

データ流通市場から見た 研究データへの期待

株式会社日本データ取引所 事業企画部 部長 上島邦彦

(English is on page24-46.)

本日の話題 (10min)

- 自社紹介
- ライセンスについて考える前に
- 弊社から見た研究データとの接点
- 産業界における検討と工夫
- 終わりに

(English is on page24-46.)

データ流通市場における役割

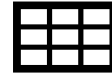
加工レベルでみた
データの種類



観測値



補正值・推計値



集計・統計表



図表・グラフ



論文・レポート

データ提供に
関わる人物



データ主体



データ取得者



データ制作者



データ出版者
(カタログ作成者)

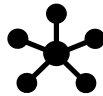


カタログ管理者
(販売代理者)

データ移転に関
わる人物



端末事業者



ネットワーク
提供者



データ保管者



IoTプラット
フォーム事業者



分析・可視化
アプリ提供者

データ利用に関
わる人物



データ品質
評価者



流通サービス
提供者



ブローカー
(購入代理者)



データ消費者



出資者

市場運営に関
わる人物



メディア・
情報提供者



取引システム
管理者



監査・監視者



決済・清算
システム提供者



取引履歴
管理者

研究データも、取扱いを準備中……。



データ流通とプライバシー



世論・流行・意識調査



広告・メディア・芸術



ジャーナリズム・図書情報



出産・子育て・教育



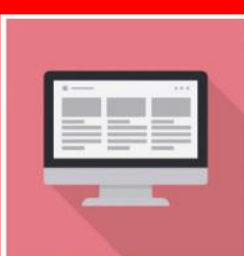
家庭と暮らし



病気と健康



住まいと安全



分析技術・学術情報



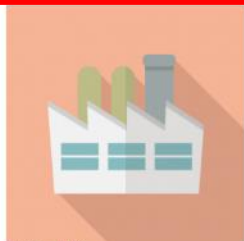
政治とお金



企業経営とビジネス情報



経済指標・エネルギー



製造・工業



地理・道路交通・自動車



農林水産・食糧生産



気象・地球環境・宇宙

ライセンスについて考える前に

「契約」のカジュアルダウンが進む

契約実務は、スマートに、ローコストに

- ネットネイティブが勤労世代に（LINE PAY, メルカリ人気）
- リーガルデザイン、リーガルハックの台頭（例：クラウドサイン）

増えつつける「利用規約に同意する」

- 受信箱に大量の「プライバシーポリシー改訂のお知らせ」
- データ分析ツールを試すたびに、利用規約を読むことに？
（API管理ツール、セルフサービスBI、機械学習…etc.）

「自由 vs. 慣習」を上手く調和させたい

契約自由の原則

- 契約するかどうかを決める自由
- 相手を選ぶ自由
- 内容決定の自由
- 方法の自由

商慣習の醸成

- 対話・交渉コストは減らしたい
- 事実上の選択肢
- 悪意・過誤の排除
- 標準契約、定型文

どこまでオープンにできるかは場合による

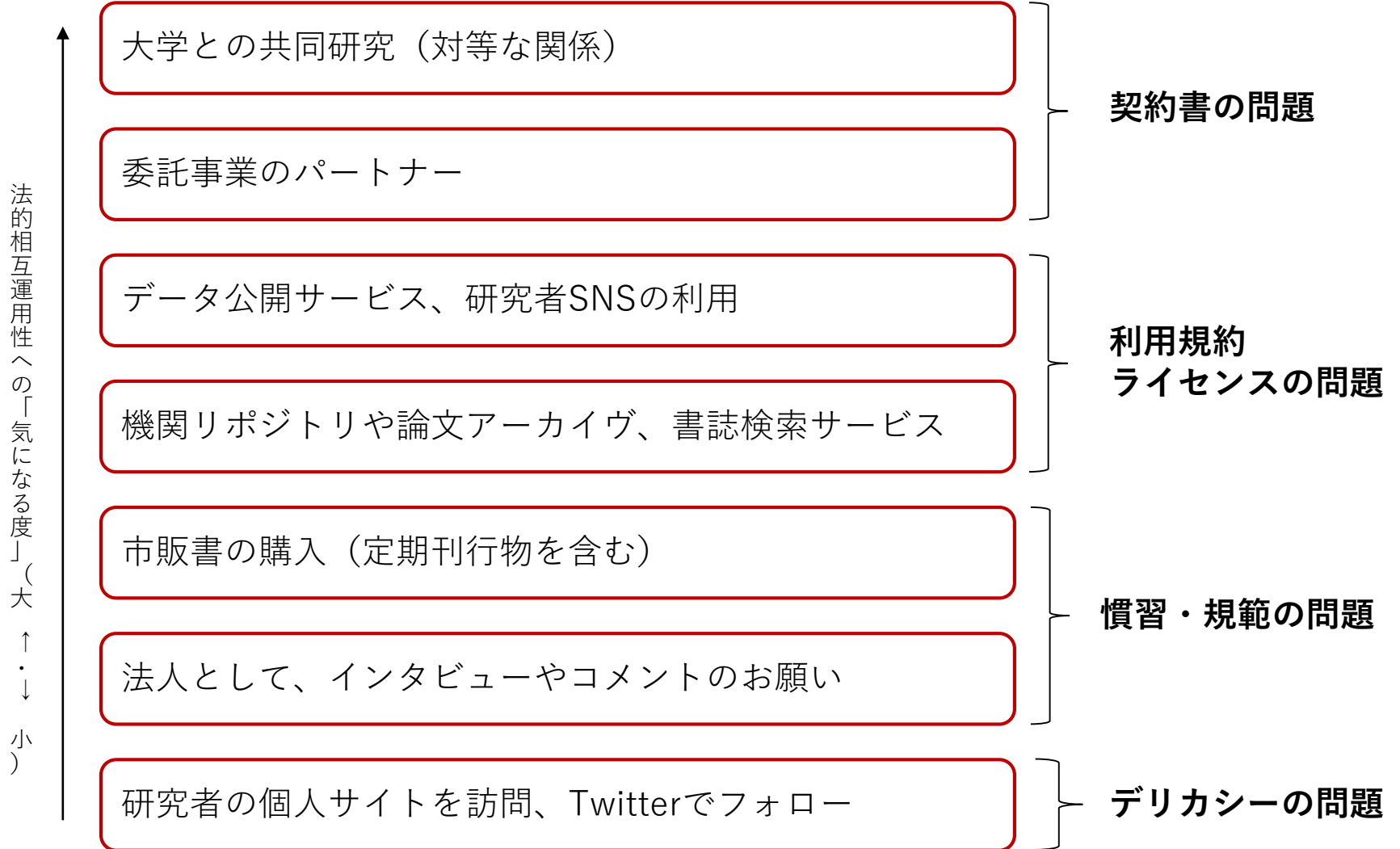
秘匿性の高さ(大↑・↓小)

