



# TEEDA

-Treasuring Every Encounter of Data Affairs-

## ワークショップ

東京大学 大学院工学系研究科 助教

早矢仕 晃章

# ポストコロナ社会における観測データ設計の集合知化

システム創成学専攻: 早矢仕晃章 助教, 上原直 (M2), 長谷大輔 (M1), 大澤幸生 教授

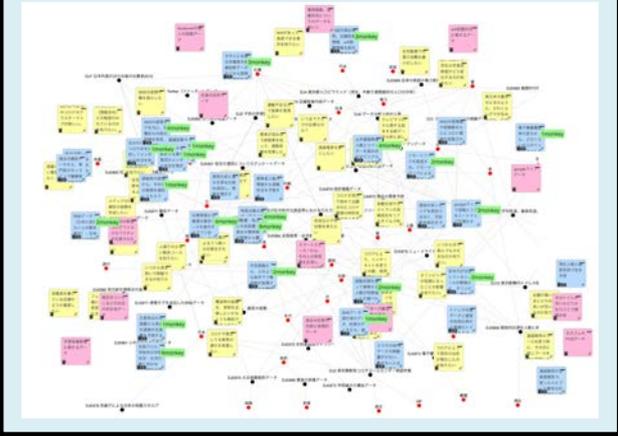
異分野データ連携とコミュニケーション支援によって、目的に応じたデータ設計・収集・共有と意思決定プロセスを支援



COVID-19現象の理解と対策のため、分野を横断したデータ流通・共有の重要性が理解されることとなったが、データの信頼性と質、そして未だデータ化されていない事象を如何に観測するのかという新たな問題が顕在化

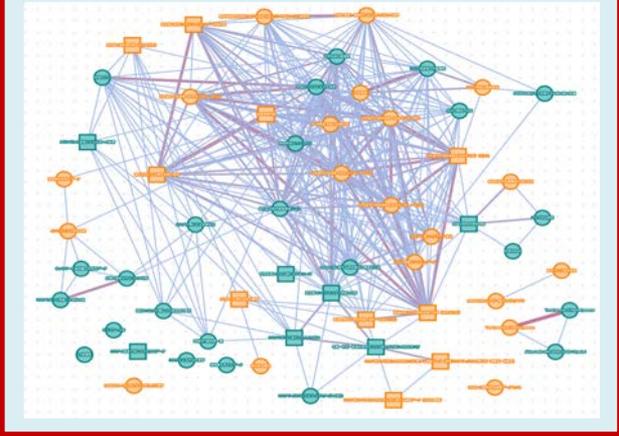
## Web IMDJ (Web-based Innovators Marketplace on Data Jackets)

コロナ禍におけるニーズの表出化とデータの組み合わせによる解決方法をオンラインで議論するプラットフォームを展開。



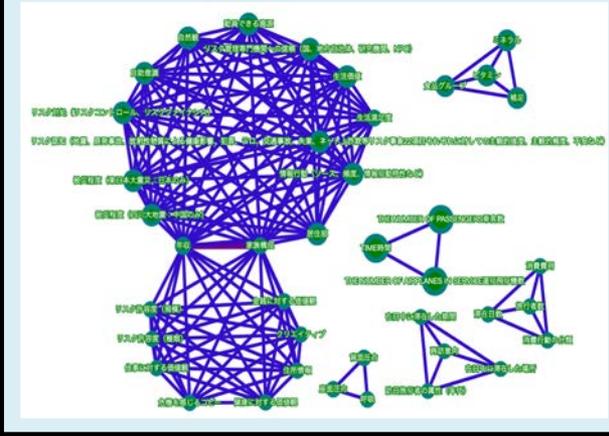
## TEEDA (Treasuring Every Encounter of Data Affairs)

