

産業技術総合研究所

渡邊 宏 (計量標準総合センター 物質計測標準部門)

山下雄一郎 (計量標準総合センター 物質計測標準部門)

小島 功 (情報・人間工学領域 研究戦略部)

概要

- 産総研の研究データベースの現状
- doiテストプロジェクトへ参加
- 実験内容・参加DB紹介
- まとめ

産業技術総合研究所の研究データベース

研究情報公開データベース

- http://www.aist.go.jp/aist_j/aist_repository/riodb/index.html
 - Web上で公開している研究成果のデータベース
 - 上記リストで約50+
 - 他にも研究ユニットが研究プロジェクト内で独自に作成・提供している
 - 多くはWebサービスとして検索や可視化などを提供



本プロジェクトへの参加

産総研におけるオープンデータ推進の中の1活動

- 産総研データバンク構想：データベースの高度化による利用促進
 - 研究DBのLOD(Linked Open Data)化、外部DBとの連携
 - ツールの高度化(分散SPARQL検索など)
 - CKANカタログの構築・インターフェイスの高度化
 - データベースの統合・最適化
 - doiの付与の検討
- doiテスト
 - 内部で呼びかけ：材料系の2つのDBからご協力いただく
 - 有機化合物のスペクトルデータベース (SDBS)担当：渡邊宏
 - <http://sdfs.db.aist.go.jp/>
 - 熱物性データベース：担当：山下雄一郎
 - <http://tpds.db.aist.go.jp/>
 - 全産総研的な動きにすべく、引き続き内部アピールや協力拡大を図っていく必要がある

試験検討項目

1. Doiのsuffix(10. 0.14977/05.XXXXXXX) の決め方
2. Doiを付与する単位
3. Landing Page の自動作成など、対応するシステムの改修の必要性等

1. 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)

2.doiを付与する単位

- 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.db.aist.go.jp/

- 産総研の研究情報DBで最も大規模かつ著名なDB (1997~Web公開)
- トップ画面(検索) 検索結果

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:

SDBS番号:

検索

検索

SDBS Search Results: 1 - 20 out of 172 hits

SDBS No	Molecular Formula	Molecular Weight	MS	CNMR	HNMR	IR	Raman	ESR	Compound Name
4544	H2O	18.0	N	N	Y	N	Y	N	water
40012	HLiO·H2O	23.9	N	N	N	Y	N	N	lithium hydroxide monohydrate
94	CH2O	30.0	N	N	Y	N	N	N	formaldehyde
94	CHDO	31.0	N	N	N	N	N	N	formaldehyde-d
94	H2NO	32.0	N	N	N	N	Y	N	nitroxide radical
40008	HNAO	40.0	N	N	N	Y	N	N	sodium hydroxide
40008	CHNO	43.0	N	N	N	N	N	Y	isocyanic acid
40008	HBO2	43.8	N	N	N	Y	N	N	metaboric acid
8337	CH2NO	44.0	N	N	N	N	N	N	carbamoyl radical
10523	CH2O2	46.0	Y	Y	Y	Y	Y	N	formic acid
3694	CHLiO2·H2O	52.0	N	Y	Y	Y	N	N	lithium formate
9122	C2H2NO	56.0	N	N	N	N	N	Y	2-(hydroxymethyl) radical
40010	HKO	56.1	N	N	N	Y	N	N	potassium hydroxide
4126	C2H2O2	58.0	N	N	N	Y	N	N	glyoxal
4421	CH2OSi	59.7	N	N	N	N	N	Y	SiOCH2 radical
40019	H2MGO2	58.3	N	N	N	Y	N	N	magnesium hydroxide

SDBS Information

SDBS No.: 4544

Molecular Formula: H₂O

Molecular Weight: 18.0

CAS Registry No.: 7732-18-5

Derivatives:

Spectral Code:

Compound Name: water, hydrogen oxide

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

Spectral Code:

1H-NMR: in CD₂Cl₂ at 27C

1H-NMR: in C₆D₆

1H-NMR: in CD₃CN

1H-NMR: in CD₃CN-d₃

1H-NMR: in DMSO-d₆

1H-NMR: in DMSO-d₆

1H-NMR: in CDCl₃

1H-NMR: in Acetonemid₂

1H-NMR: in C₂D₂Cl₄

1H-NMR: in C₂D₂Cl₄

Raman: 4890 A 200 M liquid

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

Spectral Code:

1H-NMR: in CD₂Cl₂ at 27C

1H-NMR: in C₆D₆

1H-NMR: in CD₃CN

1H-NMR: in CD₃CN

1H-NMR: in DMSO-d₆

1H-NMR: in CDCl₃

1H-NMR: in Acetonemid₂

1H-NMR: in C₂D₂Cl₄

1H-NMR: in C₂D₂Cl₄

Raman: 4890 A 200 M liquid

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。

3. Landing Pageの生成などシステムの対応 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 screenshot displays the TPDS-web interface with several components highlighted by blue boxes and arrows:

- Search Results Table:** A table showing search results for 'molten_mixture'. The table has columns for 'Material Name', 'Property', and 'Direction'. The data includes:

Material Name	Property	Direction
molten_mixture	Amount-of-substance fraction of substance 1 is	
molten_mixture	Mass diffusion coefficient	
- Graph:** A 2D graph showing the relationship between Temperature (K) on the x-axis and Density (kg_m-3) on the y-axis. The x-axis ranges from 298.15 to 323.15 K, and the y-axis ranges from 9.61E2 to 1.00E3 kg_m-3. The graph shows a downward-sloping curve for 'Water,Liquid'.
- Navigation Elements:** A '2D Graph' button and a 'Graph' button are visible, indicating the interactive nature of the data display.

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からアクセスした場合の情報の断片化が懸念
 - 表示システムの再検討必要？

The screenshot shows a web browser window with two panes. The left pane displays a landing page for a material with the following information:

- Title: Thermoreflectance signal for TiN on synthesized quartz substrate, G1, L11, 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.
- Type:Abstract
- 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, L11, Specimen1
 - Property: Thermoreflectance signal
 - Data Registrant: Yuichiro Yamashita
 - Link to Research data: Table Data

The right pane shows a data table with the following content:

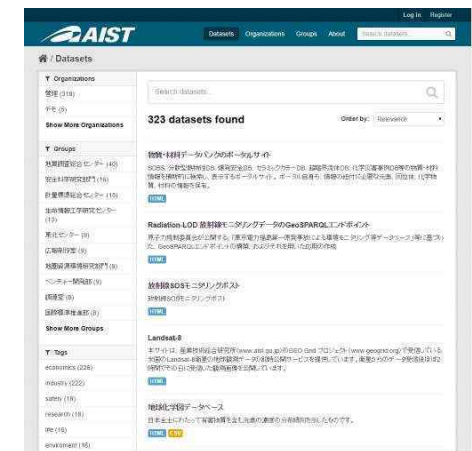
Thermoreflectance signal (1)	Time (s)
1.0094E+00	-6.0000E-08
9.8057E-01	-5.9500E-08
1.0431E-01	-5.9000E-08
1.0670E-01	-5.8500E-08
8.6681E-01	-5.8000E-08
8.9515E-01	-5.7000E-08
9.5738E-01	-5.6500E-08
9.3185E-01	-5.6000E-08
8.4892E-01	-5.5500E-08
9.1557E-01	-5.6000E-08
7.8808E-01	-5.4500E-08
7.8513E-01	-5.4000E-08
8.7153E-01	-5.3500E-08
9.2497E-01	-5.3000E-08
9.1053E-01	-5.2500E-08

A blue arrow points from the 'Table Data' link in the landing page to the data table in the right pane.

CKAN? as a Landing page provider

- CKANのページをLanding pageに使う

- CKAN : オープンデータでのカタログサイトのデファクト
- メタデータ項目=Dublin core
 - -> これをJaLC/DataCiteのメタデータ項目に拡張(可能)
- まだ作業中(未テスト)
 - CKANそのものが内部にしか公開されていない。



- Dspace?

- こちらの方が拡張性および機能が高い
- doiがなくても、handleでそもそも十分?

まとめ

基本的には意図した通りのdoiの付与が可能と思われる。

doiを付与すること(登録など対JaLCの作業)に加え、

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。

- Landing page の作成と公開
 - 対応システムの改修・システム内でのdoiの利用
 - 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。
- たぶんどの組織でもニーズはおおむね共通と思われる。
 - 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
- 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。
 - 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？

引き続き検討をしつつ正式な登録へ進んでいきたい。

- **所内体制の整備**