性質	どんなデータか	どう流通しうるか
Private (私的な)	<ul style="list-style-type: none"> 法人/国家の秘密情報 個人のプライバシー 	<ul style="list-style-type: none"> 秘密保持と人材受入 明示的な本人同意
Available (利用できる)	<ul style="list-style-type: none"> 契約・交渉の対象 営利と非営利の間 	<ul style="list-style-type: none"> データ自体の流通 利用権限や書誌の流通
Open (公表された)	<ul style="list-style-type: none"> 普及・浸透の対象 公表 ≠ 権利放棄 	<ul style="list-style-type: none"> 公衆送信可能化 引用、転載、要約
Public (公共財)	<ul style="list-style-type: none"> 誰もが知るべきこと 負担と便益を分け合う 	<ul style="list-style-type: none"> オープンデータ、パブリックドメイン 匿名の告発

👉 研究データも、性質ごとにうってつけの「オープン」があるはず。

弊社から見た、研究データとの接点

権利帰属・ライセンス表示が気になるとき



産業界における検討と工夫

それぞれのデータ、それぞれの悩み

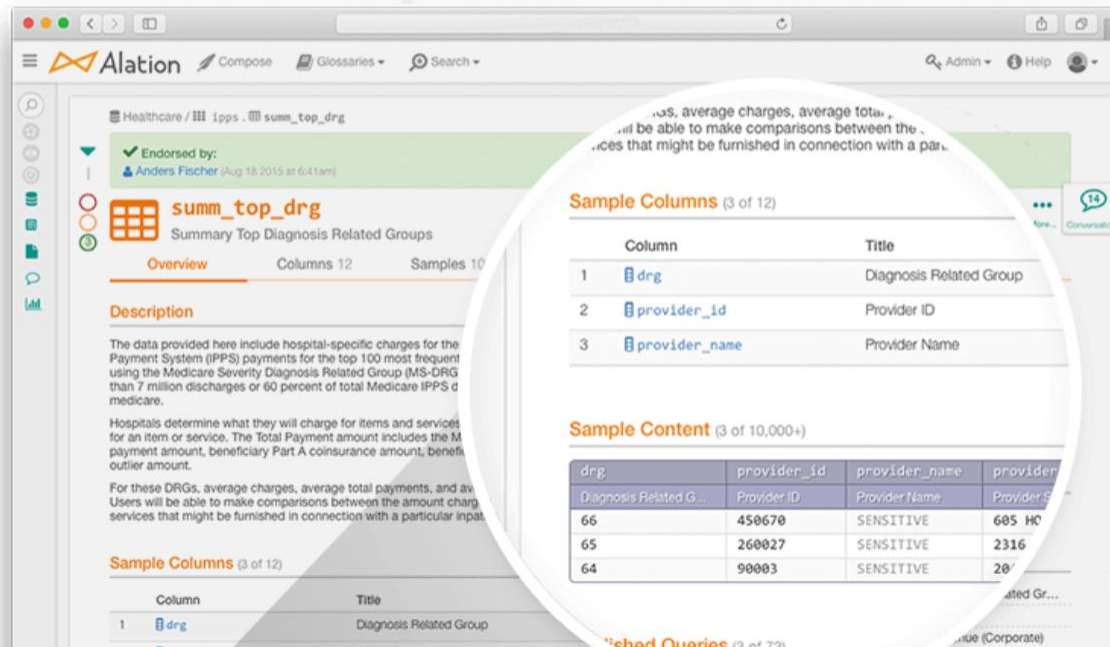
性質	よくある悩み
Private (私的な)	<ul style="list-style-type: none"> セキュリティ確保のため、データが部外持出できない 明示的な本人同意を得るための労力が割けない
Available (利用できる)	<ul style="list-style-type: none"> 権利許諾が煩雑で、義務・責任が曖昧になりがち データ単体では儲からないが、サービス化も大変
Open (公表された)	<ul style="list-style-type: none"> 公表準備や問合せ対応、ページの管理・更新で泥沼 偏った解釈や要約をされて、あらぬ誤解を招いた
Public (公共財)	<ul style="list-style-type: none"> 身銭を切る価値・意義が、経営陣の腑に落ちない FAQ「儲かるの?」「役に立つの?」

