

第2回オープンサイエンスデータ推進ワークショップ
12月7日@京都大学理学研究科セミナーハウス

産業技術総合研究所における オープンデータの現状と方向

産業技術総合研究所
情報・人間工学領域
研究戦略部
小島 功

lsao.kojima@aist.go.jp

概要

● 産総研の研究データ公開の現状

- ▶ サービスとしての提供：民間サービス等との関連

● 2次利用促進のための試み

- ▶ 情報技術(IT)の研究テーマ関連

● doi付与テスト参加

- ▶ 課題の抽出

● 議論

- ▶ 研究活動がクラウド化・統合化→公的セクタの位置？
- ▶ オープン化・標準化・ボトムアップ？

**産総研＝議論中
(というか決まっていない)
ぜひ課題の共有と議論が
できれば、、**

産総研研究データ公開の現状

産総研における研究(成果)データベース

● 原則

- ▶ 広くオープンデータとして成果公開
- ▶ インターネットでサービスとして提供 ★★★★★以上
- ▶ 無償（権利関係や予算、作成の経緯等から制限されるもの有）
- ▶ ライセンス：個々のデータにより異なるが、おおむねCreative Commons(CC-BYなど)に収束の方向性

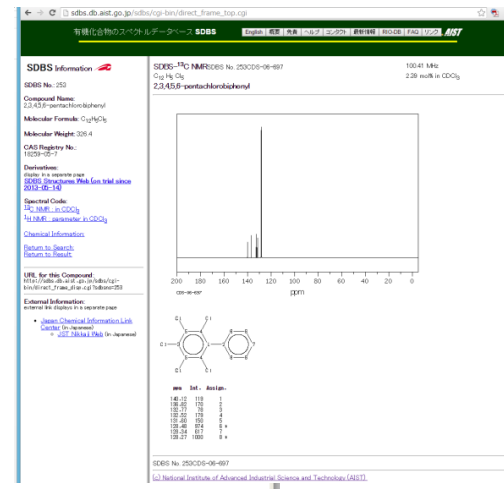
● 概要：

- ▶ 研究成果DB（広報等で把握されて紹介されているもの等）
 - ◎ http://www.aist.go.jp/aist_j/aist_repository/riodb/index.html
 - ◎ 6分野、50個以上
- ▶ その他研究プロジェクトの成果として公開されている物多数
 - ◎ 衛星アーカイブ、バイオ統合DBで構築提供されているもの等
 - ◎ 機関レポジトリ、業務系のオープンデータ（経産省本省に目録を提供）を合わせると300を超える？
- ▶ 個々のデータベースは、地質系（地質図等）や物質系（SDBS等）、ライフサイエンス（糖鎖等）などで高い評価を得ている

有機化合物のスペクトルデータベース (計測標準研究部門)

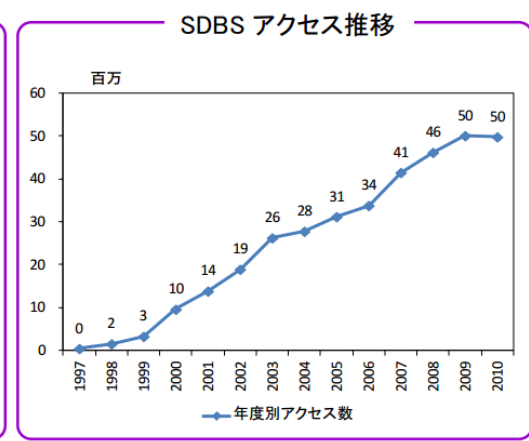
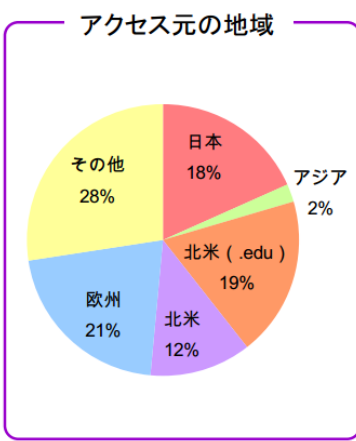
(産総研で最も有名でかつ歴史のあるデータベース)

- 1997年からWeb公開 <http://sdfs.db.aist.go.jp>
 - ▶ ○高い評価と利用の拡大
 - ▶ ○継続的な整備改良で最新データを使いやすく提供
 - ▶ ✕コンテンツ維持・更新に大きな手間
 - ▶ ?労力に見合った評価を得ているか？



Integrated Spectral Data Base System for Organic Compounds

- 年間アクセス数 H22年度 **5千万**件
- 産総研の研究情報公開DB (RIO-DB) アクセス数 **No1**
- RIO-DBのアクセスに占める割合 H22年度 **87%**
- スペクトル新規公開件数1000件/年
測定・評価後に追加:職員等 5名体制



化学分析、大学教育、材料研究等に幅広く利用

SDBS自体の知名度が上がっていることに加え、繰り返し利用している研究者等 (リピーター) が増えていることにより、毎年アクセス数が増加

地質情報データベース(地質調査総合センター)

<http://www.gsj.jp/researches/geodb/>

- オンライン化した地質図、衛星データ(画像)、岩石や断層などの情報、地質調査結果や関連文献など
- ○ 高機能: 複数のデータベースに対し、地図上で統合的に検索・表示することが可能
- ○ 地質の業務としてデータベース整備が可能
- × 他の分野とのデータベースの連携等に課題

The image displays a composite of three screenshots related to the Geological Survey of Japan's (GSJ) geology information database.

- Top Left:** A screenshot of the main website <https://www.gsj.jp/researches/geodb/index.html#category2>. The page features a navigation menu with categories like 'Geology Information Database', 'Geology Information Service', and 'GSJ Links'. A central section titled '地質情報データベース (GEO-DB)' provides an overview of the database's research results and data collection.
- Top Right:** A screenshot of the '地質図Navi' (Geology Navigator) application interface. It shows a satellite map of a region with a red rectangular area of interest. The interface includes search fields, position coordinates, and a list of search results.
- Bottom:** A larger screenshot of the '地質図Navi' application showing a detailed geological map of a coastal area. The map is overlaid with various data layers, including geological structures, topography, and land use. A sidebar on the left contains a list of data layers and search options, while a right sidebar shows detailed information for the selected layer.