コロナ禍で実際に共有されたデータと足りないデータの情報を集合知化し、未だデータ化されていない事象を検討可能に。



## VARIABLE QUEST

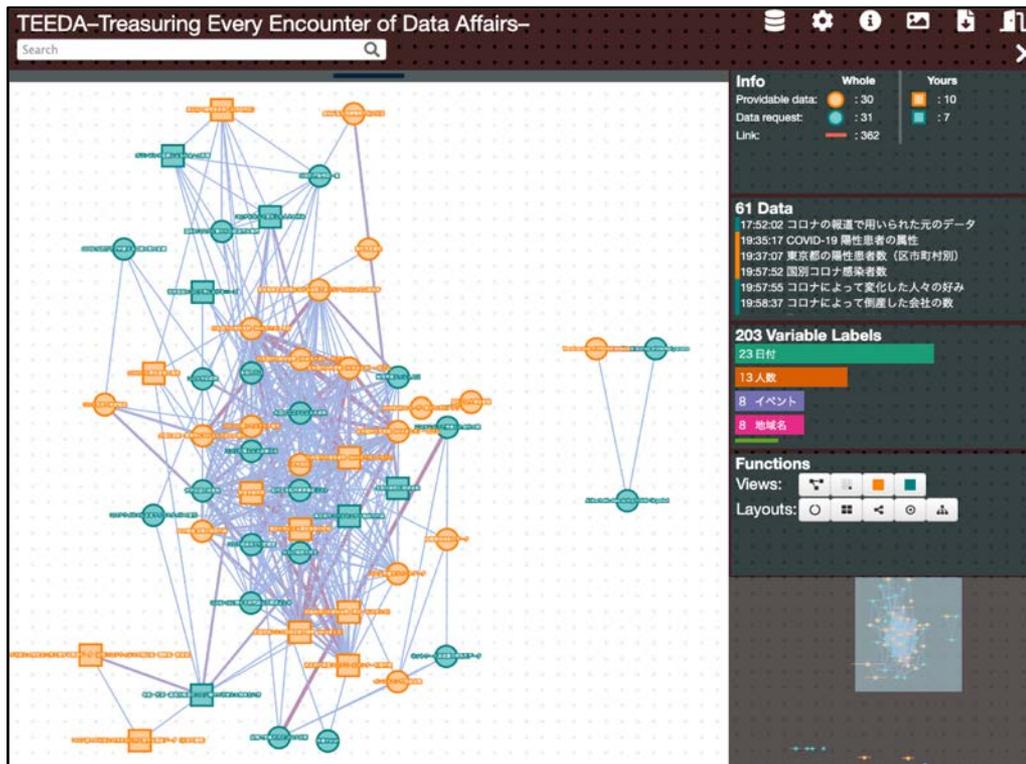
変数が未知のデータ概要から、当該データに含まれ得る変数を推定することで、未観測事象のデータ設計を支援。



# ワークショップで用いるツール

## TEEDA

–Treasuring Every Encounter of Data Affairs–



<http://teeda.data-marketplace.org/>

# データ

データ = 変数と値の集合

(例) スーパーマーケットの購買履歴データ

変数 (パラメータ、属性)

年	月	日	顧客ID	購入品目	支払金額
2017	11	1	AAAAA	人工知能学会誌	2592
2017	11	1	BBBBB	ペン、りんご、パイナップル	1080
⋮	⋮	⋮	⋮	⋮	⋮
2017	11	30	YYYYY	スナック、するめ、ビール	2536
2017	11	30	ZZZZZ	ラーメン、ナタデココゼリー	867