秘匿性の高さ(大 ↑ ・ ↓ 小)

さまざまな工夫 (1/7)



Have One Trustworthy Source of Data



What schema is this in?
What columns are available?

Automated Data Catalog

Alation indexes all of your data for business use. Technical metadata is surfaced where it is most relevant to the data consumer - in-context with data definitions and sample data. And machine learned patterns of usage are surfaced alongside the technical details,

👉 データの概要情報だけを検索・管理できるAlation

さまざまな工夫 (2/7)

The screenshot shows the data.world website interface. At the top, there is a search bar with the text "Search data.world" and a user profile for "kunihiro ueshima @ueshima". The main content area is divided into sections: "Recently added in environment" featuring a dataset titled "SGS-LTER Long-Term Monitoring Project: Vegetation Co... 1999" with an "OPEN" button, and "Trending this week" featuring a dataset titled "2018/W23: U.K. Gender Pay Gap" with an "OPEN" button. On the right side, there are sections for "RECENTLY UPDATED (0)", "TOP TAGS" (listing tags like education, funding, pocillopora, statbank, census with counts), and "RECENT DISCUSSIONS" (listing discussions like "2018/W23: U.K. Gender Pay Gap" and "Viz Review"). The left sidebar contains navigation options like "Home", "Library", "Bookmarks", "Requests", and "EXPLORE" with sub-items like "Features Overview", "Tutorials & Docs", "data.world Forum", "Who to follow", "Integrations", and "Invite people you know".

👉 タイムライン画面でデータ投稿できるdata.world

さまざまな工夫 (3/7)

Microsoft is acquiring GitHub! [Read our blog](#) and [Satya Nadella's post](#) to learn more.

Features Business Explore Marketplace Pricing

Search GitHub

Sign in or Sign up

How developers work

Support your workflow with lightweight tools and features. Then work how you work best—we'll follow your lead.

[New to GitHub? See how it works](#)



Code review



Project management



Integrations



Team management



Social coding



Documentation



Code hosting

👉 プライベートリポジトリ (有料) が作れる設計図共有サイト GitHub

さまざまな工夫 (4/7)



PRODUCTS

PRICING

CUSTOMERS

RESOURCES

EVENTS



FREE TRIAL

SIGN IN

DOCS

- Get Started
- Code Samples
- Policy Reference
- API Proxy Cookbook

SUPPORT

- Forums
- Support Portal

DISCUSSIONS

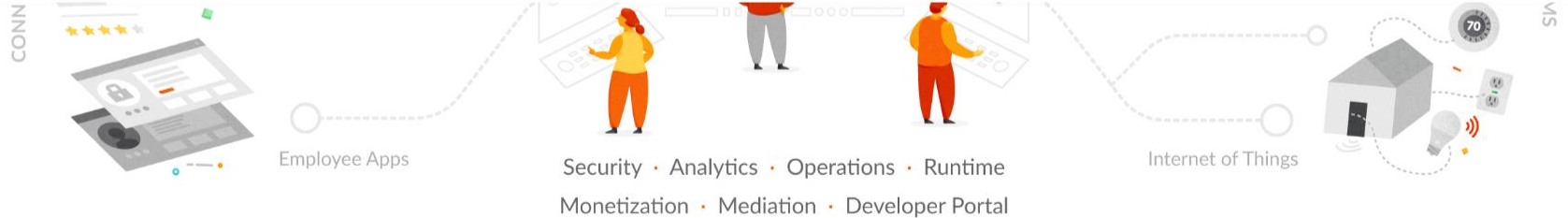
- Business of APIs
- General
- New Customers

BLOG

- Digital Business
- API Technology

GALLERY

- Ebooks & Reports
- Customer Stories
- Webcasts
- Videos



LEARN MORE

👉 API管理やセキュリティ設定を代行してくれるApigee

さまざまな工夫 (5/7)

We use cookies on kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using kaggle, you agree to our use of cookies.

Got it

Learn more

kaggle

Search kaggle



Competitions

Datasets

Kernels

Discussion

Learn



Sign In

Welcome to Kaggle Datasets

The best place to discover and seamlessly analyze open data



Discover

Use the search box to find open datasets on everything from government, health, and science to popular games and dating trends.