産総研における研究(成果)データの位置

- 「産業技術」総合研究所＝産業/ビジネスへの貢献が期待されている
 - ▶ ○データを利用した新産業などの創出への貢献(二次利用の促進)
 - ▶ ×一部データの有料化やデータ公開自身をビジネス化する等の課題
- 多くの分野をカバーする単一の研究機関
 - ▶ ○分野横断的なデータ連携等が容易
 - ▶ ×個々の研究分野の独自性と全体としての統合の両立
- 標準活動が重要な業務の一つ（JIS、計量標準等）
 - ▶ ○標準化に積極的に貢献することでの優位性の実現
 - ▶ ×標準化の時間や手間が増えて肝心のサービス化や利用促進が遅れかねない

データの公開と二次利用

公開にかかる現状

- 多くの公的(研究)機関が研究データを公開
 - ◎ WDS(World Data System)
 - ◎ DIAS
 - ◎ GEO Grid/産総研地質情報データベース
 - ▶ 殆どが無料で公開
- データを置くだけでなく、検索等サービスとして実現・提供
 - ◎ データ管理だけでなくサービスの維持費用がかかっている。
 - ✦ データは公的資金による研究成果→サービスの維持費用は？
 - ▶ サービスで提供する機能や互換性の問題
 - ◎ 単に公開だけでは良くない→オープンデータの5つ星
 - ✦ しかし高度化には費用がかかる
 - Linked Data化を誰がするか？
 - ◎ サービスの互換性
 - ✦ データに互換性があってもサービスは互換性がない場合が
 - フォーマットは同じだが検索の仕方が違う、、等

企業によるオープンデータの無料公開

- 公的データ（でなくても）独自にサービス化して無料提供
 - ▶ Amazon Public Data Set
 - ◎ NASA modis/landsatなどが自由に使える。
 - ▶ Google Skybox for Good
 - ◎ 有料衛星データの一部を無料で提供
- 特徴：データ提供とは別に費用を回収する仕組みがある
 - ▶ フリーミアム、広告
 - ▶ 他の有料サービスと連携
- 公的機関(大学・研究所)の役割
 - ▶ **データ提供者：公的資金の研究成果のオープンデータ化**
 - ▶ **サービス提供者：民間と競争になっていく**
 - ◎ むしろ民間に任せるべき？(市場化テスト)
 - ◎ **どこが切れ目か？**等を考える間もなく既に競争？

IT（＝サービス実装の技術）の研究開発方針

● ビジネス創出への貢献

- ▶ ★ データを使ったサービスの高度化や構築支援
- ▶ 応用研究者の支援
- ▶ クラウド的な運用サービスの提供
- ▶ ★ 既存のクラウドサービスとの連携

● 分野横断的なデータ連携：ボトムアップなデータ・サービス連携

- ▶ 適切な標準規格の採用による相互互換性の実現
- ▶ デファクトスタンダード（CKAN等）の採用による一覧性・到達性向上

● 標準化:ベンダーロックインを避ける

- ▶ ★ 個別分野の標準と全体としての連携・一貫性の実現
- ▶ 地理空間におけるOGC標準と、分野横断標準のLinkedData等の整合

● データ生成者との連携：ニーズとシーズのミスマッチを避ける

- ▶ ★ doi試行プロジェクトなど協業を通じた解消
- ▶ LODチャレンジ等による技術ショーケースの実現とアピール

二次利用の促進&ビジネス創出

データの二次利用：✕ データの検索、ダウンロード、引用
○ データを加工して付加価値を創出

🌐 技術ボトルネック：

- ▶ 分散・大量データから着目すべき対象のデータを発見する
 - 🌐 => 分散問い合わせなどの検索技術（ワーキングセットの発見と抽出）
- ▶ 大規模のデータを効率的に解析する
 - 🌐 => 大規模機械学習技術
- ▶ データ解析など付加価値サービスの構築を容易にする
 - 🌐 => ワークフローやラピッドプロトタイピング

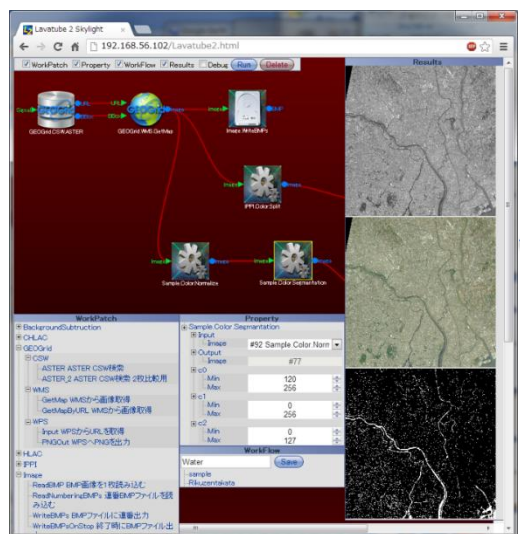
🌐 研究開発：

- ▶ 分散検索：分散SPARQL検索技術 Aderis-Hybrid
 - 🌐 <http://aderis.linkedopendata.net>
- ▶ 大規模機械学習技術：Hivemall
 - 🌐 <https://github.com/myui/hivemall>
- ▶ ワークフロー：Lavatube