値

# データジャケット Data Jacket

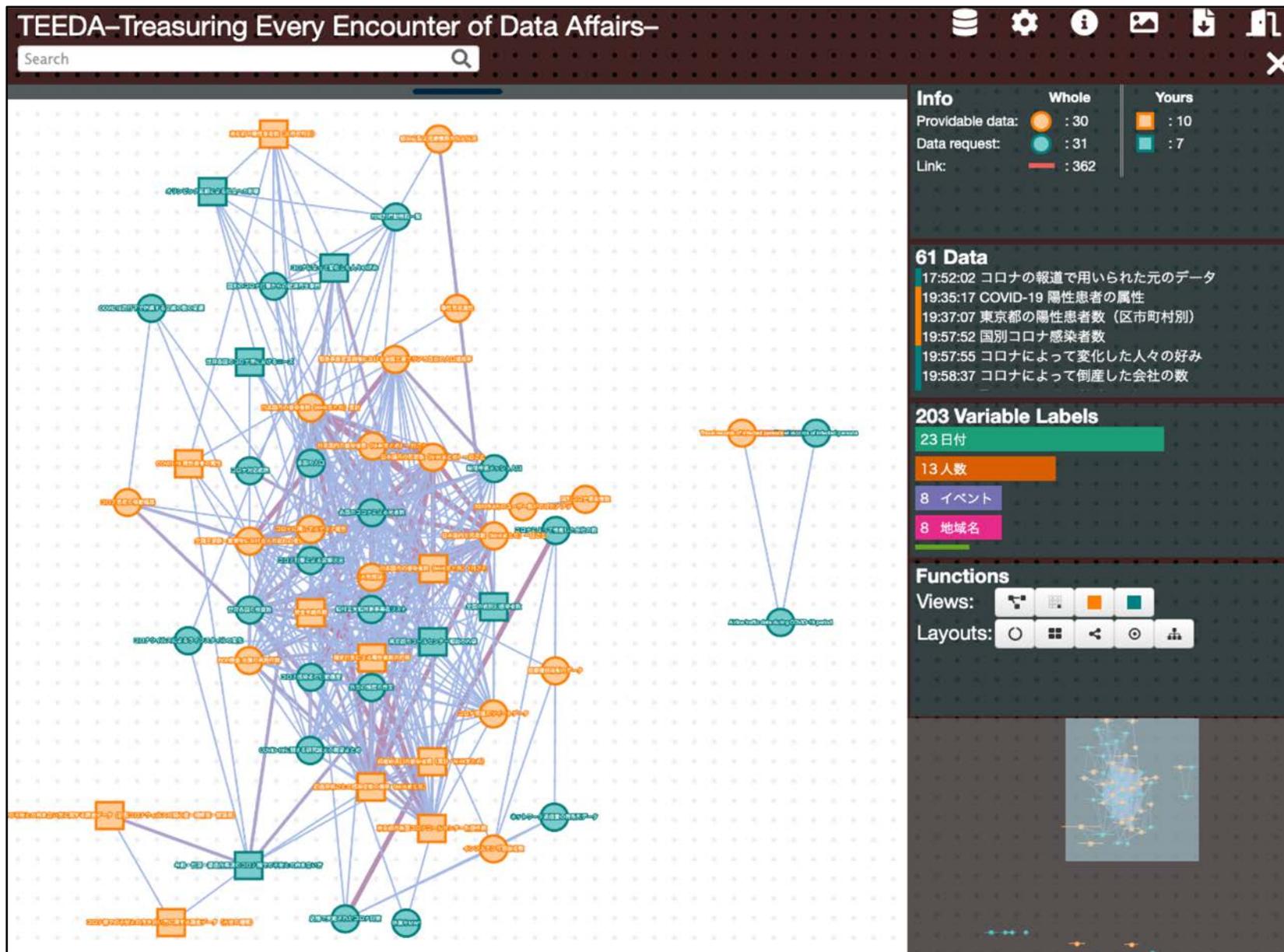
- データジャケット (DJ) は人間が読むことを前提としたデータの概要情報
- 自然言語で記述され、構造化されている
- データの中身ではなく、概要情報 (変数名、保存形式、収集方法など) を共有し、データの利用価値を検討可能にする
- 個人情報を含む共有不可能なデータでも、DJにすることでセキュリティ上のリスクを低減させて情報が共有可能となる
- 構造化によって、人間だけでなく計算機においても可読化

年	月	日	顧客ID	購入品目	支払金額
2017	11	1	AAA	人工知能学会誌	2592
2017	11	1	BBB	ペン、りんご、パイナップル	1080
⋮	⋮	⋮	⋮	⋮	⋮
2017	11	30	YYY	スナック、するめ、ビール	2536
2017	11	30	ZZZ	ラーメン、ナタデココゼリー	867



DJ No. XX【購買履歴データ】	
概要	東京都の〇〇スーパーマーケットで収集されている顧客の購買行動履歴。
収集方法・コスト	ポイントカードとPOSによって取得
共有条件	共有不可
データの種類	表形式、テキスト、数値
保存形式	CSV
分析・シミュレーション	時系列分析
変数ラベル	氏名、性別、顧客ID、支払金額、購入品目、日にち
分析結果	<ul style="list-style-type: none"><li>その日の売上の計算</li><li>今後の売上の予測と仕入れの推定</li></ul>
期待される分析	顧客の購買行動とリピート率を計算し、ロイヤルカスタマーの特定が可能かもしれない。
コメント	有効なデータの組み合わせが発見されればデータの提供あるいはコラボレーションもあり得る。

# TEEDAのインタフェース



※図はテストデータによる可視化

# TEEDAのインタフェース

ノード

リンク(エッジ)