Explore

Execute, share, and comment on code for any open dataset with our in-browser analytics tool, [Kaggle Kernels](#). You can also download datasets in an easy-to-read format.



Create a Dataset

Contribute to the open data movement and connect with other data enthusiasts by clicking "[New Dataset](#)" to publish an open dataset of your own.

Learn More

New Dataset

👉 分析者コミュニティにデータを出版できるKaggle

さまざまな工夫 (6/7)

BOOTH

商品検索...

ログイン

お買い物ガイド

作品を出品したい



商品登録する



物販商品（自宅から発送）

- 商品名を入力
- 商品情報を入力
- 価格と在庫数を入力
- 商品写真をアップロード

イラスト集販売と同時にPSDデータを購入者に提供するなど、ダウンロードデータをあわせて提供できます。



物販商品（倉庫から発送）

- 商品名を入力
- 商品情報を入力
- 価格を入力
（在庫数は倉庫サービス申込み時にご入力ください）
- 商品写真をアップロード

倉庫サービスについて
[倉庫の利用方法等はこちら](#)



ダウンロード商品

- 商品名を入力
- 商品情報を入力
- 商品写真をアップロード
- 商品データをアップロード
（合計10GBまで。常時差し替え可能。）

ダウンロード商品は、0円も設定できます。サンプルや体験版の無料配布等にご利用いただけます。

👉 物販とデータ管理を一緒にできるBOOTH

さまざまな工夫 (7/7)

BCCKS		よむ all books	書店 book store	つくる create books	検索	0 login 会員登録 BCCKSって?
EPUBの発行	データ本を作成することで、EPUB 3データも作成することができます。	○	○			
EPUB プロモーション頁の削除	EPUBの2頁目に必ず入るBCCKSのプロモーション頁を取ることができます。※奥付は変更不可	×	○			
EPUB ダウンロードの禁止	データ本を公開していても、自著のEPUBのダウンロードを禁止することができます。	×	○			
EPUB ソーシャルDRM	自著のEPUBに、ダウンロード時の利用者のメールアドレスを埋め込む形のソーシャルDRMを付与できます。	×	○			
ストア配本	BCCKSのエディタで作成されたEPUBデータを、主要電子書籍ストアに配本します。 1 申請/更新に5 チケット、削除に3 チケット必要です。	○ チケット利用	○ チケット利用			
印税シェア設定	本の印税を、共有済みの他のBCCKS会員に分配することができます。	×	○			
印税額TSVダウンロード	集計方法と期間を選択して、印税額の履歴データをTSV形式でダウンロードできます。	×	○			
無料クーポンコード発行	データ本を販売した際、献本に便利なクーポンコードを無料で発行できます。	×	○ 20コード/1冊			
有料クーポンコード発行	有料のデータ本を1冊入手できるクーポンコードを1コードあたり販売価格の30%で取得できます。	×	○			

 電子書籍のコピーガードと印税シェアができるBCCKS

国内外の法的相互運用性に関する検討

- **経済産業省「AI・データの利用に関する契約ガイドライン」**
 - データの利用権限に関する論点整理と、標準ひな型案の提示
 - 不正競争防止法、著作権法の改正・施行と並行して
- **W3C「Data Catalog Vocabulary」**
 - データの概要情報を整理する書式の項目を整備
 - ライセンス・権利の項目がある
 - 「Use Case & Requirement」には、研究データ関連の要望も
- **データ流通推進協議会 技術基準検討委員会**
 - データカタログガイドラインの作成
 - 検討対象は、データカタログ、メタデータ・API、データ品質

終わりに

諸論ありますが、詳しくは討議の時間に…

- サービス利用規約を読む／読ませることを基本としたいが……
 - 万人に、関連法の知識を求めるのは、無理筋
 - 簡潔なライセンス表示で、「はじめての理解」を促したい
 - 定型の免責事項を、研究コミュニティ内で共用するのも一手
- 標準契約書を起草・改訂するとき、データ取扱いやライセンス表示にも留意を
 - 今後は、データの取扱いも、明記いただいたほうがいいのかも
 - ライセンス表示、利用規約、標準契約のセットがあると、法務確認が捗る
 - 理想は、学術分野ごとのデファクト（デジュールでなくてもいい）
- 懸念は、「サービス管理者がやるべき仕事」になりがち問題
 - 研究者（個人）に義務・責任を負わせるのは酷かも
 - 機関・組織が担うとして、「どの法人が信頼できるか」は産業界でも課題
 - 専門人材・システム予算の費用と、期待される収益のトレードオフ

どちらのシナリオに現実味を感じますか？

嬉しい将来

- 標準的な規約・ライセンスの普及が進み、産学官民の法的相互運用性が相応に確保され、データ流通を促す商取引文化が醸成される。
- 結果として、企業・行政が学术界に投資を行いやすくなり、実務者が研究や分析に専念する時間も増える。

悲しい将来

- 産学官民ごとの都合・事情が優先され、相互運用性の十分な確保に挫折。法人（に所属する個人）の契約・交渉コストが肥大化。
- 結果として、共同研究の予算に占める法務費用の割合が増えたり、実務者が自由に使えるお金と時間が、さらに減ってしまう。

