

Ernest Satow

余也吉和漢之書庸日本以來臨習之久殆無暇日
 數年前觀三洲長先生書書格儼正真可則而
 以未見悉其諸體者深為憾焉吾友兒玉君
 梓行先生之書請而觀之盡和漢書體無可間
 然矣且哉擢內度授書法於
 明治天皇也今也日本人學歐亞文字歐亞人
 學日本文書通信於遐邇之地保親於兩邦之

間然則雖歐亞人和漢之書不可不學和漢書則
 宜法長先生之書格也是為序書韻華帖首雜
 歲紀元一千八百七拾八年明治十一年十月七日

大英國書記官薩道慈



余客日本十二年於茲學和漢文又
習和漢字稍得作贈答文章辨書
法可否兒玉少介君携長先生書
韻華帖來披而觀之書法立端正頗
極神妙聞

日本皇帝現師先生之書是可
知余不過譽也曰書卷首以告各
國人欲學和漢書法者

紀元一千八百七拾八年

日本明治十一年九月十日

佛蘭西國 書 子ユフス 介

阿 怡 成 寅 九 月 十 日 觀 于 菴 蘿 庄



韻 十二

韻 十二

Albert Charles Du Bousquet

Cher Monsieur Kodama,

Je viens d'apprendre que vous devez livrer à l'étude de tous, la méthode d'écriture de Monsieur Echoko, méthode qui vient d'être imprimée. Je suis heureux de cela et je vous écris pour vous dire combien je suis satisfait et heureux de voir que ceux qui désirent étudier les sciences pourront en tirer de grands avantages.

Monsieur Echoko, comme vous le savez, a beaucoup étudié les langues japonaise et chinoise, ancien officier au Ministère de l'Instruction publique il fut chargé comme chef de faire faire des ouvrages utiles à l'Instruction. Excellent écrivain pour les caractères chinois et japonais, il a obtenu la place de professeur de sa Majesté l'Empereur. Ses talents le mettaient au premier rang comme écrivain.

Monsieur Dubousquet, ancien officier militaire français est à même de juger ce qui est bien, parce qu'il a au Japon consacré de longues années à l'étude des langues chinoise et japonaise. Il a été surpris en voyant les écrits de Monsieur Echoko et a même ajouté, que tout étranger désireux d'apprendre et d'apprendre la langue japonaise devrait prendre l'œuvre de Monsieur Echoko pour modèle. Mon opinion dans la circonstance est que ce jugement est juste et mérité.

Je suis très satisfait, cher Monsieur Kodama d'avoir à vous communiquer ces choses qui, j'en suis persuadé vous feront plaisir.

Zouque
Professeur à l'École de droit,
Ministère de la Justice. Tokio.

研究基盤の「維持」・発展が不可欠 & 実験を実験で終わらせない更なる展開へ

後継計画の展望

➤ 更なる画像データの拡充

- ・国内外の機関が国書DBへの接続を要望！
- ・国初から近代に至る日本の書物のシームレスなデータ連携

➤ 30万点の画像からテキストへ

- ・テキスト化実証試験をうけ、テキストデータ作成推進
- ・全テキストの公開へ。できれば必要情報の抽出可能（深層学習の取組等）

➤ 古典籍メタデータの標準化/テキストの構造化の推進

- ・国立国会図書館・JPCOAR（オープンアクセスリポジトリ推進協会）等との協議によるメタデータの標準化推進
- ・TEI（Text Encoding Initiative）関連の取組

➤ 総合書物学、典籍防災学等の更なる展開

- ・整備されたデータの自然科学・社会科学分野における利活用を促進

➤ マテリアルとしての書物分析・解析技術開発

- ・書物を「内容」からの分析だけではなく、「モノ」として「研究資源」とみるマテリアル分析の推進

我が国の史資料は膨大であり、史資料を所蔵する様々な人文系諸機関等が協力して更なるデータ連携を図っていくことが重要

いかにデータを動かすか？

===> 古典籍データ駆動研究センターの設置

教育研究組織改革分（組織整備）令和4（2022）年度～

4つの研究領域からデータ駆動による課題解決型人文学へ

本計画の目的

持続的かつ汎用的な大規模データの蓄積・構築という概念を人文学分野の研究に本格的に導入し、自然科学を含む他分野にも利用可能なデータインフラストラクチャを人文学分野に築くとともに、その十全な利活用を通して、持続可能な社会構築のために異分野と協働し得るデータ駆動型による課題解決型の人文学研究を創成する。

