

JAMSTECのデータポリシー、 データマネジメントとベストプラクティス



国立研究開発法人海洋研究開発機構
付加価値情報創生部門
地球情報基盤センター
データ管理グループ

福田 和代

本日の話題

- JAMSTECにおけるデータポリシーと規程類
- 情報管理部署によるデータマネジメント
- 最先端の研究者によるデータマネジメント

海洋を軸とした地球環境全体、すなわち生命活動をも含めた地球を構成する複雑かつ多様なシステムを「海洋・地球・生命」として一体的に捉え、それらシステム
の行く末に大きな影響を及ぼす人間活動との相互影響を含めた統合的な理解を
推進し、科学的知見を有用な情報として発信していくことにより、人類社会が地球
の未来を創造していくことに貢献する。
(第4期中長期計画 2019-2025年度)

データ サンプル 調査観測データ、シミュレーションデータ、画像等
生物、堆積物、岩石、海水等の標本

キーワード 地球科学 > 海洋、古気候、固体地球、生物学的分類、生物圏
雪氷圏、大気、地表、陸水圏

施設
・
設備



「データ・サンプルの取り扱いに関する基本方針」

2007年5月制定 JAMSTEC全体のデータ・サンプルに適用

主な内容

- データ・サンプルは**人類共有の財産**
 - 研究・教育などの利用のために**広く公開**され、**将来にわたって世界中で活用**されることが重要
 - データ・サンプルは、原則として**機構に帰属**
 - 機構は保管、**公開**、**提供**を行う
 - **知的財産として保護**すべきものについては、機構の「知的財産に関する基本方針」にもとづき、データ・サンプルの積極的活用・適切な管理に努める
 - データを取得した研究者に**公開猶予期間**を与える
 - データは機構のウェブサイトにて**公開**
 - 科学的・教育目的利用の場合は**実費を除き無償**
 - 産業利用については原則として**有償**
-

関連規程類の整備

- 研究者・技術者で構成されるWGにおいて議論のうえ、データ・サンプルの取得方法で体系化した規程類を順次整備

データ・サンプルの種類によって管理・公開担当が異なる

	データ 取得	メタデータ 管理/保存/公開	データ 管理/保存	データ 公開
航海調査 ※1	研究者	情報管理部署	情報管理部署	情報管理部署
機構外船舶 係留系、陸域調査	研究者	情報管理部署 ※2, 3	研究者	研究者 ※2
シミュレーション	研究者	所属部署	研究者/所属部署 ※4	研究者
実験データ	研究者	所属部署	研究者	研究者