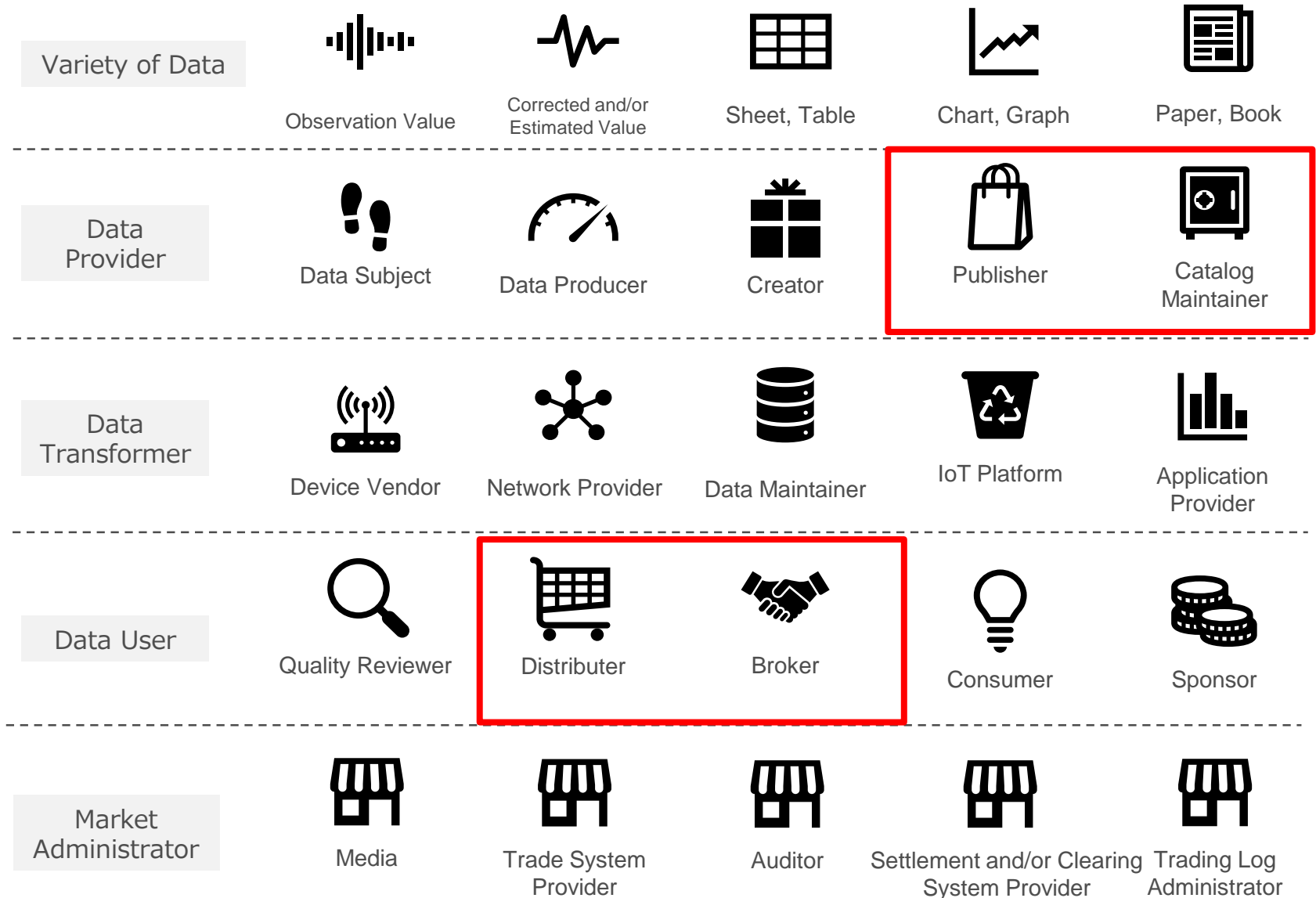
Expectation to Research Data From **Data Market**

Japan Data Exchange, corp

Topic (10min)

- **About us**
- **Before You think about License**
- **When We Concern about Rights and Licenses**
- **Try and Error in Business**
- **Ending**

Role of us in Data market



Handling Valuable Data (Including Research Data)



データ流通とプライバシー



世論・流行・意識調査



広告・メディア・芸術



ジャーナリズム・図書情報



出産・子育て・教育



家庭と暮らし



病気と健康



住まいと安全



分析技術・学術情報



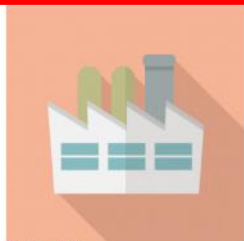
政治とお金



企業経営とビジネス情報



経済指標・エネルギー



製造・工業



地理・道路交通・自動車



農林水産・食糧生産



気象・地球環境・宇宙

Before You think about License

Trends: Contract's Casual Down

Getting Smarter, Cost less

- Generation Z enters the business (Spreading LINE PAY, Mercali…)
- Rise of Regal Design, Regal Hack (Ex: CloudSign)

Countless “Agree of Term of Service”

- Lots of "Notice of revised privacy policy“ in Mail box(GDPR!)
- Every time We try new tool, should we Every time “read” Term of Service? (API Managemtn, Self Service BI, Machine Learning…etc.)