課題解決型人文学とは？

人間社会の課題に対し、多種多様な人の営みを対象とする学問分野である人文学の立場から総合知を構築し、その解決を試みる新しいタイプの人文学。

人と生物、人と自然など、ヒト以外のさまざまなモノとも共生しうる社会、及び資源とエネルギーが循環するレジリエントな共生社会の構想を支える基盤となる。

人間社会の
課題解決

データ駆動による
課題解決型人文学



データの
集積・変換

データインフラ
ストラクチャの構築

多分野にわたる総合的研究を
目指す基盤として
データインフラストラクチャを構築



データの
解析・変換

人文系データ分析
技術の開発

日本語で記された人文系データの
機械処理のための技術向上と
研究支援を推進



データ
の利活用

コンテンツ解析
からの展開

人文学の範囲を超える、
自然科学・社会科学分野における
利活用を促進

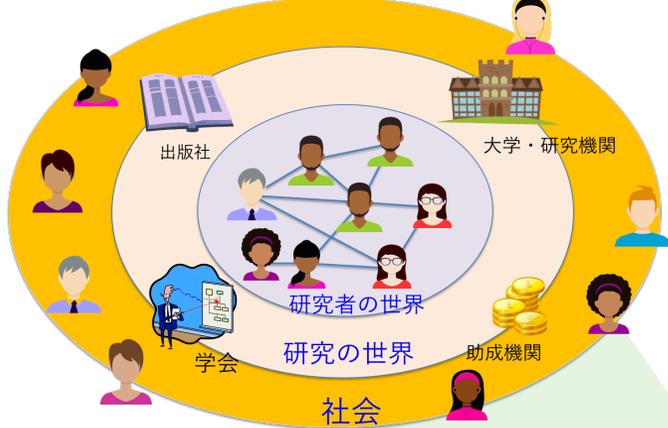


新たな
データ開拓

マテリアル
分析・解析

マテリアルとしての分析
・解析技術の開発を行い、
情報抽出を推進

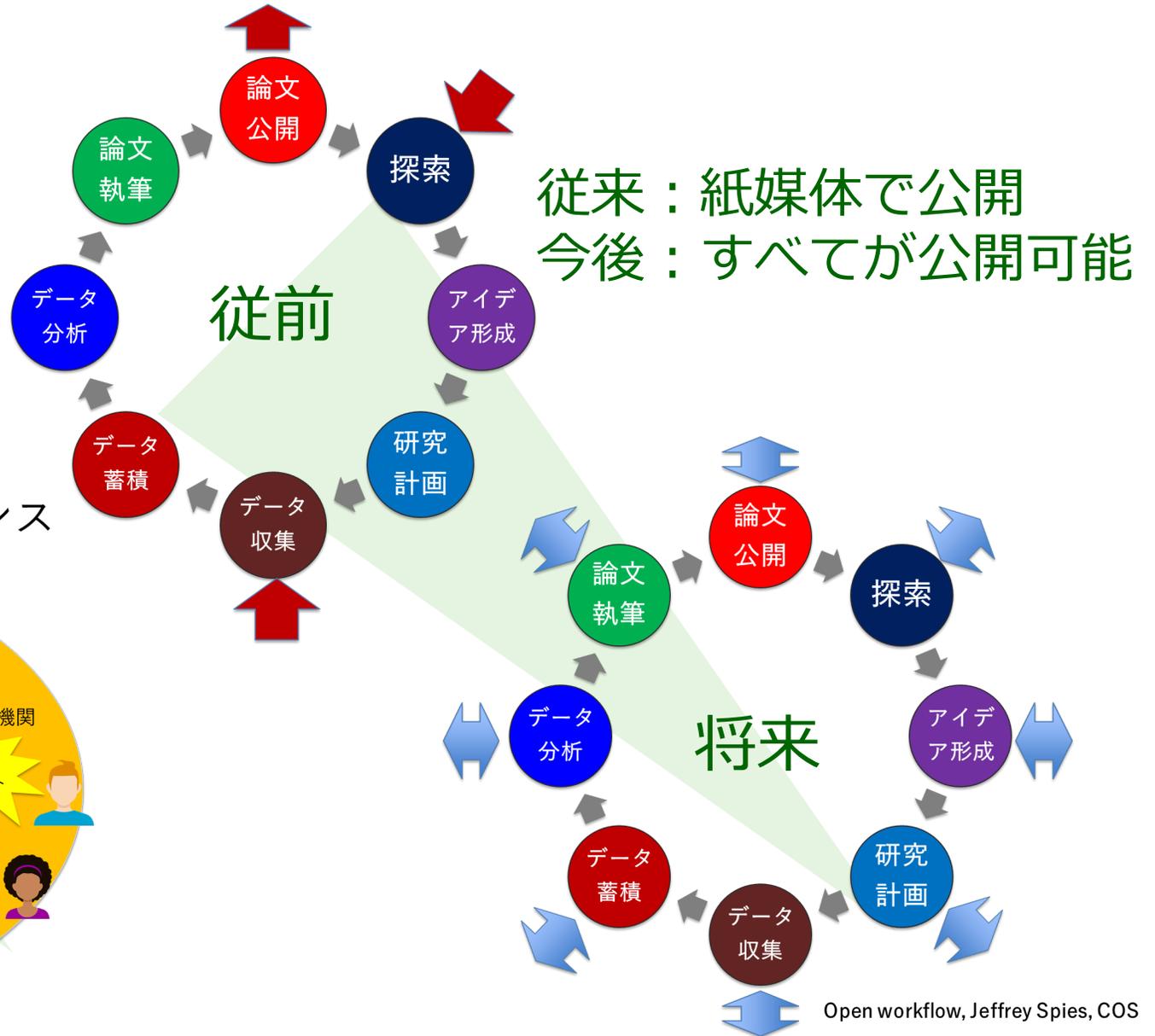
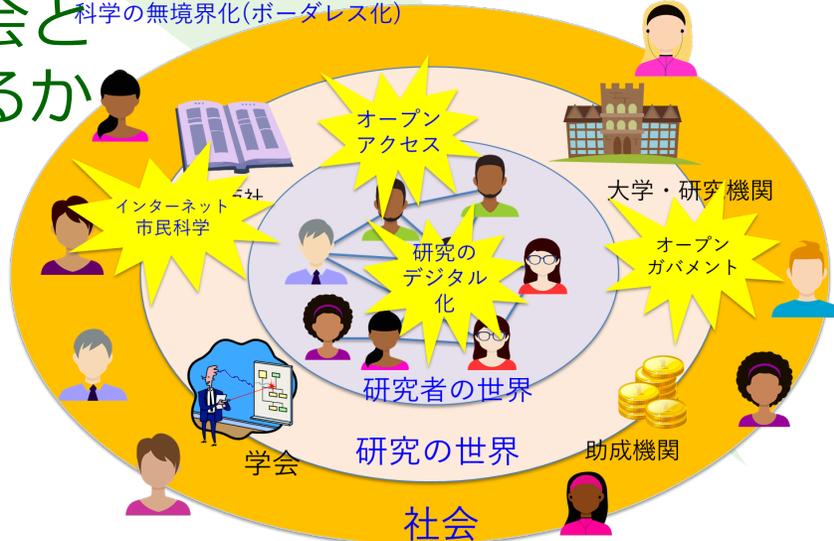
研究と社会