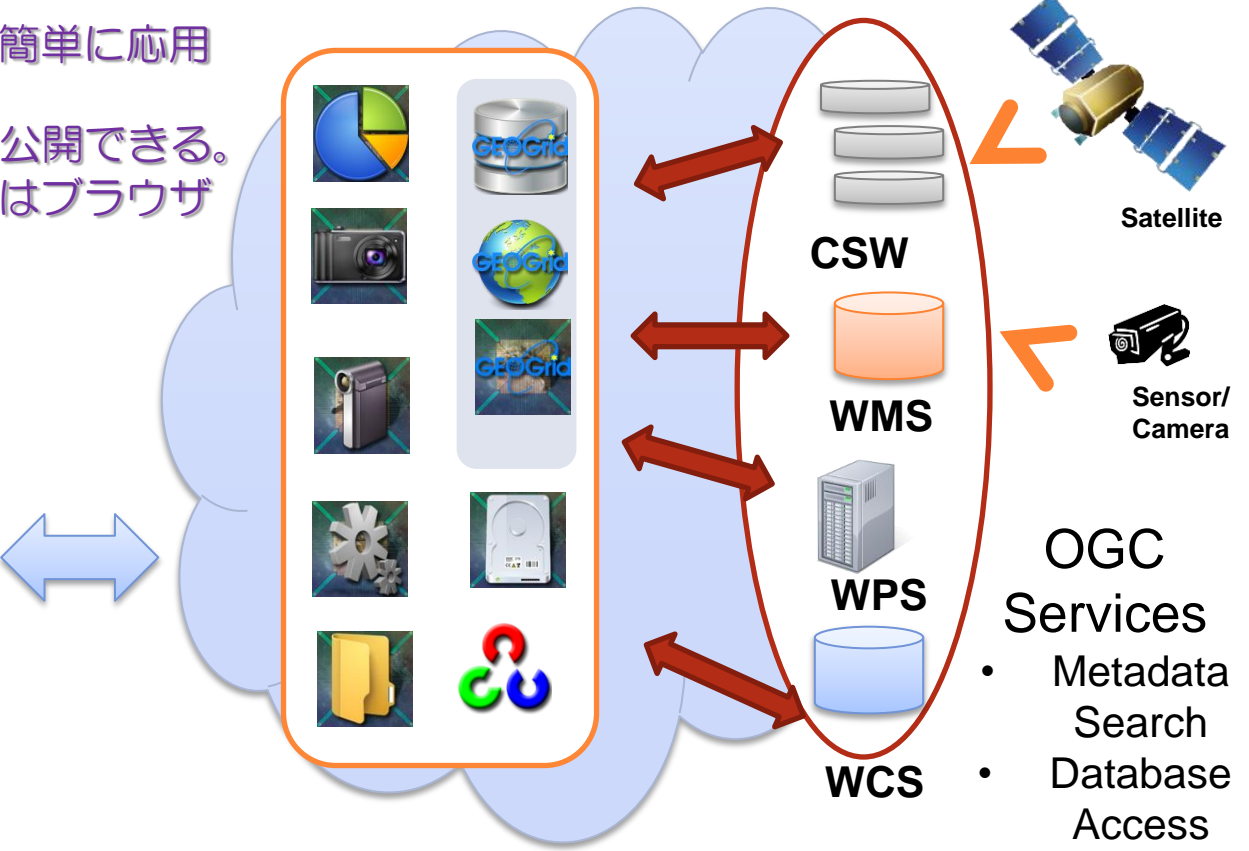
1:サービス構築・提供の容易化：ワークフロー構築サービスの提供: Lavatube

- 公的データのアクセス基盤と、それを使った解析プラットフォームのクラウド提供
- 事業者はこの上で独自の付加価値処理を構築して商用化
- 試行錯誤的なデータ検索や、アプリケーションのラピッドプロトタイピングが重要

1. 既存の公開データを使って簡単に応用を構築できる
2. 作った物をサービスとして公開できる。
3. 処理はクラウド上で、操作はブラウザ上で



Browser Interface(HTML5)



Workflow Parts

- Metadata Search
- Database Access
- Processing

研究課題

対話性：サーバで進行中の画像処理をクライアントで表示させたい。

- 多数のビデオ画像を並列に転送するのが難しい
 - ▶ →WebSocketの効率的利用

高速化：ワークフローの処理の間で動画像を受け渡す必要がある

- 普通に渡すとリアルタイム処理が難しい
 - ▶ →ストリーム処理の実装を工夫

精度：2時期衛星画像の正確な位置合わせ

- ▶ →色、位置、範囲のずれなどの吸収

事例: 変化検出 ⇒クラウドでの提供と事業化FS

(BtoBtoC)

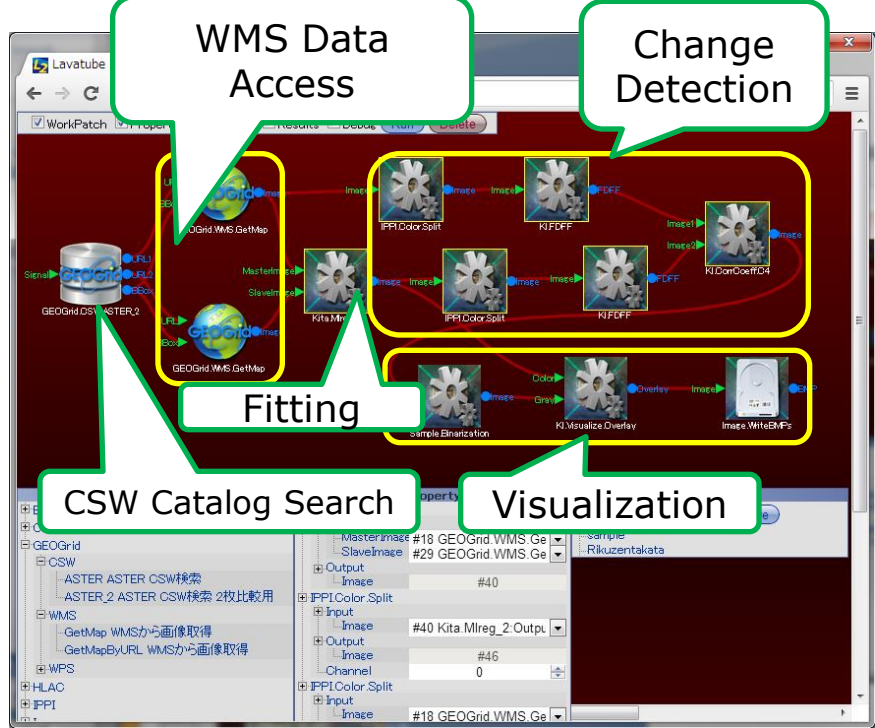
事業化に向けての課題を研究開発中

1. 特定領域をメタデータ検索 (OGC CS-W)
2. 2 時期の画像データの取得 (OGC WMS)
3. 位置合わせ、色合わせ等の処理
4. 変化検出
5. 可視化 (browser / google earth)