Freedom or Custom, That is a question.

Freedom of Contract

- Whether Contract or not
- Whom to Contract
- Which Term to Contract
- How to Contract

Custom for Business

- Reduce negotiation Cost
- De facto Option
- Avoid malice and error
- Standard Contract

How far can Data be open

Types	What kind of Data	How We Publish
Private	<ul style="list-style-type: none"> Confidential Information Personal Privacy 	<ul style="list-style-type: none"> Non-disclosure Agreement, Acceptance of Researchers Explicit Agreement
Available	<ul style="list-style-type: none"> sometimes commercial at other times non-profit 	<ul style="list-style-type: none"> Trading Data it self Trading Rights or Catalog
Open	<ul style="list-style-type: none"> Hoped to the diffusion Publish, but not Copy-left 	<ul style="list-style-type: none"> Enable public Transmission Quote, Reprint, Summary
Public	<ul style="list-style-type: none"> Everybody can to know Share benefits and costs 	<ul style="list-style-type: none"> Open Data or Public Domain Anonymous Accusation

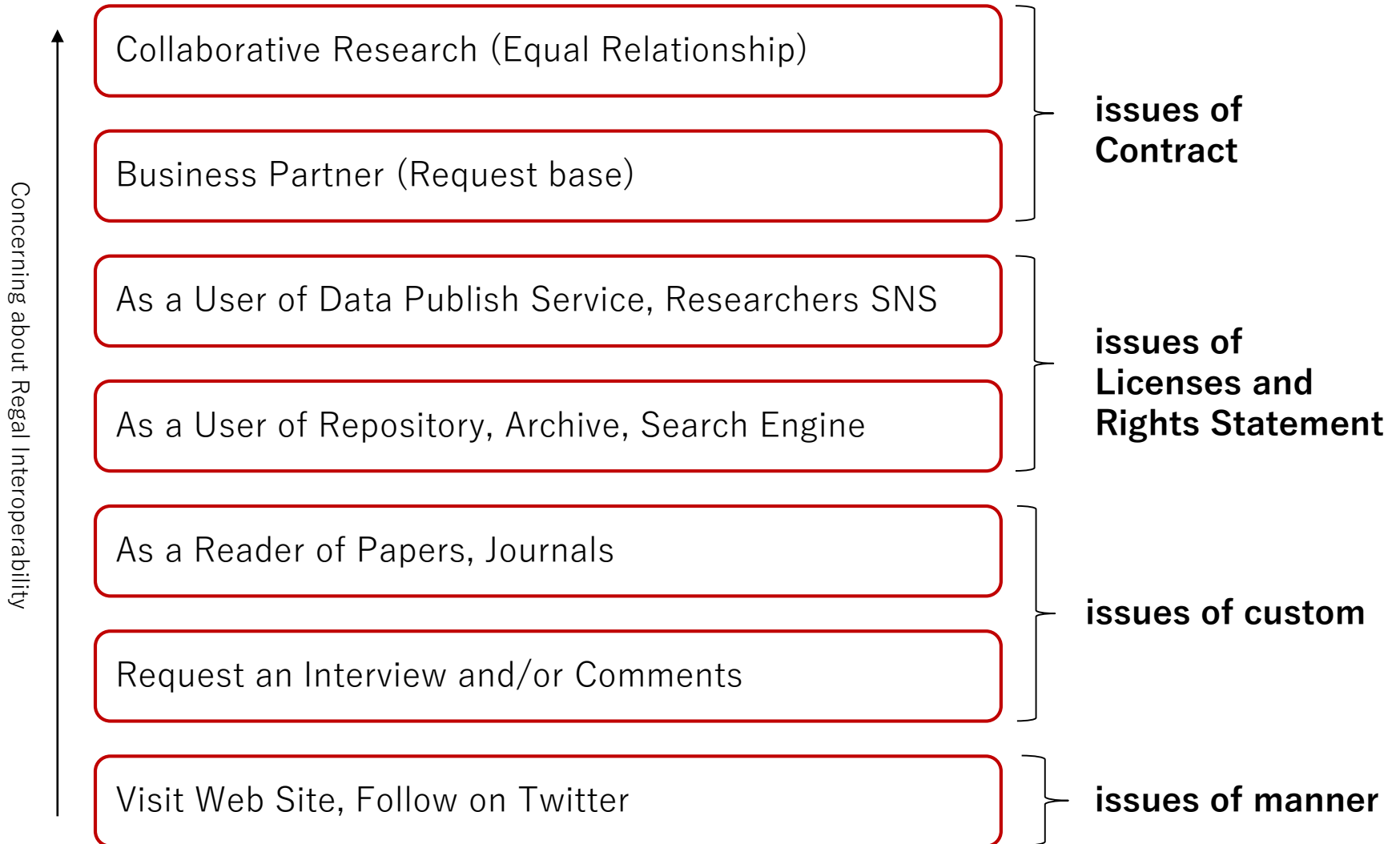
Concerning Privacy, Confidentiality



There must be best choices for any research data.

