

# ソーシャルビッグデータ グループ活動報告

# ソーシャルビッグデータグループ

越前功（客員教授，社会データ構造化センター/国立情報学研究所）

相原健郎（客員准教授，社会データ構造化センター/国立情報学研究所）

小須田祐介（特任専門員，社会データ構造化センター）

二瓶美紀（特任専門員，社会データ構造化センター）

斎藤舞（特任専門員，社会データ構造化センター）

## 目的

1. 社会調査データ、公的マイクロデータ、ソーシャルビッグデータのアーカイブ整備
2. 新たな調査データ収集システムと解析システムの開発
3. 実証的データに基づく人文社会科学の発展と政策立案の実現のための研究基盤の発展

## 事業

### 共通課題

- ・データ共有プラットフォームの整備（システム構築、コンプライアンス等）
- ・データ構造化による利用性向上（データクレンジング、リレーション等）
- ・研究利用促進のためのコミュニティ形成（共同利用コンソーシアムの運営等）

1. 社会調査関連事業  
＞全国共同調査ネットワーク形成によるデータ収集、データ整備と公開
2. 公的マイクロデータ事業  
＞公的統計データの整備と共同利用システムの発展、オンラインのデータ解析システムの研究開発
3. ソーシャルビッグデータ事業  
＞リアルタイムの社会変化予測基盤開発、地方のソーシャルビッグデータの専門家養成

## 主な活動

- ① 共同利用データの管理方法ならびにプラットフォームの整備
- ② 実社会データ共有基盤の開発

# 活動概要

## ■ 共同利用データの管理方法ならびにプラットフォームの整備

### • これまでの活動

- Webビッグデータの活用、既存デバイスのセンサライズに関する研究
- 個人情報保護に関する法的な側面からの検討
- データ活用がなされる共同研究活動の推進
- 各種、外部資金にもとづく研究活動でのデータ共有にまつわる課題の検討
- 共同利用データ管理プラットフォームの試行的な開発とその課題に関する検討
- 権利に関する規定等の調査とチェックリスト化のための基礎的開発

### • 本年度の活動

- 共同利用データ管理プラットフォームに関する以下機能およびシステムの開発
  - 共有時の規定チェック機能の開発
  - 運用を想定したシステム仕様の改修ならびに管理者操作システムの開発