研究と社会とオープンサイエンス

研究者は社会と繋がっているか

科学の無境界化(ボーダレス化)

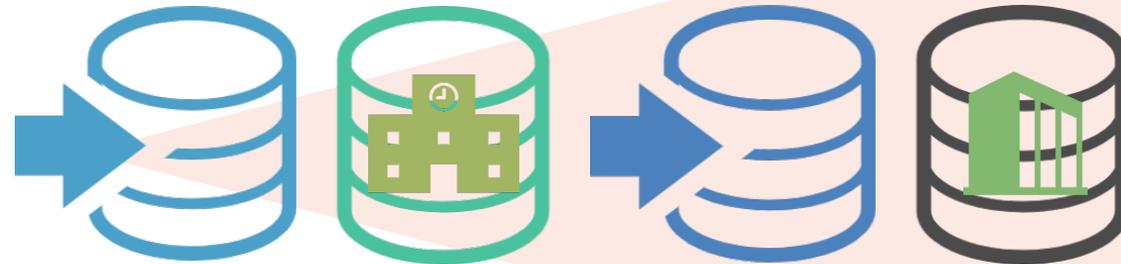


【現状】バラバラにDB存在



DX（デジタルトランスフォーメーション）は、明確に定められた目的を達成してこそ意味のあること。

デジタル化だけが目的ではない。
ここで出来ること出来ないことを明確に意識化する必要あり（今現在、何が出来て、何が出来ないか）



プレゼン

オンライン
ジャーナル



ワンストップで出来ること／出来ないこと
代替案を構想できるか

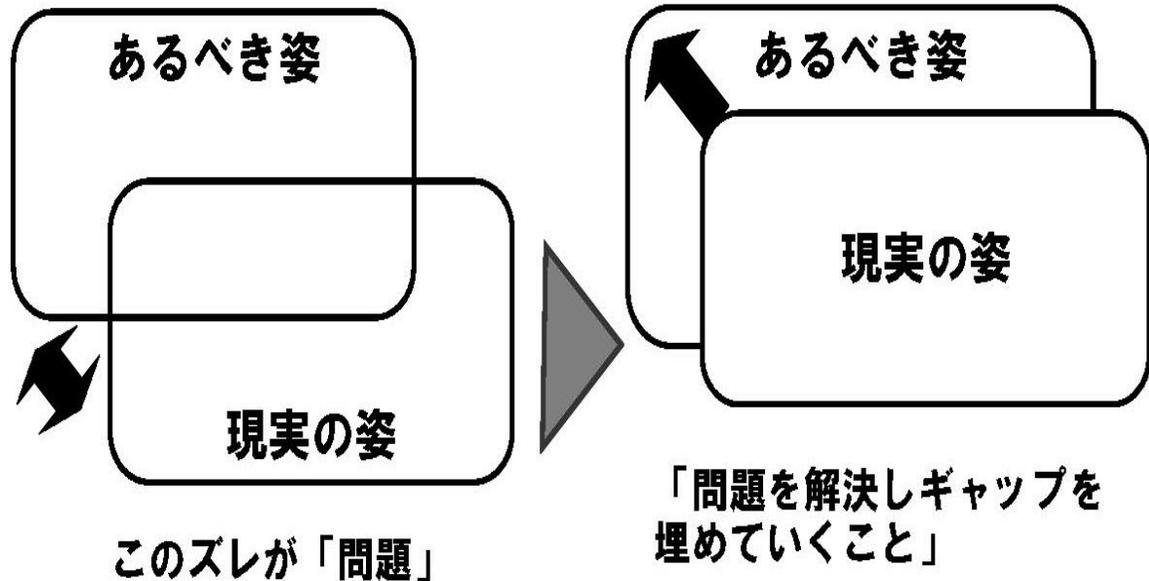


論文

問題解決 (ソリューション) の定義