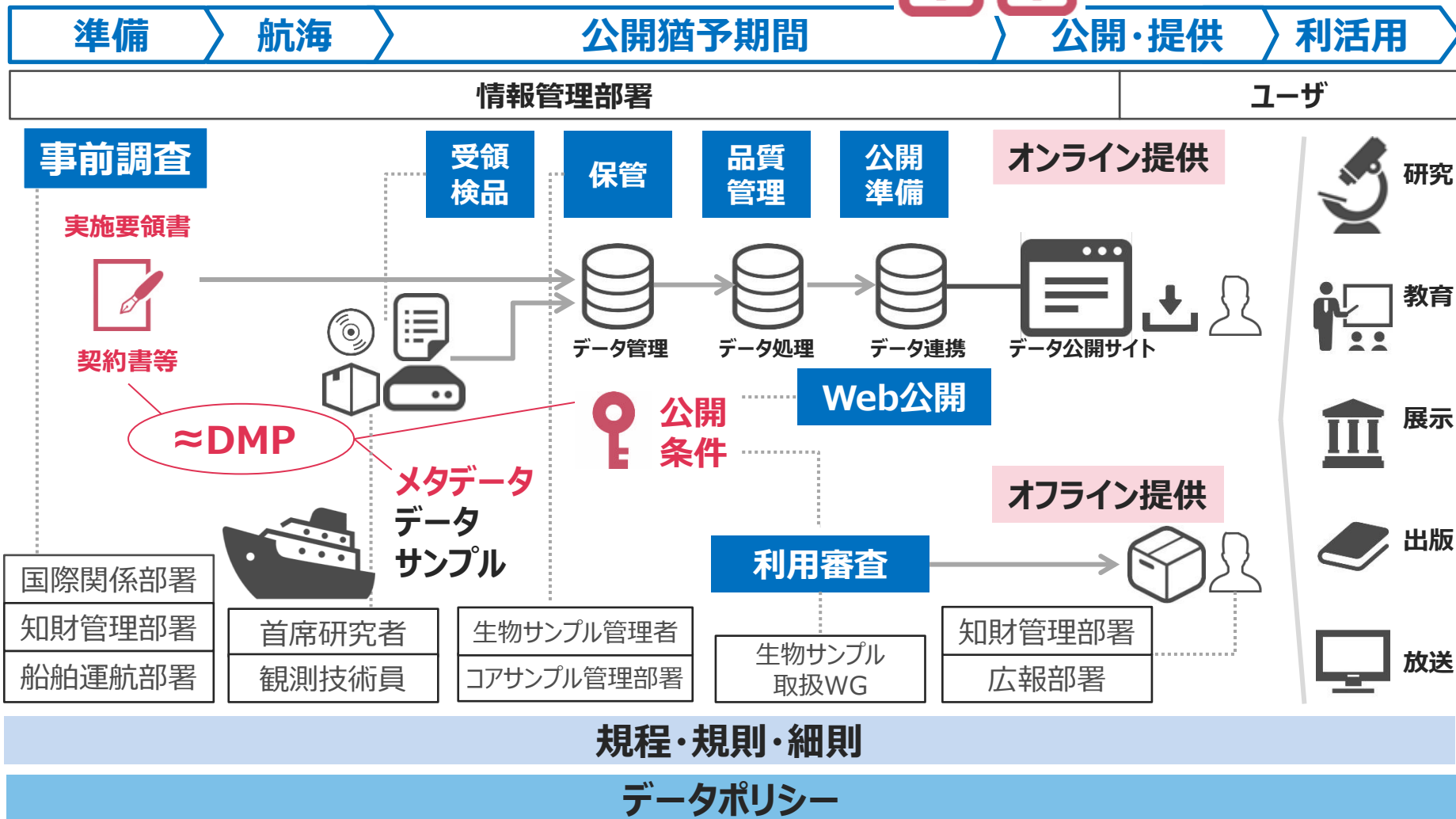
※1：JAMSTEC船舶による航海（「ちきゅう」のJAMSTEC独自航海含む）今年度より、学術船による共同研究航海も含まれる。

※2：情報管理部署提供のシステム（観測、データカタログなど）による公開を含む。

※3：情報管理部署によるメタデータ公開は、陸域調査、係留系など、一部実施。

※4：公開されるデータのみ、所属部署が管理保管を行うと明示されている。

データマネジメントは出港前から開始される



公開条件はメタデータ・データの受領状況とともにデータベースで管理

機構内ルール

- ・ 規程類で決められた公開猶予期間

機構外ルール

- ・ 国連海洋法条約
- ・ 海洋保護区
- ・ 特別採捕
- ・ 鉱物探査
- ・ 遺伝資源へのアクセスと利益配分 (ABS)
- ・ 機関間契約 (IA, MOU)

特別な契約

- ・ データの取り扱いに関する特別な契約
- ・ 外部資金による研究計画

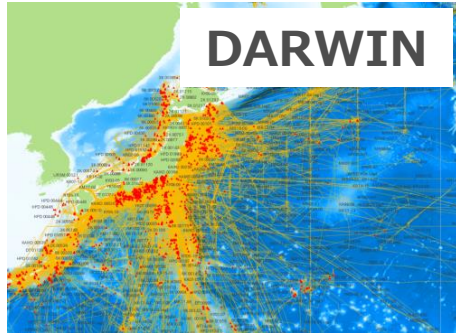
各航海の
MSR申請
海洋保護区での実施許可
特別採捕許可
ABS申請
IA,MOU等の記載事項を確認