When We Concern about Rights and Licenses

When We Concern about Rights and Licenses



Try and Error in Business

Any Data, Some Trouble

Types	business sectors often worry about
Private	<ul style="list-style-type: none"> • Data can not be taken out to ensure security • Not afford to get “explicit” agreement
Available	<ul style="list-style-type: none"> • Rights handling is more and more complicated • difficult to monetize data itself, but not easy to convert it into a service
Open	<ul style="list-style-type: none"> • hard days of never ending write, edit, publish, notice, comment response, page update... • Biased interpretation, Wrong summary
Public	<ul style="list-style-type: none"> • Do We monetize it ? • Is it Useful ?

Concerning Privacy, Confidentiality

A Case of Try (1/7)



Have One Trustworthy Source of Data

The screenshot shows the Alation interface for a dataset named 'sum_top_drg'. A circular callout highlights the 'Sample Columns' and 'Sample Content' sections.

Sample Columns (3 of 12)

Column	Title
1 drg	Diagnosis Related Group
2 provider_id	Provider ID
3 provider_name	Provider Name

Sample Content (3 of 10,000+)

drg	provider_id	provider_name	provider...
Diagnosis Related G...	Provider ID	Provider Name	Provider S...
66	450670	SENSITIVE	605 HC
65	260027	SENSITIVE	2316
64	90003	SENSITIVE	20

What schema is this in?
What columns are available?

Automated Data Catalog

Alation indexes all of your data for business use. Technical metadata is surfaced where it is most relevant to the data consumer - in-context with data definitions and sample data. And machine learned patterns of usage are surfaced alongside the technical details,



👉 Data Catalog(Alation)

A Case of Try (2/7)

The screenshot shows the data.world website interface. At the top, there is a search bar with the text "Search data.world" and a navigation menu with "Add" and a notification badge with the number "2". The user profile for "kunihiko ueshima @ueshima" is visible on the left, with an "Add photo" button and a "View your profile" link. The main content area is divided into "Recent" and "Featured" sections. The "Recent" section displays a dataset titled "SGS-LTER Long-Term Monitoring Project: Vegetation Co... 1999" with an "OPEN" button. Below it, there is a "Trending this week" section featuring a dataset titled "2018/W23: U.K. Gender Pay Gap" with an "OPEN" button. On the right side, there are sections for "RECENTLY UPDATED (0)", "TOP TAGS" (listing tags like education, funding, pocillopora, statbank, census with their respective counts), and "RECENT DISCUSSIONS" (listing discussions like "2018/W23: U.K. Gender Pay Gap" and "Viz Review" with their respective counts).

☞ Timeline (data.world)

A Case of Try (3/7)

  Microsoft is acquiring GitHub! [Read our blog](#) and [Satya Nadella's post](#) to learn more.

X



Features

Business

Explore

Marketplace


Pricing

Search GitHub

Sign in or Sign up

How developers work

Support your workflow with lightweight tools and features. Then work how you work best—we'll follow your lead.

 [New to GitHub? See how it works](#)



Code review



Project management



Integrations



Team management



Social coding



Documentation



Code hosting



Private Repository (GitHub)

さまざまな工夫 (4/7)



PRODUCTS

PRICING

CUSTOMERS

RESOURCES

EVENTS



FREE TRIAL

SIGN IN

DOCS

- Get Started
- Code Samples
- Policy Reference
- API Proxy Cookbook

SUPPORT

- Forums
- Support Portal

DISCUSSIONS

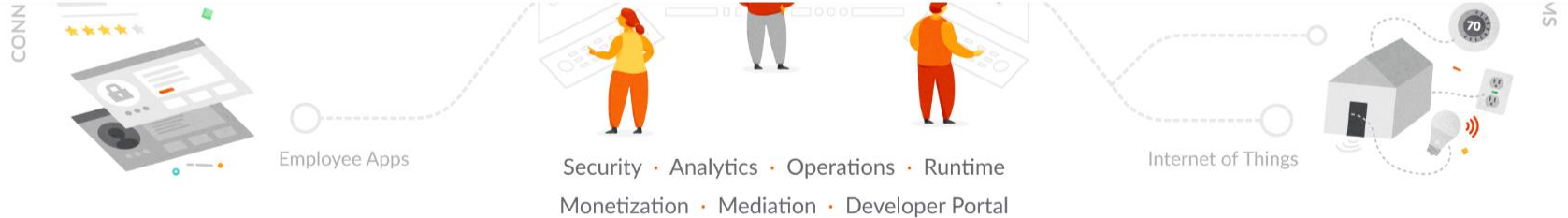
- Business of APIs
- General
- New Customers

BLOG

- Digital Business
- API Technology

GALLERY

- Ebooks & Reports
- Customer Stories
- Webcasts
- Videos



LEARN MORE

 API Management (Apigee)

さまざまな工夫 (5/7)

We use cookies on kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using kaggle, you agree to our use of cookies.