陸前高田

仮設住宅用地等の造成



Workflow constructed with Lavatube



2011.06



2012.06



Result

防波堤の復旧

2：標準規格に基づくデータ統合とその高度化

- 福島原発事故に関わる放射線データの測定とDB化（原研）
 - ▶ 土壌モニタリング、航空機によるモニタリング、カーサーベイ、(リアルタイム)モニタリングなど多種多様なデータを整備
 - ▶ インターネットでダウンロード可能(CSV, XML形式)
- 国際標準に基づいたデータベース連携環境の実現（産総研）
 - ▶ Google Mapなどで表示するための国際標準API提供
 - ▶ 高度な解析を行うためのデータ連携フレームワーク構築

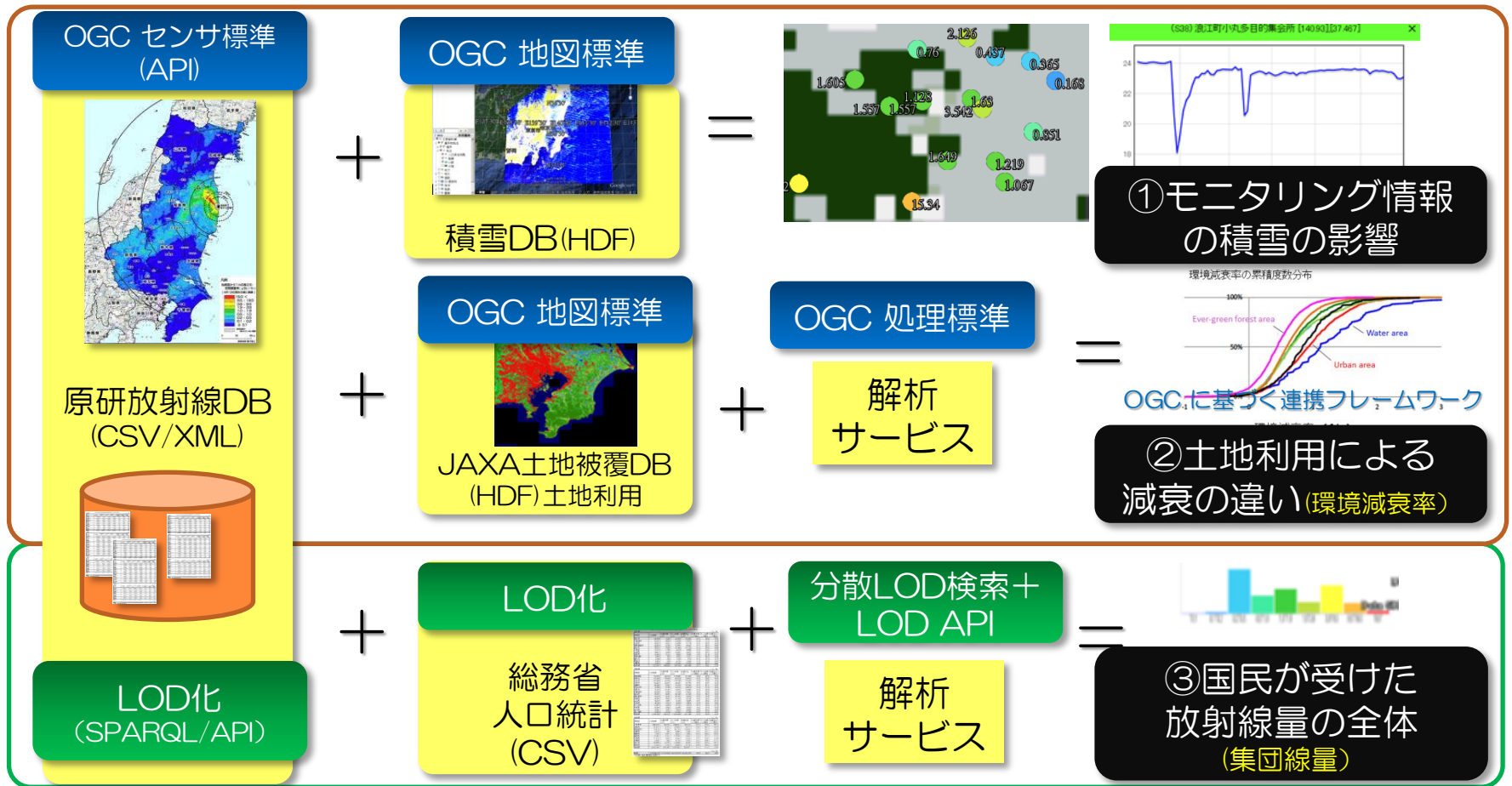


(※) 原子力規制委員会「東京電力福島第一原子力発電所事故による環境モニタリング等データベースの構築事業」
 (独) 日本原子力研究開発機構（原研）より再委託として実施 (H25)
 引き続き原研からの委託事業として推進中 (H26,27)

産総研が実現するオープンアクセス基盤



- 多種多様な科学技術データを連携できる基盤
- データ解釈(解析・応用)プログラムの構築が容易



研究課題

データ統合：異種のCSVファイルのデータ統合

● ボトムアップに作ったので属性の名前がバラバラ

- ▶ →自動的に推論して統合先を示唆
- ▶ →人が確認してデータ統合
- ▶ →その結果を知識ベース化して後のファイルに適用

高速化：データ検索と表示の高速化

● 標準規格(OGC)のアクセス手法をそのまま実装すると検索/表示が遅くて使えない.