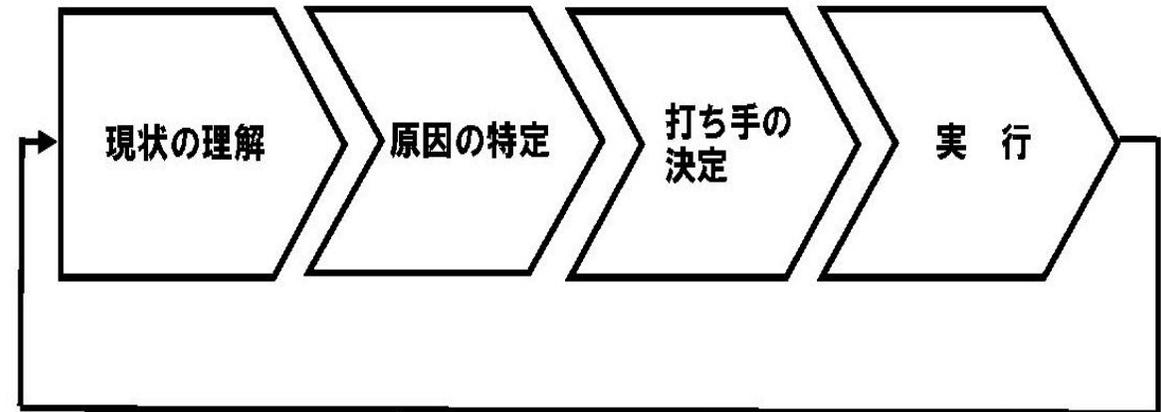
= 「あるべき姿と現実の乖離、逸脱 (ギャップ) を埋めていくこと」

要するに、「現実」を「あるべき姿」に持っていくこと。これを《問題解決 (ソリューション)》という。



思いつきだけでは解決に繋がらない
まず現状認識を！

問題解決への流れ



《問題解決の流れ》

- ①現状の理解→②原因の特定→
③打つ手の決定→④実行→①へ (繰り返し)
誰のため、何を対象としての問題解決か？

ネットの世界は膨大

今後求められる能力は？

見つけ出す能力 (情報検索能力)

⇔ 検索ワードが思い浮かべられない

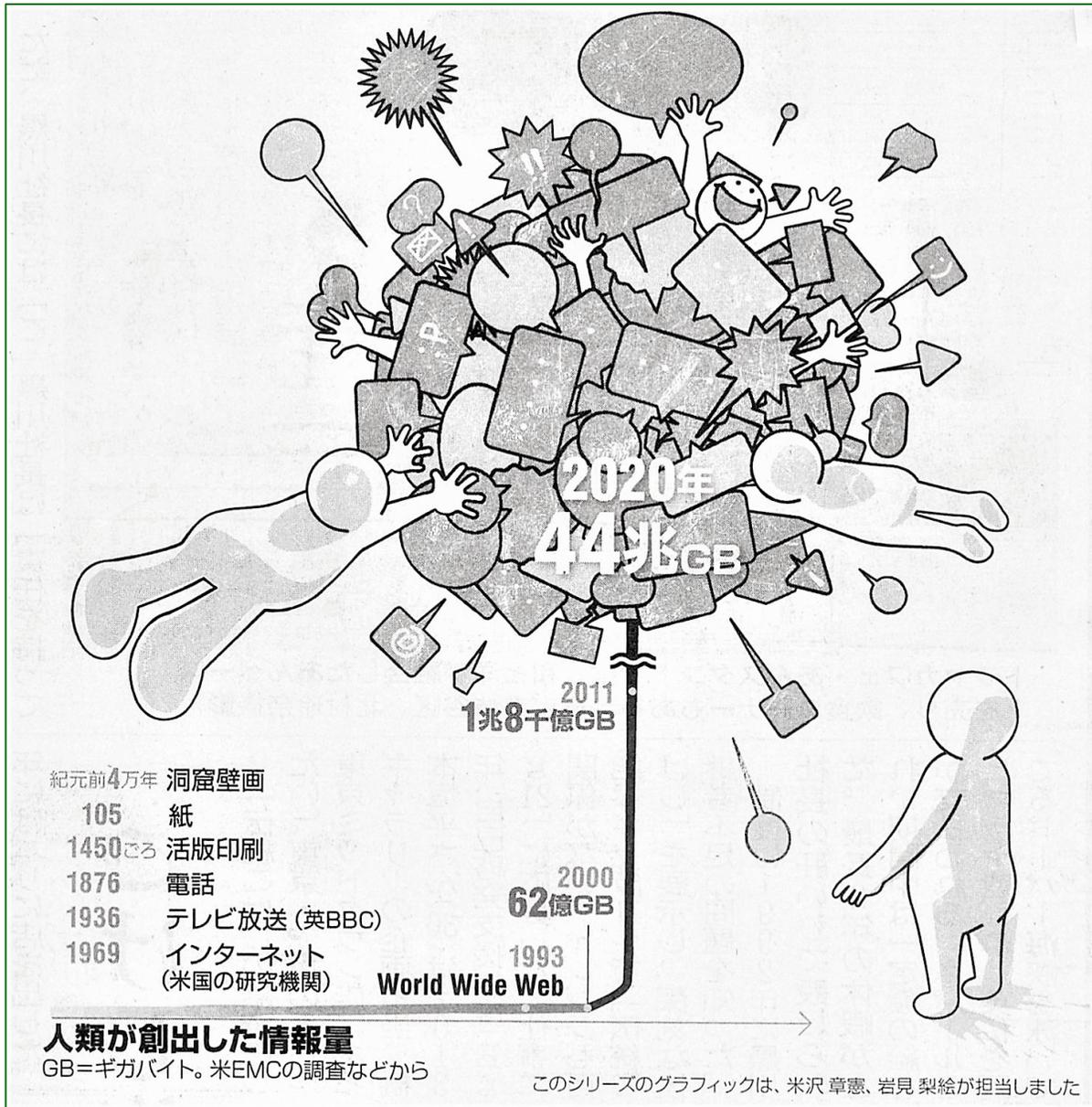
メディアリテラシー (情報判断能力)