©JAMSTEC/NHK

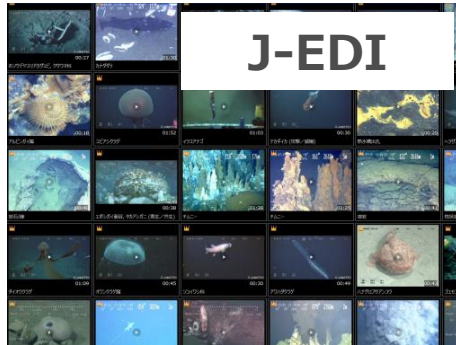
例：映像・画像の著作権表示

公開提供 1980年代以降の航海・潜航に関するメタデータ・データファイルを**WEB公開**
生データ、高解像度映像、サンプルは**オフラインで提供・貸出**



航海	1,929件
潜航	5,961件
観測データ	13,914件
サンプル	69,384件

(DARWIN 2019-01-30)



映像総公開時間	約38,700時間
画像総公開数	約159万枚

(J-EDI 2019-01-29)

利活用 科学的・教育的利用

データ解析、サンプル分析、学会発表、論文執筆、航海計画策定、博物館展示、教材、...

産業利用

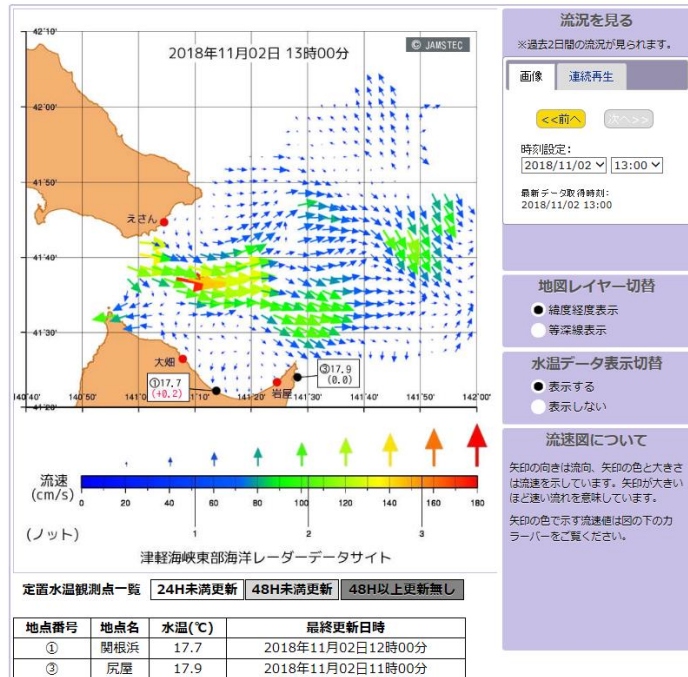
放送、出版、...

表面流マップの準リアルタイム公開・・・漁業者にも有用な情報提供



津軽海峡東部海洋レーダーデータサイト

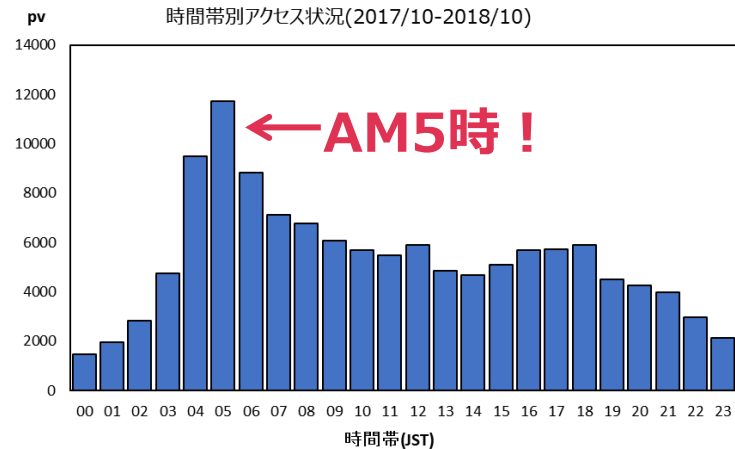
MIO Ocean Radar data Site for the Eastern Tsugaru Strait (MORSETS)
<http://www.godac.jamstec.go.jp/morsets/j/top/>



