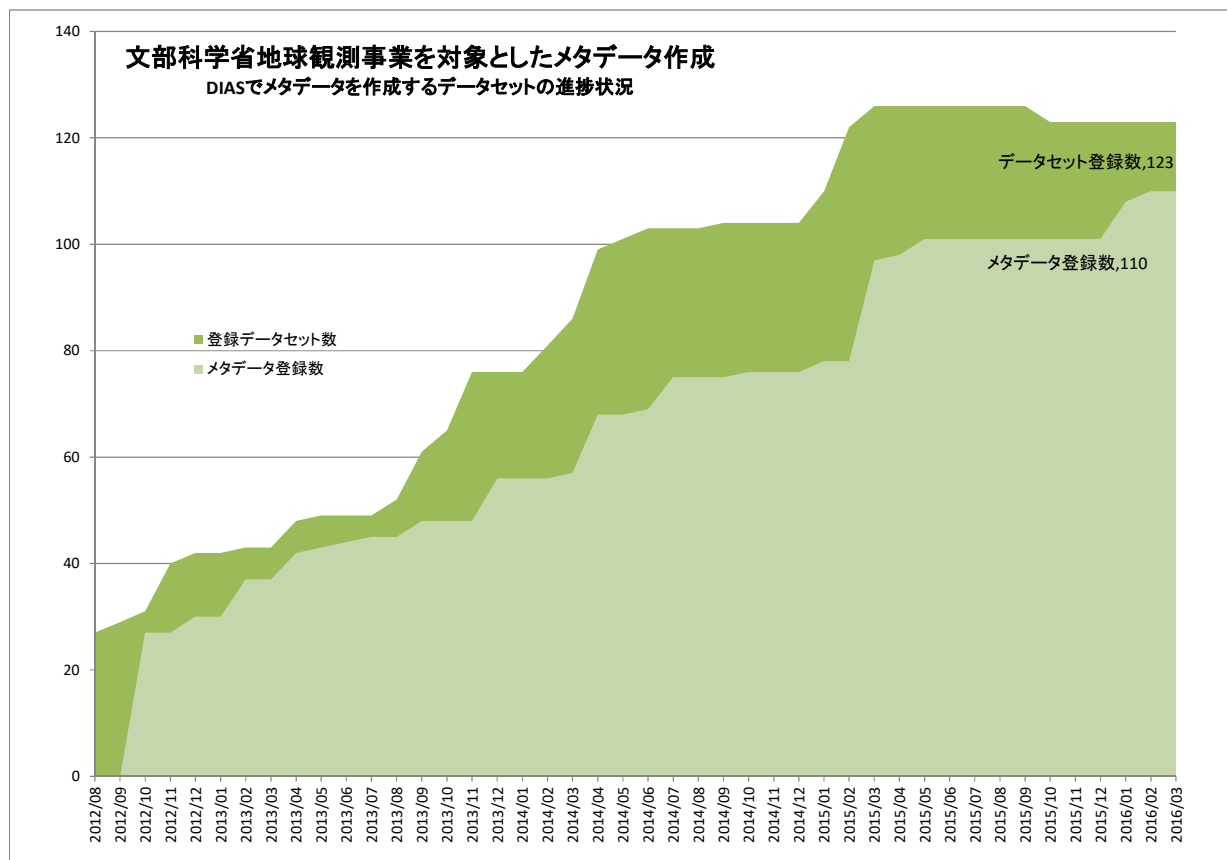
- PANGAEA: 無料。最大ファイルサイズは100MB
- Figshare: 無料のサービスは1GB(ファイルサイズは250MB)
- The World Data Center for Climate (DKRZ): モデル出力はこちらとPANGAEAから案内を受けた。
- COPDESS: Journal of Geophysical Research (JGR) では、地球科学のレポジトリとして最近これを挙げている。
- DIAS
- しかし、データ量が爆発的に増えている昨今、特に数値モデルの出力データを研究に用いる研究者には、**リポジトリから提供される容量が不十分な場合**が多い。

DIASデータ登録手続き <http://www.diasjp.net/data-provider/>



2011年度文部科学省地球観測の実施計画にリストされている事業を対照に
研究データの所在確認、メタデータ収集・メタデータ未作成データへのメタデータ作成を実施(2012年から2016年)