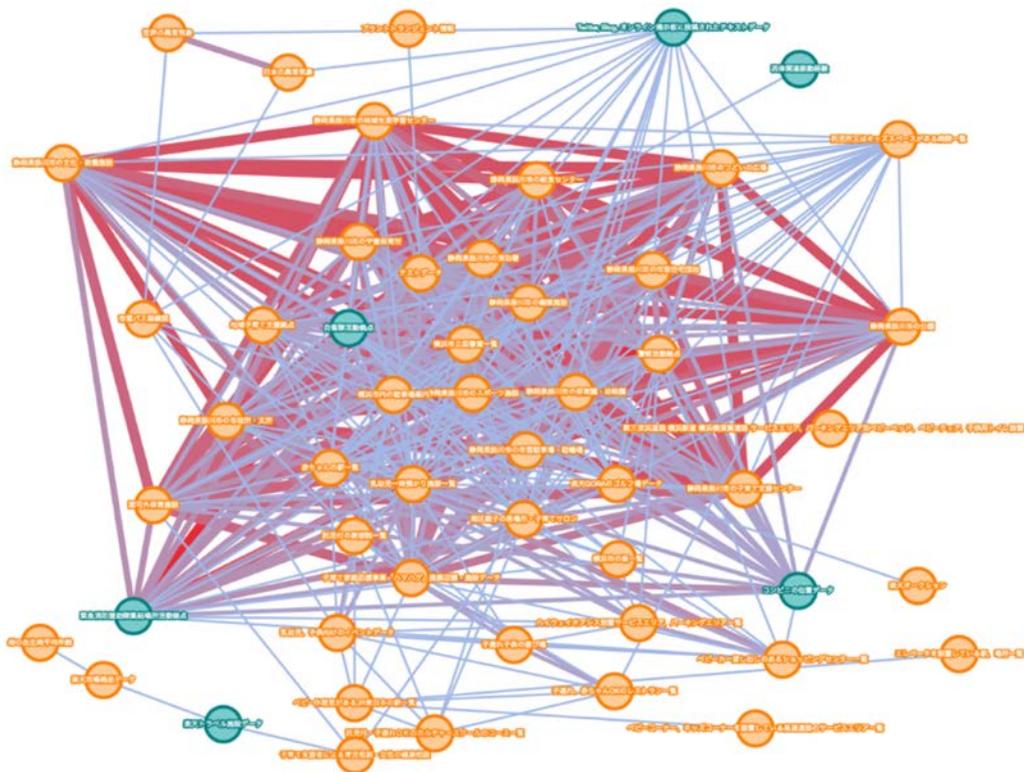
: 入手したいデータ



: 提供可能データ



: データ同士が共通する変数を有している場合にリンクを張る  
(共通する変数の数に応じて太くなる)



## ヒント :

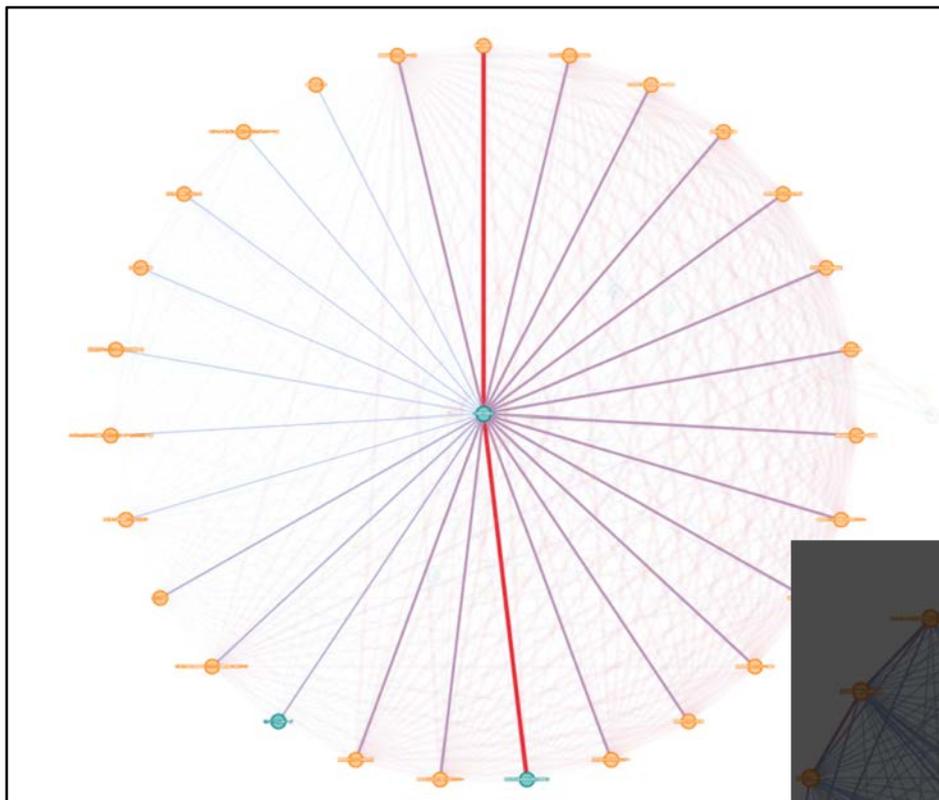
COVID-19現象の理解と対策のため、

- どのようなデータが必要なのか、
- どのようなデータが提供されているのか、

という情報を入力してください。

講演や研究発表を聞いて関心の高まったデータについて入力いただいても構いません。

# TEEDAのインタフェース



ノードをクリックすることで、  
変数で紐付いている関連データ  
との関係が理解できる

入力されたデータの詳細情報も  
閲覧可能

A screenshot of the TEEDA interface showing detailed information for a specific data point. The interface is in Japanese and includes a title, a summary, a list of variables, data types, and storage formats.