- ▶ → 検索エンジンの技術を使って高速化
- ▶ → 事前に検索結果をあらかじめ計算、準備して応答を高速化
- ▶ → 標準形式やプロトコルを工夫

3 : LODの分散検索 (分散SPARQL) : ADERIS-Hybrid

オープンデータの分散検索を可能にする技術

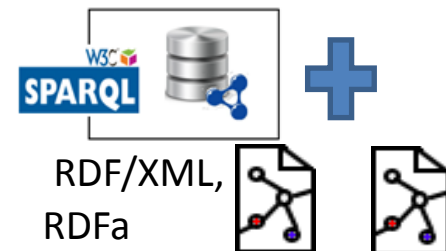
- 標準問い合わせ言語SPARQLを拡張することで、ユーザが検索処理の「制限時間」を設定可能に
- 指定した制限時間内で、できるだけ多くの結果を返す「ベストエフォート」型の検索
- 検索結果とともに信頼性指標を算出して返すことで、検索結果の評価を可能に
 - 最終的な結果個数予測
 - それに対する信頼性 (自信度)
 - 得られた答えの網羅性や鮮度

開発したソフトウェアをサービスとして一般公開
<http://aderis.linkedopendata.net/>

- 各国 (日本政府を含む) で推進されるオープンデータを利用した統合的な検索システムの実現
- さまざまな研究分野のデータベースをつなぐ知識探索アプリケーションの実現
- オープンデータサービスのSI事業等への技術移転

Hybrid

- Increased coverage
- Fresh results
- Diverse results



Best-effort

- Time is important (complete answering over the Semantic Web is difficult, unpredictable)

```
SELECT DISTINCT *
WHERE {
  ?paper <http://data.
  <http://data.sem
  ?paper <http://swr
  ?p_rdfs:label ?n
}
```

WITHIN 5 seconds

Specified when writing SPARQL query

4 : Hivemall:簡便でスケーラブルな機械学習基盤

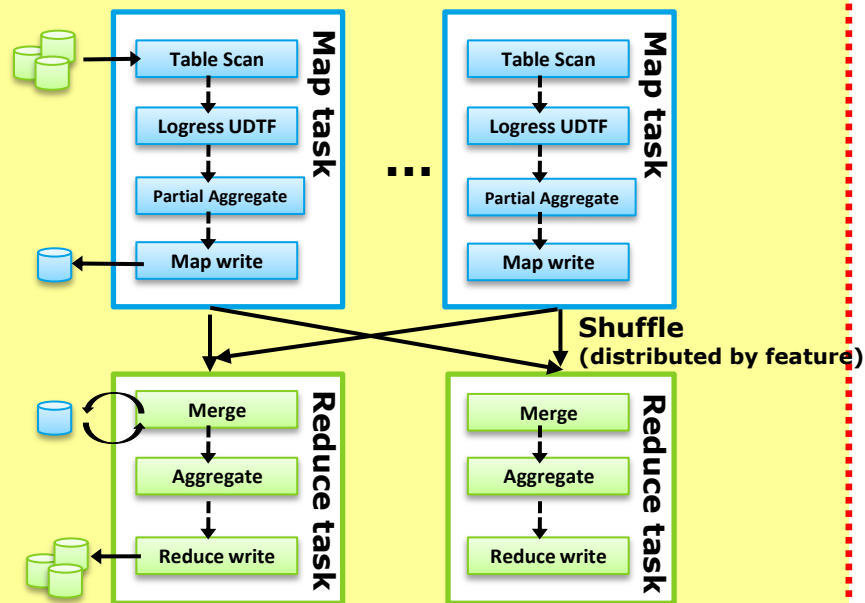
Apache Hadoop/Hive上で動作する、プログラミングフリーな機械学習基盤

- 簡便な問い合わせ言語HiveQLだけで機械学習のすべての手続きが完結
- データ数、特徴の次元数、学習モデルのサイズ、計算資源に対してスケーラブル
- 最新のオンライン機械学習アルゴリズムをサポート
- Amazon EMRを始めとする商用クラウドサービスでも利用可能

- 産業界と連携し、マーケティング、Webをはじめとする様々な応用に展開
- ビッグデータ処理基盤のデファクト標準への貢献

```
SELECT
  feature, avg(weight) as weight
FROM (
  SELECT logress(features, label, ...) as (feature, weight)
  FROM train
) t
GROUP BY feature;
```

ロジスティック回帰の学習例



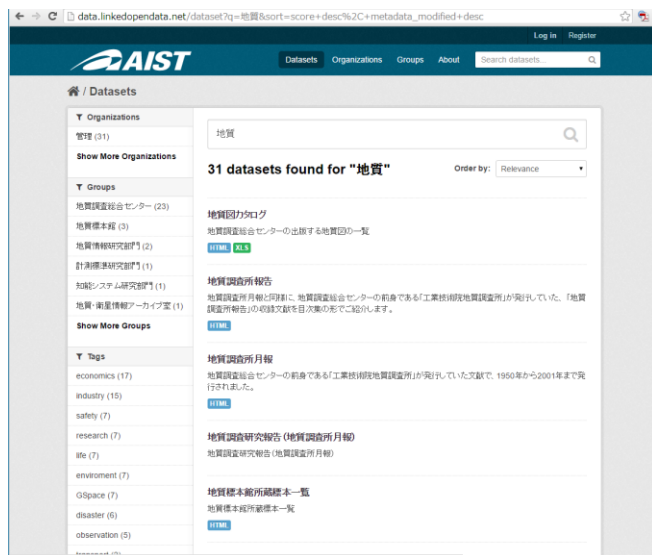
ロジスティック回帰の内部処理



InfoWorld Bossie Awards 2014: Best of Open Source
Big Data Toolsを受賞
<https://github.com/myui/hivemall>