Got it

Learn more

kaggle

Search kaggle



Competitions

Datasets

Kernels

Discussion

Learn



Sign In

Welcome to Kaggle Datasets

The best place to discover and seamlessly analyze open data



Discover

Use the search box to find open datasets on everything from government, health, and science to popular games and dating trends.



Explore

Execute, share, and comment on code for any open dataset with our in-browser analytics tool, [Kaggle Kernels](#). You can also download datasets in an easy-to-read format.



Create a Dataset

Contribute to the open data movement and connect with other data enthusiasts by clicking "[New Dataset](#)" to publish an open dataset of your own.

[Learn More](#)

[New Dataset](#)



Direct Data Publishing (Kaggle)

さまざまな工夫 (6/7)

BOOTH

商品検索...

ログイン

お買い物ガイド

作品を出品したい



商品登録する



物販商品（自宅から発送）

- 商品名を入力
- 商品情報を入力
- 価格と在庫数を入力
- 商品写真をアップロード

イラスト集販売と同時にPSDデータを購入者に提供するなど、ダウンロードデータをあわせて提供できます。



物販商品（倉庫から発送）

- 商品名を入力
- 商品情報を入力
- 価格を入力
（在庫数は倉庫サービス申込み時にご入力ください）
- 商品写真をアップロード

倉庫サービスについて
[倉庫の利用方法等はこちら](#)



ダウンロード商品

- 商品名を入力
- 商品情報を入力
- 商品写真をアップロード
- 商品データをアップロード
（合計10GBまで。常時差し替え可能。）

ダウンロード商品は、0円も設定できます。サンプルや体験版の無料配布等にご利用いただけます。



E-commerce with Data upload service (BOOTH)

さまざまな工夫 (7/7)

BCCKS		よむ all books	書店 book store	つくる create books	検索	0 login 会員登録 BCCKSって?
EPUBの発行	データ本を作成することで、EPUB 3データも作成することができます。	○	○			
EPUB プロモーション頁の削除	EPUBの2頁目に必ず入るBCCKSのプロモーション頁を取ることができます。※奥付は変更不可	×	○			
EPUB ダウンロードの禁止	データ本を公開していても、自著のEPUBのダウンロードを禁止することができます。	×	○			
EPUB ソーシャルDRM	自著のEPUBに、ダウンロード時の利用者のメールアドレスを埋め込む形のソーシャルDRMを付与できます。	×	○			
ストア配本	BCCKSのエディタで作成されたEPUBデータを、主要電子書籍ストアに配本します。 1申請/更新に5チケット、削除に3チケット必要です。	○ チケット利用	○ チケット利用			
印税シェア設定	本の印税を、共有済みの他のBCCKS会員に分配することができます。	×	○			
印税額TSVダウンロード	集計方法と期間を選択して、印税額の履歴データをTSV形式でダウンロードできます。	×	○			
無料クーポンコード発行	データ本を販売した際、献本に便利なクーポンコードを無料で発行できます。	×	○ 20コード/1冊			
有料クーポンコード発行	有料のデータ本を1冊入手できるクーポンコードを1コードあたり販売価格の30%で取得できます。	×	○			

 Social DRM, Revenue Sharing of e-book(BCCKS)

Regal Interoperability

- **METI “ Contract Guidelines on the Use of AI and Data”**
 - Discuss on authority of data and suggest a standard model
 - In response to amendment and enforcement of the Unfair Competition Prevention Act, Copyright Law
- **W3C “Data Catalog Vocabulary”**
 - International Standard for Data Catalog
 - There are issues about License, Rights
 - I recommend to read “Use Case & Requirement”
- **Data Trading Alliance Technical Standard Committee**
 - Data Catalog Guideline (Now Creating…)
 - Scope: Data Catalog, Meta Data, API, Data Quality

Ending

Summary

- **Can't We recommend to read All the Terms of Service**
 - Impossible for everyone to seek expert knowledge
 - Encourage "understanding for the first time" with a clear license
 - It's good to create a standard disclaimer in each research community
- **Pay attention to data rights and license, When drafting or revising the standard contract about copyrights**
 - Please clearly state the handling of data (not only copyrights)
 - Business Legal will be happy if they can easily get 3 documents (licenses, terms of use, and standard contract)
 - It's nice there is a de facto each research community (de jure is unnecessary)
- **The problem is every-stakeholder tends to say "it's not my job"**
 - It's cruel to let the researcher take responsibility
 - It's hard to decide "who can trust", even if the organization plays the role
 - There is a trade-off between expenses of human resources and/or systems and expected earnings

Which scenario feels reality?

Happy end

- Standard terms · Licenses are becoming increasingly popular, ensuring legal interoperability of industry, academia, public and private sectors appropriately, and fostering custom that encourages open science.
- Companies and governments invest more in academia, researchers can devote more time to research

Unhappy end

- Stakeholders can't harmonize, interoperability can't be secured. The contracting and negotiation costs of the corporation are enlarged.
- More research budgets will be spent on legal affairs, which further reduces the money and time that researchers might use.