**自衛隊活動拠点**

**概要**  
自衛隊活動拠点の位置情報です（静岡県地域防災計画より）

**変数ラベル**  
市町名 名称 使途 所在地 緯度 備考 最終更新日 経度

**データの種類**  
数値 テキスト

**保存形式**  
CSV・XLSなど

**共有条件**  
一般的に共有してよい

※図はテストデータによる可視化

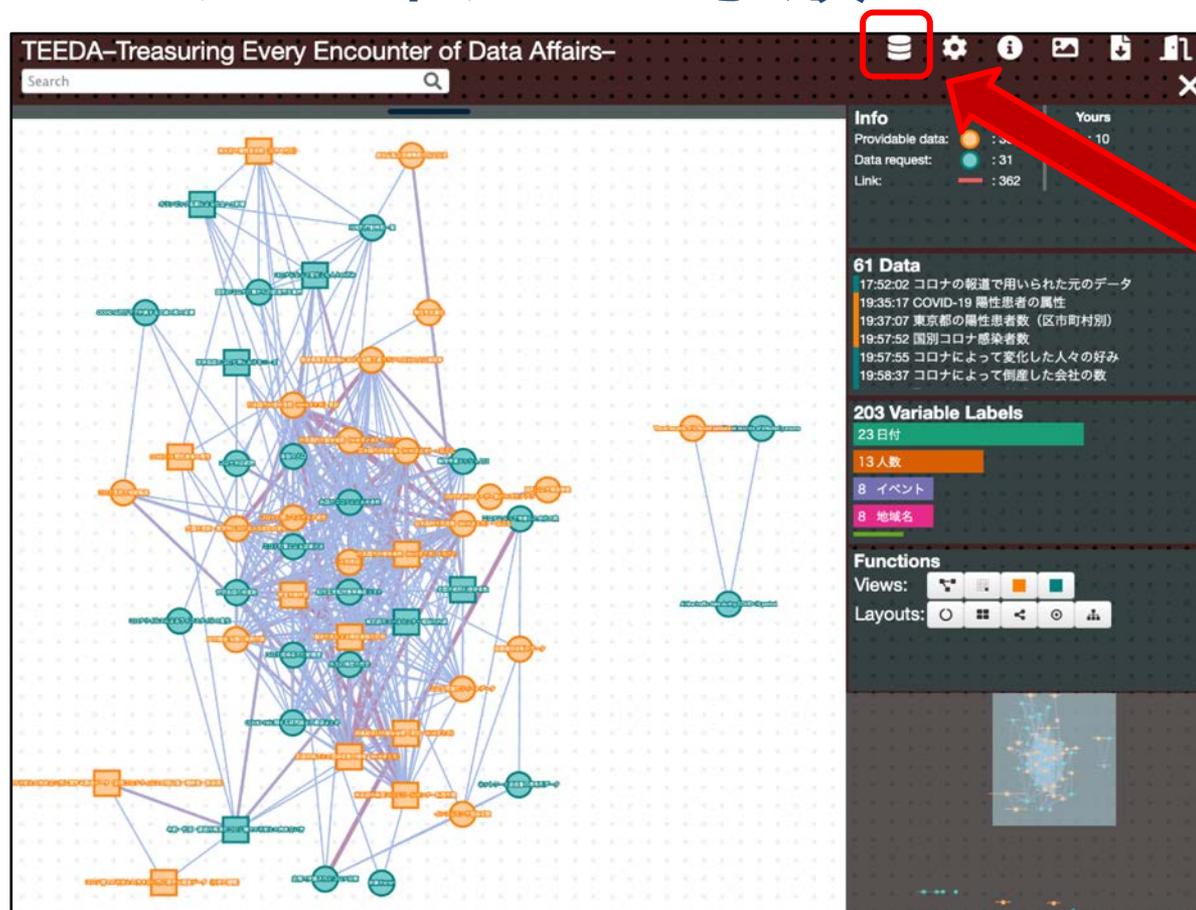


# ワークショップの手順

A screenshot of the TEEDA website's login interface. It features four main login options at the top: 'Go to Open TEEDA', 'Create New Room', 'Login', and 'Admin Login'. Each option has a corresponding form with input fields for 'Username', 'Password', and 'Room ID'. Below these is a 'Special Login for the Workshop in Systems Innovation' section, which is highlighted with a red rounded rectangle. This section includes the text '公開ワークショップ「ポストコロナ時代のシステム創成-シミュレーションとデータ-」用ログイン' and a 'Username:' input field. A large red arrow points upwards from the text below towards this 'Special Login' section. The background of the interface is dark with a grid pattern and some abstract light-colored graphics.