表面流マップは、研究データという位置づけのみならず、おそらく漁業者や現場を航行する船舶などにとっても有用な情報。



Web公開に対する反応

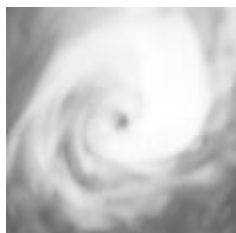


オンラインコンペティション開催による熱帯低気圧・タマゴ検出の精度向上

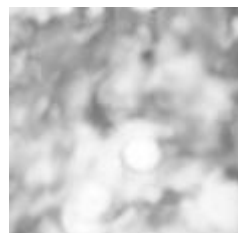
概要：

雲解像モデルNICAMのデータにラベル（熱帯低気圧・タマゴ、それ以外）を付与して公開し、分類精度の高さを競うコンペティション形式でのイベントを開催した。

正例と負例の画像数が大きく異なる場合（不均衡データ）のクラス分類



熱帯低気圧・タマゴ
(正例：約7万枚)



それ以外
(負例：約217万枚)

開催期間：2018年8月28日～10月26日

(入賞者らによる報告会を2019年2月20日に開催)

データ提供：NICAMチーム

(海洋研究開発機構、理化学研究所、東京大学)

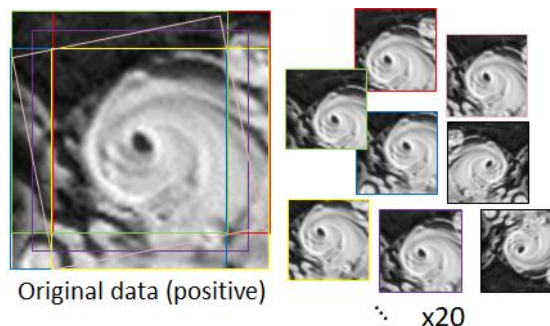
スポンサー企業 (入賞者への賞品提供)：

NVIDIA、G-DEP、アーク情報システム、CRAY

主催：海洋研究開発機構

効果：

- 2ヶ月間の開催期間において、210名（3293件の投稿）の参加があり、JAMSTECのデータへの興味の高さがうかがえた。
- 上位入賞者の分類学習モデルでは、従来手法と比べて空振り率を約70%から30%程度へと大幅に低減させることに成功。



学習データの選択手法やディプラーニングの層設定の最適化等が提案

(気象学としてはタブーと思われる) 台風画像の回転等によって、学習データ数を増やす等の技術も



<https://signate.jp/competitions/134>

データポリシー制定から10年以上経過

- データ・サンプルをとりまく状況に応じて、関連規程類・運用フローを整備
 - 情報管理部署による航海調査のデータ・サンプルの公開・提供
 - 研究者による新しいタイプのデータ公開
- 様々な目的のデータ・サンプル利活用
- 研究の発展、 研究分野の裾野拡大

課題

- 観測機器やモデルの高精度化によるデータ容量の増加
- オープンサイエンスへの対応
 - 論文に紐づくデータの公開
 - データ・サンプルへの識別子付与
- 組織改変やプロジェクト終了に伴うデータ公開の継続
-

これからのデータマネジメントとは？ ベストプラクティスとは？

参考URL

スライド3

データ・サンプルの取り扱いに関する基本方針

http://www.jamstec.go.jp/j/database/data_policy.html

スライド6

観測メタデータ管理システム

<https://doi.org/10.5918/jamstecr.18.53>

スライド7

航海・潜航データ・サンプル探索システム (DARWIN)

<http://www.godac.jamstec.go.jp/darwin/j>

深海映像・画像アーカイブス (J-EDI)

<http://www.godac.jamstec.go.jp/jedi/j/>