⇔ 一元的な情報のみに晒される現実

探す楽しみを如何に創りだすか

= 教育の課題

= 研究者の課題



段階	公開の状態	データ形式例	参考)Linked Open Data 5star
1段階	オープンライセンスの元、データを公開 make your stuff available on the Web (whatever format) under an open license	PDF、JPG	OL – Open License (計算機により参照できる(可読))
2段階	1段階に加え、コンピュータで処理可能なデータで公開 make it available as structured data (e.g., Excel instead of image scan of a table)	xls、doc	RE – Readable (Human & Machine) (コンピュータでデータが編集可能)
3段階	2段階に加え、オープンに利用できるフォーマットでデータ公開 make it available in a non-proprietary open format (e.g., CSV instead of Excel)	XML、CSV	OF – Open Format (アプリケーションに依存しない形式)
4段階	Web標準(RDF等)のフォーマットでデータ公開 use URIs to denote things, so that people can point at your stuff	RDF、XML	URI – Universal Resource Identifier (リソースのユニーク化、Webリンク)
5段階	4段階が外部連携可能な状態でデータを公開 link your data to other data to provide context	LoD、RDF スキーマ	LD – Linked Data (データ間の融合情報が規定。検索可能)

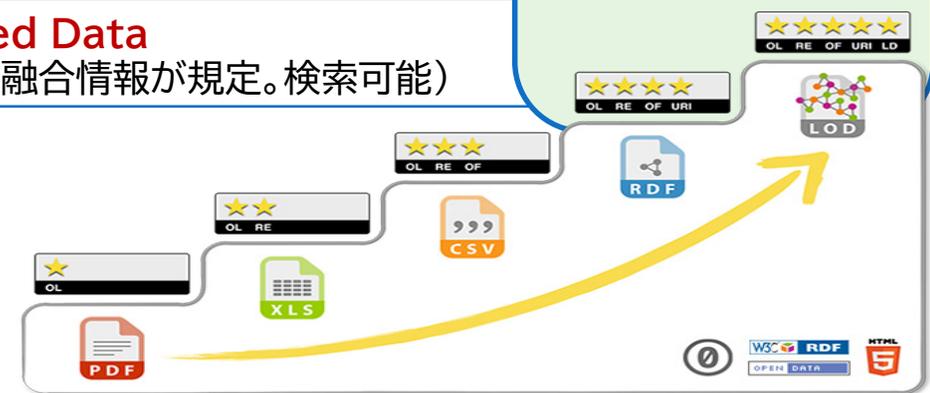
人が理解するための
公開文書(編集不可)

公開文書
(編集可)

機械判読可能な
公開データ

オープンデータの5つの段階

出典:★ Open Dataのサイト(<http://5stardata.info/>)
およびTim Berners-Lee氏のLinked Dataに関する提言ページ
(<http://www.w3.org/DesignIssues/LinkedData.html>)を参考に作成。



<https://force11.org/info/the-fair-data-principles/>

FORCE11
The Future of Research Communications and e-Scholarship

Code of Conduct Login [Join FORCE11](#)

Posts Membership Groups Conference FSCI Events Info

The FAIR Data Principles

Translations

Japanese – <https://doi.org/10.18908/a.2019112601> (added 31.01.2020).

Preamble

One of the grand challenges of data-intensive science is to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows. Here, we describe FAIR – a set of guiding principles to make data Findable, Accessible, Interoperable, and

FAIR原則 (「THE FAIR DATA PRINCIPLES」 和訳)

目

公開する「THE FAIR DATA PRINCIPLES」の正式な日本語訳

は、課題解決に適した科学データや関連するアルゴリズム、ワーク
 統合、解析し、人間や機械の円滑な知識発見を助けることである。こ
 データを「Findable (見つけられる)」、「Accessible (アクセス
 利用できる)」、「Reusable (再利用できる)」にするための一連の

FAIR原則:

データ駆動型サイエンスにおいて、研究データを Findable, Accessible, Interoperable, Re-usableな状態にしておくことが重要。

To be Findable:(見つけられるために)
 eternally persistent identifier(識別子)、
 メタデータ、リポジトリ

To be Accessible:(アクセスできるために)
 standardized communications protocol,
 認証

To be Interoperable:(相互運用できるために)
 共通フォーマット、broadly applicable
 language

To be Re-usable:(再利用できるために)
 license、来歴、標準

甲	きのえ	子	ね
乙	きのと	丑	うし
丙	ひのえ	寅	とら
丁	ひのと	卯	う
戊	つちのえ	辰	たつ
己	つちのと	巳	み
庚	かのえ	午	うま
辛	かのと	未	ひつじ
壬	みずのえ	申	さる
癸	みずのと	酉	とり
		戌	いぬ
		亥	い