TEEDAサイト (<http://teeda.data-marketplace.org/>) にアクセス  
ユーザー名（自由にご記入ください）を入力し、Loginをクリック

# ワークショップの手順



データ入力アイコンをクリック

本ワークショップ中に各人で、以下のデータについての情報をTEEDAに入力してください。

- ① どのようなデータを入手したいか
- ② どのようなデータが提供可能であるか

# ワークショップの手順

## Data Registration

Thank you for your participation in TEEDA Workshop.  
TEEDA stands for "Treasuring Every Encounter of Data Affairs."  
In this workshop, we will visualize the data you want to obtain and the data you can provide.  
Please discuss in the group and enter information on the data.

---

**Which data do you register?**

data request

providable data



data request(欲しいデータ)またはprovidable data(提供可能データ・公開されているデータ)を選択し、データ概要情報を入力

本ワークショップ中に各人で、以下のデータについての情報をTEEDAに入力してください。

- ① どのようなデータを手に入れたいか
- ② どのようなデータが提供可能であるか

# ワークショップの手順

## 入手したいデータの入力

- データの名前 \*
- 変数名\*  
(カンマ(半角)「,」で区切って入力して下さい)
- データの利用目的

## ヒント：

- 変数名はデータ連結のコアとなる要素です。
- 可能な限りたくさんのデータを入力してください。
- 可視化サイトも閲覧し、他の参加者がどのようなデータを入力しているのか見てみましょう。

## 提供可能データの入力

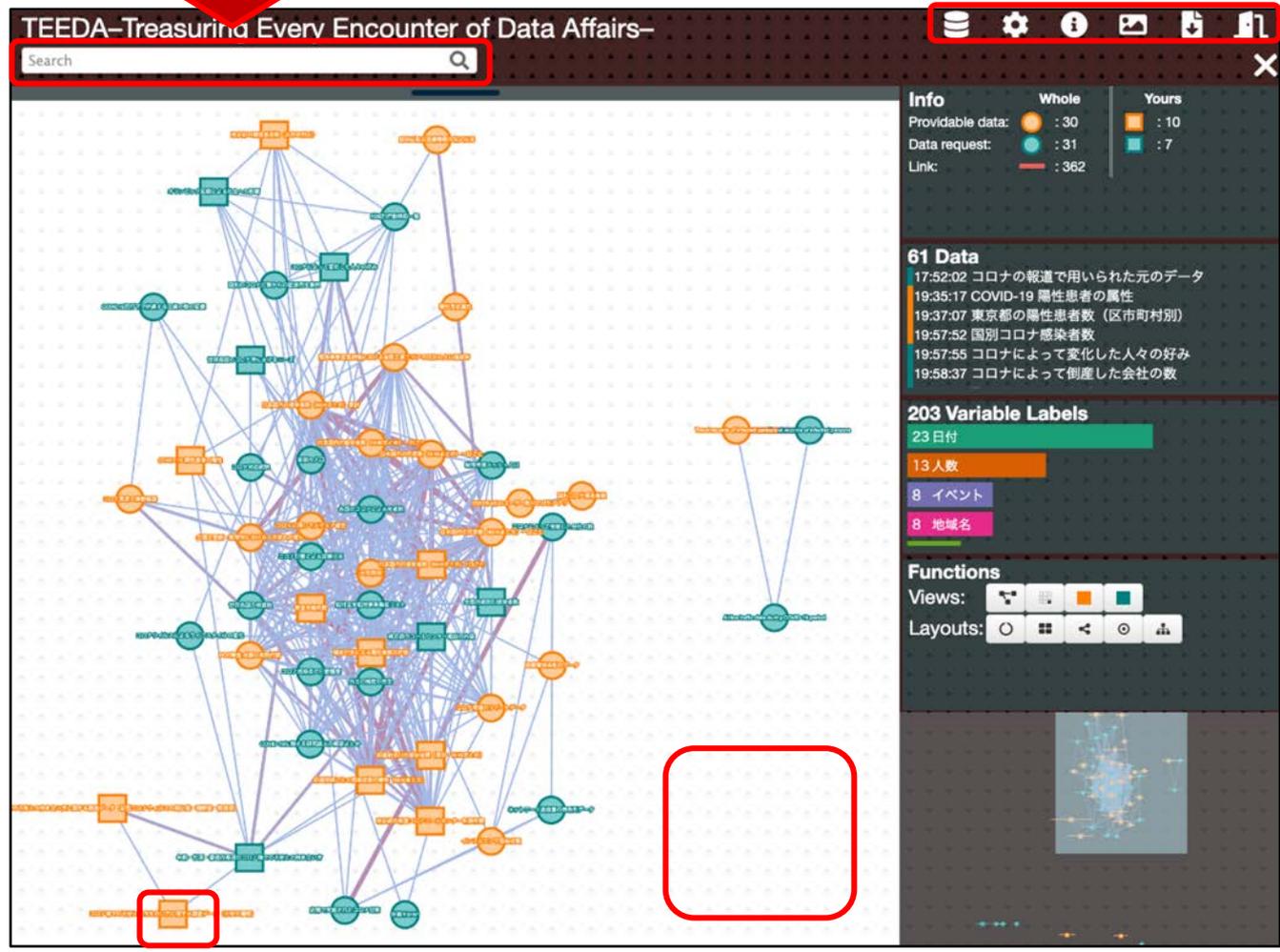
- データの名前 \*
- 変数名\*  
(カンマ(半角)「,」で区切って入力して下さい)
- データの概要
- 共有条件
  - 一般に共有可能、購入により共有可能、共有不可能など
- データの種類
  - 時系列、テキスト、数値、画像など
- データの保存形式
  - RDB、マークアップ形式、txt、RDF、PDFなど