研究データの管理について対象事業に問合せてデータを把握する。
データの把握、メタデータ作成に開始から2年かかっている。



メタデータ作成は、研究者の知識が大前提。

登録できるデータが明確になるのに時間がかかる。

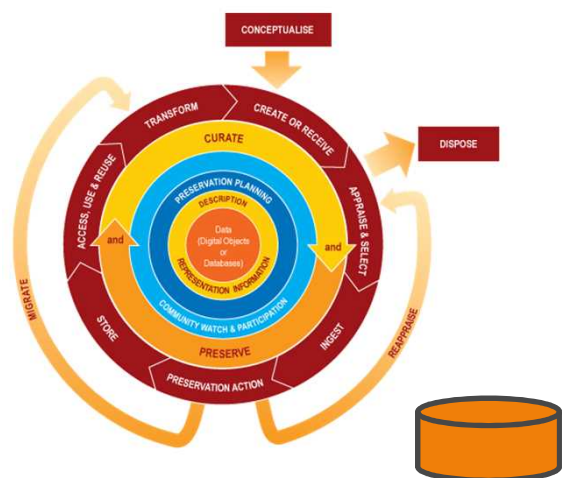
プロジェクトが解散すると連絡先が変わり、メタデータ作成者変更、追跡など支援が必要。

根気強く研究者と連絡をとらないと進まない。

- **公開・共有されるデータは常にアクセス可能で最新の問合せ先を維持する必要がある**
 - ・ 公開されているサーバが故障なく維持・管理されている
 - ・ データ利用者からの問合せへの対応のため、メタデータは最新の状態に更新されている
 - ・ データに誤りがあった場合、データ提供者に伝え、メタデータへの記載など対処方法を検討する
- **データの利用規約・アクセス規約の変更に対応する必要がある**
 - ・ 公開に関する規則が毎年変更となり、メタデータの文章変更をすべてのデータについて行ってきた。
- **データ提供者が省庁の場合、人事異動で担当者が頻繁に変更になる**
 - ・ たくさんのメタデータの維持管理を引き継ぐのが大変
 - ・ 連絡先が人事異動に影響されない書き方を推奨するようになった。
- **データ提供者が研究者の場合**
 - ・ 複数研究者の所属が異なっていて共同研究をしている場合、研究プロジェクト終了後は所属が変わって連絡が取れなくなる場合がある。

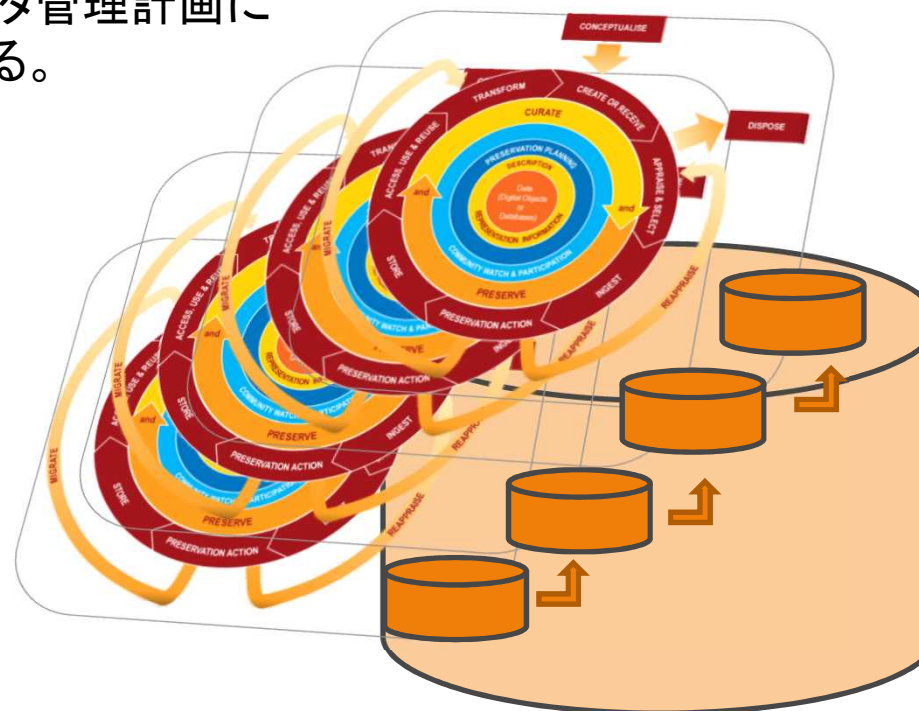
データベース運営の立場から

研究資金団体が要求するひとつの研究データ管理計画に加え
時間経過と共に継続する研究データ管理計画に
対応するモデルを考える必要がある。



1つの研究ライフサイクル

研究データ管理計画



同一テーマで継続する研究のライフサイクル

研究データ管理計画
研究データ管理計画
研究データ管理計画
研究データ管理計画

図 : DCC Curation Life Cycle model <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

Japan Open Science Summit 2018

研究データ管理を考える ～データリポジトリのサービスとCoreTrustSeal認証～

(研究データ利活用協議会リポジトリ小委員会)

14:15～15:45 中会議場A(2F)

研究データの保存・公開と信頼性の関係について、
関心のある方と共に議論したい。