Time Information System **HuTime** English

[解析ソフトウェア](#) |
 [時間基盤情報](#) |
 [資料・参考文献](#) |
 [リンク](#) |
 [このサイトについて](#)

[ホーム](#) > [時間基盤情報](#) > [暦変換](#) > 暦変換サービス

暦変換

和暦

書式指定の前に型を選択 自動

寛延二己巳歳正月吉日

暦

書式指定と型

変換

暦日(始点)

グレゴリオ暦(先発グレゴリオ暦)

C.E. 1749-02-17

[暦について](#)

[書式について](#)

左側のボックスに 平成25年4月1日 など、日付を表すテキストをそのまま入力し、変換ボタンを押して複数の日付を入力する場合は、改行して次の行に日付を書いてください。100行までは入力できます。

Copyright © 2012-2020 Tatsuki Sekino. All rights reserved.



<http://www.hutime.jp/basicdata/calendar/form.html>

蔵書印ツールコレクション 画像検索 部首検索 データベース検索

篆字画像検索 (AI篆字認識)

画像から篆字を検索します

試してみる

蔵書印ツールコレクション 画像検索 部首検索 データベース検索

ヘルプ

蔵書印データベース検索システム

蔵書印文

蔵書印主

備考等

形状

印影外郭

書体

印文行数

所蔵先

10,000 件中 1-10 件を表示 (全件数: 44,949 件)

画像	蔵書印文	蔵書印主	色	陰陽	形状	書名	所蔵先	典拠資料
	藤波家蔵書	藤波家	朱	陽	長方形	伊勢大輔集	国文研初雁, W	国文学研究
	青洲文庫	渡辺青洲	朱	陽	長方形	一休骸骨	国文研, W	国文学研究

<https://seal.dhii.jp/>



異体字セレクトセレクト (β) 文A ▾ 説明

CHISE で「卿」を表示 GlyphWiki で「卿」を表示 Codepoints で「卿」を表示

表示	コード	セレクト	画像	コレクション	識別名
↑挿入 卿 コピー	U+537F	-	卿	基底文字	CJK UNIFIED IDEOGRAPH-537F
↑挿入 卿 コピー	U+537F	U+FE00	卿	標準異体字	CJK COM
↑挿入 卿 コピー	U+537F	U+FE01	卿	標準異体字	CJK COM
↑挿入 卿 コピー	U+537F	U+FE02	卿	標準異体字	CJK COM
↑挿入 卿 コピー	U+537F	U+E0100	卿	Adobe-Japan1	CID+137
↑挿入 卿 コピー	U+537F	U+E0101	卿	Adobe-Japan1	CID+169



使い方の説明

<https://digitalnagasaki.hatenablog.com/entry/2017/01/18/030751>

大量の文章データ（テキストデータ）から、有益な情報を取り出すことを総称して**テキストマイニング**と謂う。自然言語解析の手法を使って、文章を単語（名詞、動詞、形容詞等）に分割し、それらの出現頻度や相関関係を分析することで有益な情報を抽出する。

Voyantツール

<https://voyant-tools.org/?lang=ja>

User Local AIテキストマイニング

文書一覧 > 太宰治『走れメロス』 > 解析結果

1つの文書を解析 | 感情分析AI | 文書要約 | 本文検索 (ベータ版) | 解析設定

ワードクラウド | 単語出現頻度 | 共起キーワード | 2次元マップ | 係り受け解析 | 階層的クラスタリング

ワードクラウド
スコアが高い単語を複数選び出し、その値に応じた大きさに表示しています。単語の色は品詞の種類で異なり、青色が名詞、赤色が動詞、緑色が形容詞、灰色が感動詞を表しています。単語をクリックすることで、詳しい本文をご覧いただけます。

共起ネットワークは、単語が共通に出現する関係（共起関係）を円と線で表示した図。どんな単語が同じテキスト内に出現しやすいかが感覚的に把握できるもので、一つ一つの文書を精読する方法とは異なる観点を呼び込む。→使い方を間違えない（発想の第一段階）

単語出現頻度
文章中に出現する単語の頻出度を表しています。単語ごとに表示されている「スコア」の大きさは、与えられた文書の中でその単語がどれだけ特徴的であるかを表しています。



Voyantツール

シラース | コーバスターム | リンク | 本文 | TermsBerry | トレンド | ドキュメントの語

七草草紙そもへ正月七日に、野に出で、七草をつみて、みかどへ供御に備ふるといふなる由来を尋ぬ...