\*: 必須項目

# その他の機能

検索バー

データ登録、スクショ、ダウンロードなど



入力されたデータ概要のサマリー

ノード長押しで詳細情報を閲覧

フィールド長押しでワードクラウド閲覧

# TEEDAワークショップについて

COVID-19現象の理解と対策のため、

- どのようなデータが必要なのか、
  - どのようなデータが提供されているのか、
- という情報を入力してください。

講演や研究発表を聞いて関心の高まったデータについて入力いただいても構いません。

入力いただいたデータ概要情報と可視化については、本ワークショップの総合討論にてご紹介します。

ぜひ、積極的なご入力をお願い致します。

さらなるCOVID-19現象の理解と対策、データ利活用のため、

- ◆ TEEDAワークショップで入力いただいた情報及び当該データについて、ワークショップ後にぜひご相談させてください。
- ◆ 入力いただいた情報は報告及び研究のため、個人を識別可能な情報を除いて活用させていただきます。

# ポストコロナ社会における観測データ設計の集合知化

システム創成学専攻: 早矢仕晃章 助教, 上原直 (M2), 長谷大輔 (M1), 大澤幸生 教授

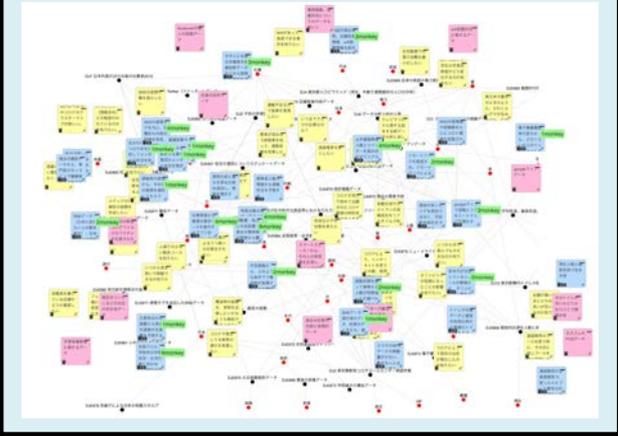
異分野データ連携とコミュニケーション支援によって、目的に応じたデータ設計・収集・共有と意思決定プロセスを支援



COVID-19現象の理解と対策のため、分野を横断したデータ流通・共有の重要性が理解されることとなったが、データの信頼性と質、そして未だデータ化されていない事象を如何に観測するのかという新たな問題が顕在化

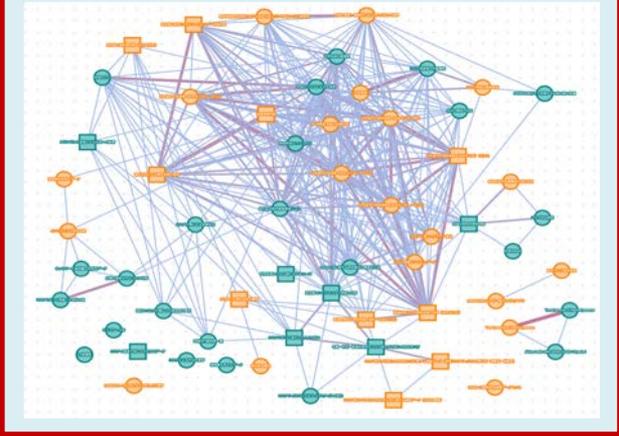
## Web IMDJ (Web-based Innovators Marketplace on Data Jackets)