5:CKANとその高度化

- ▶ CKAN：オープンデータのデファクトカタログ
→GUIを改良してデータへの到達性を上げる
Doiシステムとも連携予定



組織構成を基調に、内包されているデータセットが表示されます。

創薬関連の膜たんぱく質データベース(SEVENS)

SEVENS summarizes GPCR (G-protein coupled receptor) genes that are identified with high accuracy from 59 eukaryote genomes, by a pipeline integrating such software as a gene finder, a sequence alignment tool, a motif and domain assignment tool, and a transmembrane helix predictor. This treats a larger data space (than that in currently available other databases), which should include not only the expressed sequences but also the newly identified sequences that cannot be detected by in vivo experiments, although they definitely exist on the genome sequence and are just waiting for the opportunity to express their functions. SEVENS can provide the infrastructure of general information of "GPCR universe" for comparative genomics.

タイトル	創薬関連の膜たんぱく質データベース(SEVENS)
対象地域	
開始日	
終了	
作成頻度	
備考	

CKAN AIST opendata-ml@aist.go.jp | CC BY4.0

まとめ1：（オープンデータの2次利用促進）

情報技術の研究成果：それなりな評価

→使えるサービスとして提供しないと実際のインパクトがない。

- 研究成果である技術をサービス化して競争して行きたい（が事業化までは困難）。
- 無料提供 → 公的サービスとしてはGoodだが、いずれ運営コストを回収すべきという話が出る。
- フリーミアムモデルを取れる民間にまかせては？
 - ▶ 広告・お試し無料

データの公開後評価

公開後評価：データに対する引用・評価の仕組み

データにdoi(digital document identifier)を付与していく。

- JST/JaLCLにおいて実験プロジェクトが進行中（産総研も参加）

● メリット：論文のようにデータを識別/引用できるようになる。

● 課題：様々なポリシーを決める必要があるのでは？

- ▶ doiの単位は？
 - ▶ 引用の方法は？
 - ▶ 更新前のデータが引用された場合は？
 - ▶ データの一部を引用したい場合は？
 - ▶ Landing Page（データを紹介するページ）の作り方とその形式は？
- などなど、

● ポリシー：必ずしも合意の必然はないが、折角なので互換性を高めたい。

- ▶ 合意・決定したものを日本スタンダードとして(例えばRDAのData Citation WGで)主張していくことができる(かも)
- ▶ 日本はオープンデータ・オープンサイエンスで出遅れているので、目立てる分野や重要なポイントを戦略的に見きわめてアプローチして行く必要がある。

doiの基盤と公開データの質

- データ公開のための何らかの基盤を実現する必要がある。
 - ▶ 組織内でdoiを与えたりLanding Pageを作って公開
 - ▶ 安定的なデータの置き場
- 公開データ＝普通は公開組織が質を(暗黙的に)保つ
 - ▶ Dspace等のソフトウェアには公開承認フローが付随
 - ▶ WDSも同様(加入＝ある種の質を満足することを示す)
- データジャーナル：データを査読して公開 (Science Data等)
 - ▶ 出版費用が必要(オープンジャーナルのデータ版)
 - ▶ その代りdoiの付与等を全部行う
 - ▶ 査読メカニズムとその質の評価はまだまだこれから
- JST/jaLC実験プロジェクトでやっていること自体が民間出版社にアウトソースされるかもしれない。

JaLC/Doiテストプロジェクトへの参加

- データベースの「引用」を実現する基本的なしくみ。
- 産総研の現状に即してみると、色々な課題が考えられた。
 - ▶ ネーミング・管理体系
 - ▶ doiを与えるデータ単位と利便性・管理の手間等の関連
 - ▶ 対応するシステムの改修の必要性

1 : doiを付与する単位と既存システムの改修

データに付与するというより「サービスの一断面」に付与
→システムの挙動と併せて確認の必要がある。

● できるだけ細かい方が引用には便利

- ▶ 検索の結果でもあるのでシステムの挙動と対応が取れないといけない。
- ▶ Landing Pageやメタデータの準備も増える。

● 同時に、DB全体でも参照できると良い

→いくつかの単位で試験を行って挙動を確認する

SDBS(有機化合物のスペクトルデータベース)

<http://sdb.s.db.aist.go.jp/>

トップ画面(検索)

SDBS化合物・スペクトル検索

化合物名(英語名・日本語名):

分子式:

分子量:

CAS登録番号:

SDBS番号:

検索

検索結果

SDBS Search Results: 1 - 20 out of 172 hits

SDBS No	Molecular Formula	Molecular Weight	MS	CNMR	HNMR	IR	Raman	ESR	Compound Name
4544	H ₂ O	18.0	N	N	Y	N	Y	N	water
40012	HLIO H ₂ O	23.9	N	N	N	Y	N	N	lithium hydroxide monohydrate
84	CH ₂ O	30.0	N	N	Y	N	N	N	formaldehyde
69	CHDO	31.0	N	N	N	N	N	N	formaldehyde-d
40008	H ₂ N ₂ O	32.0	N	N	N	N	Y	N	nitroxide radical
40013	HNAO	40.0	N	N	N	Y	N	N	sodium hydroxide
40006	CHNO	43.0	N	N	N	N	Y	N	isocyanic acid
40008	HBO ₂	43.8	N	N	N	Y	N	N	metaboric acid
8337	CH ₂ NO	44.0	N	N	N	N	Y	N	carbamoyl radical
10523	CH ₂ O ₂	46.0	Y	Y	Y	Y	Y	N	formic acid
3694	CHLIO ₂ H ₂ O	52.0	N	N	Y	Y	N	N	lithium formate
8122	C ₂ H ₂ NO	56.0	N	N	N	N	Y	N	cyano(dioxymethyl) radical
40010	HKO	56.1	N	N	N	Y	N	N	potassium hydroxide
4126	C ₂ H ₂ O ₂	58.0	N	N	N	Y	N	N	glyoxal
4421	CH ₂ OSi	58.0	N	N	N	N	Y	Y	SiOCH ₂ radical
40019	H ₂ MGO ₂	58.3	N	N	N	Y	N	N	magnesium hydroxide