七草草紙そもへ正月七日に、野に出で、七草をつみて、みかどへ供御に備ふるといふなる由来を尋ぬに、もろこし楚國のかたはらに、大しうといふ者あり。かれは親に孝ある者なり。既に、はや百歳に及ぶ父母あり。膝などもかきみ、目などもかすみ、言ふことも聞えず。さるほどに若いければ、大しうこの朽ちはたる御姿を、見参らすたびに、歎き悲しむこと限りなし。大しう思ふやう

表示単語数: [スライダー]

このコーパスは1個のドキュメントを含みます。のべ語数 720 語、異なり語数 312 語。生成されたのはおよそ 10 分前。

語彙の密度 0.433

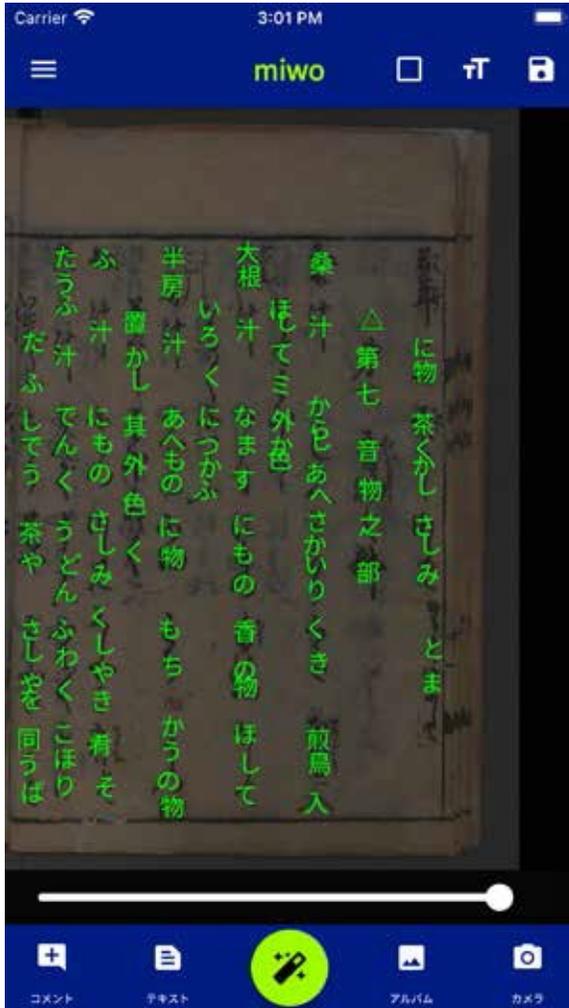
Readability Index: 6981.400

項目: [スライダー]

ドキュメント	左	単語	右
七草草紙そ...	に、野に出で、七草をつみ	て	、みかどへ供御に
七草草紙そ...	うこの朽ちは	て	たる御姿を、見参
七草草紙そ...	ふたゝび若くなし	て	たび給へ」と、仏
七草草紙そ...	に転じかへ	て	たび給へ。わが身は

Voyantツール、Stefan Sinclair & Geoffrey Rockwell (© 2024) プライバシー v 2.6.12

<https://textmining.userlocal.jp/>



くずし字認識



Bounding Box 表示



認識結果修正機能



む此約にたがふものならば賢第吾を何ものとかせんとや
たすら思ひ洗のども通りの方なしめしふの人のいふ人
一百に千りをゆくをあたはず魂よく一日に千りをもゆく
と此とわりを思ひやてみづから刃に伏今夜陰風に乗ては
る 来り菊花の約に世この心をあよれみ如へといひをはり
て涙わき出るがやし今は永きわかれなり只母公によく
つかへ治へとて座を立と見しがかき消て見えずなりにけ
る左門院廿とゞめんとすれば法風に眼くらみて行なをし
らず俯向につまづき倒れたる御に声を放て大に哭く老
母同さめ驚き立て左門がある所を見れば座上に酒瓶魚盛
たる皿ともあまた列へたるが中に臥州れたるをいそがは
しく扶起していかにととへとも只声を着て泣く 御らに言
なし老母問てふ伯氏亦穴が約にたがふ

コピー

テキスト出力

<https://21dzk.l.u-tokyo.ac.jp/SAT/>

大正新脩大藏經 テキストデータベース

संगणिकीकृतं तैशोत्रिपिटकं
Saṃgaṇikīkṛtaṃ Taiśotripitaṅkaṃ
The SAT Daizōkyō Text Database

ホーム 検索 ご挨拶 組織 利用条件 使い方 English History

大正蔵検索

AND 画像DB
 Punctuation Hangul Eng Search [settings]

Index Tools Image

TextNo. Vol. Page
 Jump

- 阿含部
- 本縁部
- 般若部
- 法華部・華嚴部
- 寶積部・涅槃部
- 大集部
- 經集部
- 密教部
- 律部
- 釋經論部
- 毘曇部
- 中觀部・瑜伽部

Introduction はじめに

Introduction

Right-side window

Two windows that appear at the right allow the reader to retrieve matches text from within the content of the D Buddhism (DDB) and BDK-SAT parallel. These windows can be moved freely area of your browser, and can be closed as necessary.

Left-side tabs

There are three tabs on the left side and **Image**.

The **Index** tab allows you to display

Digital Dictionary of Buddhi...

電子佛教辭典
 パスワードがない場合は「guest」でログインしてください。
 Users who do not have a password can log

JapanKnowledge 例文 仏教語...