コロナ禍におけるニーズの表出化とデータの組み合わせによる解決方法をオンラインで議論するプラットフォームを展開。



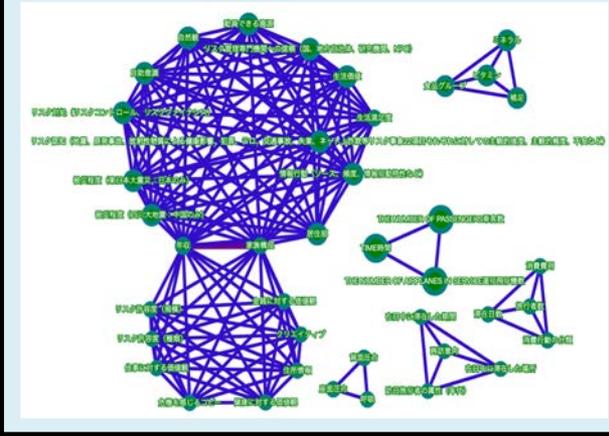
## TEEDA (Treasuring Every Encounter of Data Affairs)

コロナ禍で実際に共有されたデータと足りないデータの情報を集合知化し、未だデータ化されていない事象を検討可能に。



## VARIABLE QUEST

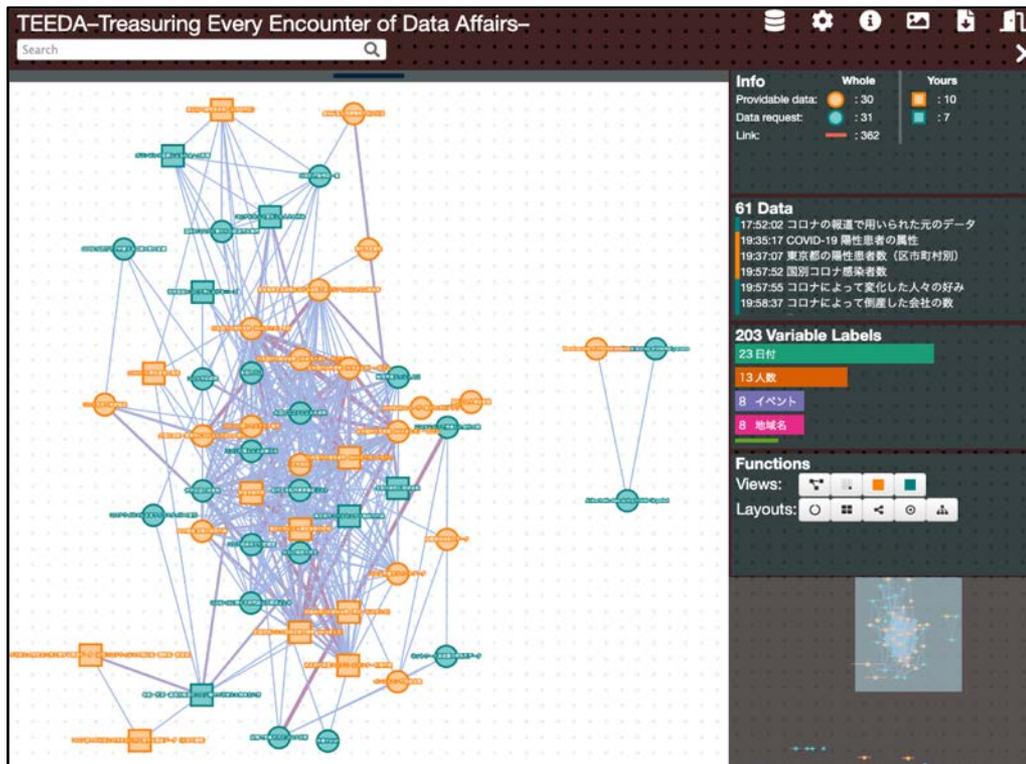
変数が未知のデータ概要から、当該データに含まれ得る変数を推定することで、未観測事象のデータ設計を支援。



# ワークショップで用いるツール

## TEEDA

–Treasuring Every Encounter of Data Affairs–



<http://teeda.data-marketplace.org/>