検索

SDBS Information

SDBS No.: 4544

Compound Name: water

Molecular Formula: H₂O

Molecular Weight: 18.0

CAS Registry No.: 7732-18-5

Derivatives: display in a separate page

Spectral Code: ¹H NMR: in CD₂Cl₂ at 23°C

化合物情報

SDBS Information

SDBS No.: 4544

Compound Name: water

Molecular Formula: H₂O

Molecular Weight: 18.0

CAS Registry No.: 7732-18-5

Derivatives: display in a separate page

Spectral Code: ¹H NMR: in CD₂Cl₂ at 23°C

スペクトル表示等(H-NMR, Raman)

SDBS Information

SDBS No.: 4544

Compound Name: water

Molecular Formula: H₂O

Molecular Weight: 18.0

CAS Registry No.: 7732-18-5

Derivatives: display in a separate page

Spectral Code: Raman: 4890 A.200 M.liquid

1つの情報が複数のフレームで構成されて表示

試験内容と結果

次の三種類のを登録する。

1. SDBSのWebサイト： ○
2. sdb sno=4544 化合物「水」の化合物情報： ○
3. 水の1H NMRスペクトル HR2014-02904NS: ×

doiとそれによって解決されるURLの組み合わせテスト

1. SDBS番号（SDBSの中で使われる・使ってほしい番号）をdoi suffixに入れる・入れない、など
 - Ⓜ 05.sdb s-4544-HR20xxxxxx
 - Ⓜ 05.sdb s-HR20xxxx
2. URLがcgiになるのでいくつかのパターンをテスト
 - Ⓜ http://sdb s.db.aist.go.jp/sdb s/cgi-bin/direct_frame_disp.cgi?sdb sno=4544
 - Ⓜ http://sdb s.db.aist.go.jp/sdb s/cgi-bin/img_disp.cgi?disptype=disp3&imgdir=hpm&fname=HR201402904NS&sdb sno=4544

結果：

- doi:ネーミング等での問題はない。
- システム側：スペクトルデータに直接到達できるcgiを設ける必要があるが、複数フレームを同期する方法がない。
- フレームを多用するシステムの場合、大きな改修が必要

2 : Landing Pageの自動生成 分散型熱物性データベース <http://tpds.db.aist.go.jp/>

物質の熱物性値（熱伝導率、比熱容量、密度、表面張力、蒸気圧など）のデータベース

● 組織に跨った分散データベース

● **ポップアップでデータ表示(検索結果&ナビゲーション結果)**

The image shows a composite screenshot of the TPDS-web interface. On the left is the main landing page with navigation links and a search bar. The central part shows search results for 'molten_mixture' with a table of data points. A blue arrow points from the table to a detailed data popup window. Another blue arrow points from the popup to a 2D graph window showing a plot of Density (kg_m-3) vs Temperature (K) for Water Liquid. The graph shows a decreasing trend of density with increasing temperature, with a sharp drop at the boiling point.

Thermofluorescence signal (I)	Time [s]
1.0054E+00	-6.0000E-08
9.8057E-01	-5.9500E-08
1.0431E+00	-5.9000E-08
1.0070E+00	-5.8500E-08
9.5907E-01	-5.8000E-08
9.6680E-01	-5.7500E-08
9.9516E-01	-5.7000E-08
9.5730E-01	-5.6500E-08
9.3185E-01	-5.6000E-08
8.4992E-01	-5.5500E-08
9.1557E-01	-5.5000E-08
7.8808E-01	-5.4500E-08
7.8513E-01	-5.4000E-08
8.7153E-01	-5.3500E-08
8.2497E-01	-5.3000E-08
9.1053E-01	-5.2500E-08
8.6292E-01	-5.2000E-08
8.4509E-01	-5.1500E-08

試験内容と結果

● 検索結果の物質にLanding Pageをプログラムから生成する

- ▶ JaLCスキーマに相当するデータ+αを表示
- ▶ Landing Pageに埋め込まれたリンクから実データをアクセス & ポップアップ

◎ 1物質・URLで試験→アクセス可

▶ **ポップアップ表示によるDB**

- ◎ Doiから別ページで表示
- ◎ 意味的な関連が理解しにくい
- ◎ 表示システムの再検討必要？

The screenshot shows a web browser with two windows. The main window displays a landing page for a material with the following details:

- Title:** Thermoreflectance signal for TiN on synthesized quartz substrate, G1, Lt1, Specimen1
- DOI:** 10.5072/05.tpds-23732
- Publisher:** National Institute of Advanced Industrial Science and Technology
- Published Date:** Year:2011
- Description:** Measurement Instrument Nanosecond thermoreflectance measurement system. Note: The data is a raw signal on the measurement.
- Related Contents:** Type:DOI, Relation: http://dx.doi.org/10.1143/JJAP.50.11RH03
- Creators:**
 - CreatorsSequence: 1, Creator First Name: Yuichiro, Creator Last Name: Yamashita
 - CreatorsSequence: 2, Creator First Name: Takashi, Creator Last Name: Yagi
 - CreatorsSequence: 3, Creator First Name: Tetsuya, Creator Last Name: Baba
- Other Information:**
 - Material Name: TiN on synthesized quartz substrate, G1, Lt1, Specimen1
 - Property: Thermoreflectance signal
 - Data Registrant: Yuichiro Yamashita
 - Link to Research data: [Table Data](#)

The 'Table data' popup window shows a table with the following data:

Thermoreflectance signal (1)	Time (s)
1.0094E+00	-6.0000E-08
9.8057E-01	-5.9500E-08
1.0431E+00	-5.9000E-08
1.0670E+00	-5.8500E-08
8.6600E-01	-5.8000E-08
8.6600E-01	-5.7500E-08
8.6600E-01	-5.7000E-08
9.3188E-01	-5.6500E-08
9.3185E-01	-5.6000E-08
8.4892E-01	-5.5500E-08
9.1557E-01	-5.5000E-08
7.8808E-01	-5.4500E-08
7.8513E-01	-5.4000E-08
8.7153E-01	-5.3500E-08
9.2497E-01	-5.3000E-08
9.1053E-01	-5.2500E-08

3 : doiの決め方

● 産総研内で複数の組織が別々に登録すると予想

- ▶ 組織で統一してdoiを与えるような体制には現在ない。
 - ◎ プロジェクトなりDB担当者が行う可能性が高い。
- ▶ 名前の衝突等を避ける必要がある
 - ◎ 名前決めの都合はデータの保有者が一番分かっている？

● 今回は仮に以下のようにポリシーで決めてみた。

- ▶ (prefix)/((もしあれば) サブ組織名) -(DB名)- (識別子)
 - ◎ 例 : <http://doi.org/10.14977/05.tdbs-23732>
 - ◎ 例2 : <http://doi.org/10.14977/05.gsj-aster-xxxx>
- ▶ DB名でsuffix空間を分け、以降はDBの担当者が決める。
 - ◎ 自主管理ルールなので他の担当者が間違える可能性は残る
- ▶ 将来サブ組織が必要ならその前に入れる。
 - ◎ 別ドメインのある組織があるから(gsj/nmij)

まとめ：（doiテスト）

doiを付与すること(登録など対JaLCの作業) 以上に、
 所内でのさまざまな業務やシステム改修が想定され、その手間の軽減が重要。

- ④ Landing page の作成と公開
- ④ 対応システムの改修・システム内でのdoiの利用
- ④ suffixの登録管理などの内部業務フロー

▶ 導入できるソフトウェアが提供されるとありがたい。

- ④ たぶんどの組織でもニーズはおおむね共通と思われる。
- ④ 独自の作り物は長期的な維持管理が問題なので、できれば既存のソフトのアドオンなりが望ましい。

▶ アウトソースでも良い

- ④ クラウド等のホスティングサービスで提供されてもよいが、landing pageがホスティング先に作られるので、ドメインだけでもaistである必要がある。
- ④ 上記の業務も含めいっそ全部特定のデータジャーナルにアウトソース？

● doi付与のメリット= doiが付くだけでは十分ではない

- ▶ データ提供者にとってデータの認知度の上がる付加価値サービス
- ▶ データ利用者にとってデータの発見の容易な付加価値サービス

議論

論点（IT研究・サービスの視点から）

- データ公開に伴う様々な課題・付加価値の提供
 - ▶ ○ サービスとして実現
 - ▶ ○ なんらかの情報技術的な研究課題解決の成果
 - ▶ × 完成度の高いサービスで使える形にしないとイケない。
 - ▶ × 単一の機能だけではユーザは使ってくれない。
- クラウド化：データの公開だけでなく研究活動そのものがクラウド上に
 - ▶ ○ ：データ解析、コミュニティ支援も含めクラウドで完結
 - ▶ ○ ：利便性の高い環境
 - ▶ × ：主に民間が提供→公的機関の立ち位置は？
 - ▶ × ：データも含め研究全体がクラウドで完結

Illumina BaseSpace

データに基づいた研究活動を支援するプラットフォーム

- Illumina社：シーケンサー（ゲノムの読み取り装置）の会社
- BaseSpace ライフサイエンス研究者のためのクラウド
 - ▶ 無料（今後有料化の計画有り）
 - ▶ データ共有/様々なオープンデータの提供
 - ◎ 秘密にしておきたいデータも保管
 - ▶ データの様々な解析手法の提供（オンラインで使える）
 - ◎ アプリの登録と相互利用が可能
 - ▶ 結果の可視化
 - ◎ 手元のPCのブラウザで解析結果を閲覧可能
 - ▶ グループ管理など研究そのものの支援
 - ◎ 研究そのものをクラウド上で便利に遂行できるようにして、研究の活発化と囲い込みを同時に行う。

地理空間分野でもesri等似たようなサービスを始めている
(ArcGIS Online/ArcGIS open data)

まとめ

- 研究を支援する・研究を行うプラットフォームはビジネスになる。
 - ▶ 共通領域・データ提供者としての産総研
 - ▶ 競争領域・サービス提供者としての産総研
- 引用の仕組みや評価の指標等もビジネスになる。
 - ▶ 便利なのは良いことだが、企業による寡占化の可能性も
 - ▶ データを持っている公的セクタが迅速に進めることもありでは？
- インターネット上での研究活動：
 - ▶ データ分散はワークしないが、分散した環境から必要なデータや収集して解析のためのワーキングセットを迅速に作る技術は必要。
 - ▶ サービスそのものは分散しているので連携が必要。
- 集中のメリット：
 - ▶ 歴然としてある
 - ▶ 便利なところにデータやサービスや人が集まってきます。
 - ▶ その中での公的セクタの役割は？
 - ○サービスへのデータ提供者
 - ?サービスそのものの提供者？

まとめ

- オープン化：オープンアクセスの重視。存在等メタデータは（必ず？）オープンにする等
 - ▶ ○データやサービスの見える化が進む。
 - ▶ ×クローズポリシーとの関係
 - ◎メタデータはオープンにしてほしいが、
- 標準化：接続性を保証し、ベンダーロックインを避けるため、標準化されたインターフェイスやフォーマットを重視
 - ▶ ○接続性が保証される。
 - ▶ ×規格がない場合の合意形成等の手間等が大きい
- ボトムアップ：多様な組織が相互に連携できるように、ボトムアップな環境構築
 - ▶ ○多様なプレーヤが連携できる。
 - ▶ ×エンドユーザにとって効率的ではない。