※本文をドラッグすると
 JapanKnowledge 例文 仏教語大辞典の見出し語を検索して、ドラッグした文字列に含まれる見出し語と読み・辞書項目へのリンクが表示されます。
 ※ 辞書項目の内容を閲覧するには Japan Knowledgeの契約が必要です。リンク先から契約方法についての説明をご参照ください。（ただし、機関契約をしている機関からはそのままアクセスできます）
 ※ この仕組みは、有用性の高い有料コンテンツと無料コンテンツとを効果的に連携させた教育研究環境の構築を目的とする。

すべて表示



9784909658142_b.jpg ファイルを開く

268e76f851e1aba8b...jpg ファイルを開く

9784909658586.jpg ファイルを開く

次世代デジタルライブラリー

全文から検索する

画像から検索する

「次世代デジタルライブラリー」は、国立国会図書館次世代システム開発研究室での研究を基に開発した機能を実験的な検索サービスです。国立国会図書館デジタルコレクションで提供している資料の中から、著作権の保護期間が満了した図書及び古典籍資料全部（約33万6千点）が検索可能です。

以下の機能を備えています。

- 「全文から検索する」では、OCRにより生成された全文テキストから資料を検索できます。令和4年3月現在の対象資料は、令和3年度デジタル化資料のOCRテキスト化事業の成果物を利用した図書資料約28万点分です。ヒットした部分はスニペット形式で表示されるほか、資料画像の上に一致箇所と前後のスニペットがピンで表示されます。
- 「画像から検索する」では、図書及び古典籍資料中から自動的に切り出された約860万点の画像・図版を検索できます。絵を選ぶことで、似たような画像を含む資料を探すことができます。
- 切り出された画像・図版には、機械学習により自動的にタグが付与され、タグを利用した絞り込み検索が可能です。

サービスURL

<https://lab.ndl.go.jp/dl/>
ソースコード

<https://github.com/ndl-lab/tugidigi-web/>

次世代デジタルライブラリー

画像から検索する

サンプル画像から

資料のタイトルや目次から

探したい画像をドラッグアンドドロップしてください。類似画像を検索できます。

© 2019- 国立国会図書館 次世代システム開発研究室

EN



近代日本文学大系 第8巻
(名作浄瑠璃集上)

書誌 目次 本文 図表

御所蔵

45コマ見つかりました。

1 2 3 < >

9コマ

目次鬼一法眼三巻巻二一四〇〇蓬屋道高大内蔵四〇一四八六新撰徳川夜討四八七五七一逆崎松矢船梅ひらがな盛衰記五七三六六菅原傳授手習鑑六八七七六三犬物船矢倉義經千本櫻七六五八八六二吉野花矢倉假名手本忠臣蔵八六三

24コマ

進行方向 (自動推定)

見開きで自動分割する

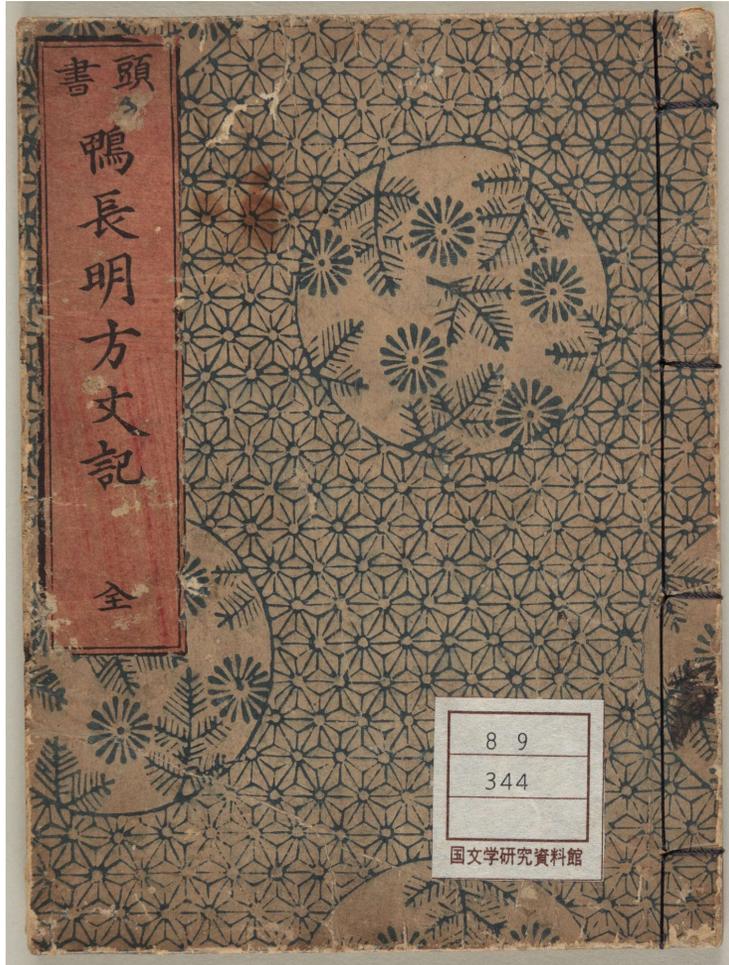
ページを白色化する

読みやすくなる

調整する

この資料の全文テキストデータ

© 2019- 国立国会図書館 次世代システム開発研究室



国文研「表紙文様集成」

www.nijl.ac.jp/pages/images/hyousimonyou.pdf



○○地に△△文様という記述

麻の葉地に小菊と若松の丸散らし文様