# 背景と目的

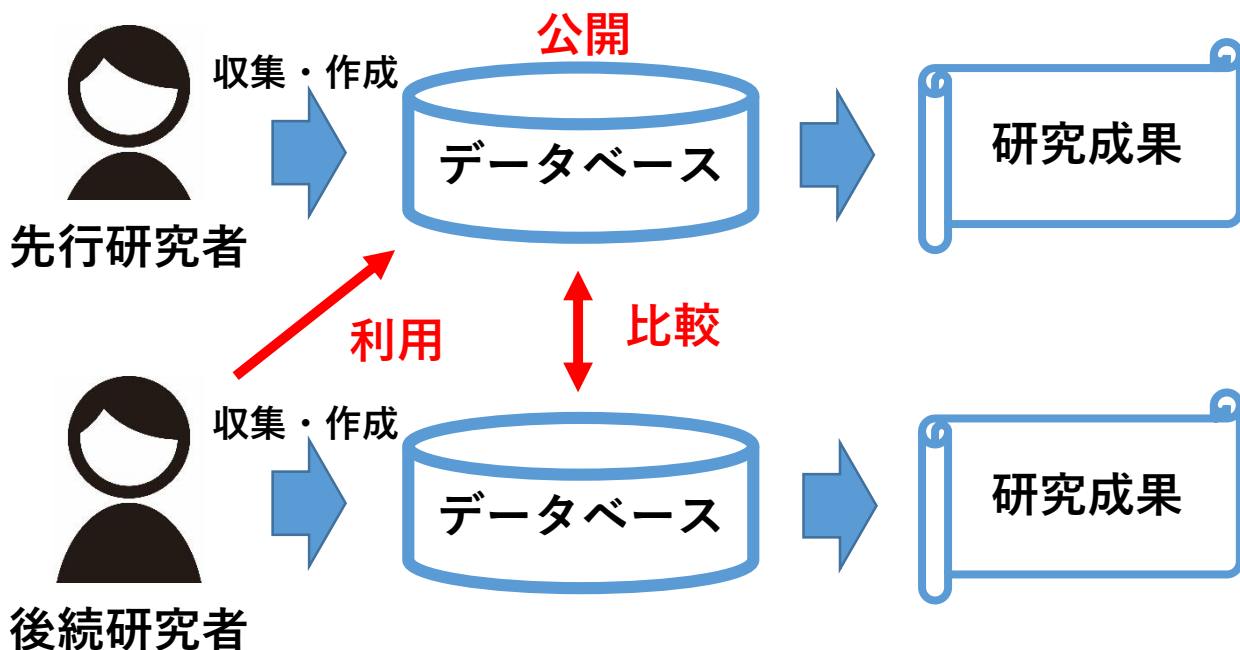
## ■ 共同利用データの管理方法ならびにプラットフォームの整備

### 背景

データベースの利用促進  
利用可能データ活用の簡便化

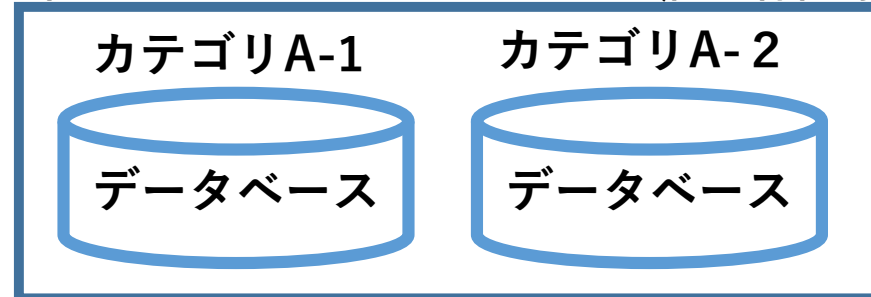
### 目標

共同利用データの管理方法の構築  
プラットフォームの整備

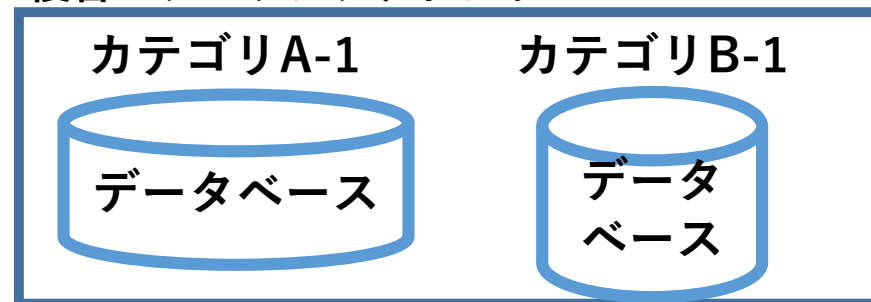


研究促進のためのプラットフォーム整備

### 単一カテゴリプラットフォーム（例：株価等）



### 複合カテゴリプラットフォーム



データ活用簡便化のための管理方法の構築

# これまでの活動①

## ■ 共同利用データの管理方法ならびにプラットフォームの整備

### これまでの活動について

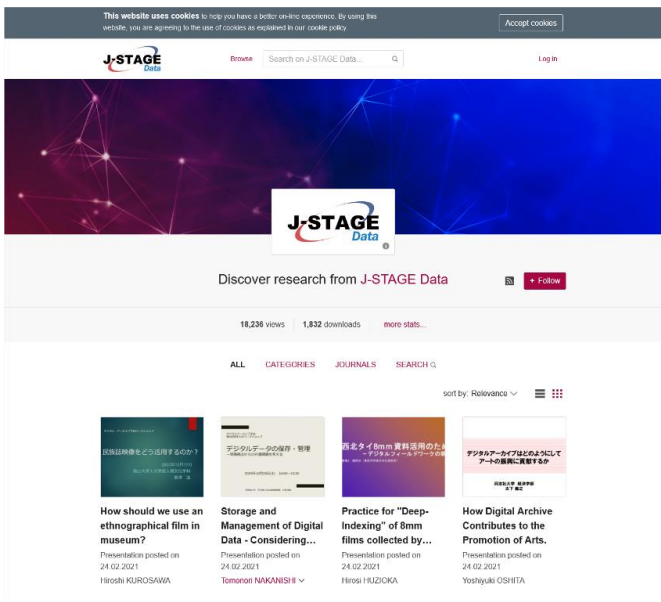
- **データ活用がなされる共同研究**  
→ 共同利用データを共有・使用にあたり考慮する必要がある点の確認
- **外部資金に基づく研究活動のデータ共有課題**
- **個人情報保護に関する法的な側面からの検討**  
→ データ共有に関する詳細課題の確認
- **共同利用データ管理プラットフォームの試行的な開発とその課題に関する検討**

# これまでの活動②

## データ共有の動向

Open ScienceおよびRDM (Research Data Management) におけるデータリポジトリの推進を背景に、それぞれの機関が主体的に研究成果およびデータの管理に対するポリシー策定、責任を負うことが必要とされ、この具体的なスキームが検討されている。

## J-STAGE Data



### 科学技術振興機構 (JST) より データリポジトリの運用が開始

- 記事・論文の根拠となったデータ
- 研究の実施過程で得られた情報およびその関連情報

ただし

(引用) 発行機関に即した具体的なデータの公開・管理ポリシーは発行機関自身で策定する必要がある。

## RDMのロードマップ

### 課題

- 発行機関の管理者ならびに練度の低い研究者 (教員を想定) においては、**リスクと事前の施策が部分的にしか見えていない**



エディンバラ大学「RDMのためのロードマップ」  
*Training* として項目より

局所的な研究リテラシーの向上またはこれに代わるものが  
研究データを取り巻くリスクの低減に寄与する

➡ 昨年度より権利に関する規定等の調査とチェックリスト化のための基礎的開発に着手

[1] 科学技術振興機構 情報基盤事業部 研究成果情報グループ, 第3回SPARC Japanセミナー2020 データリポジトリJ-STAGE Dataでのデータ公開, 2021年2月

[2] 三上 絢子 (北海道大学附属図書館), 研究データ公開実践のための課題を探る: 北海道大学での事例を通じて, 2021年2月

[3] University of Edinburgh, *IS Research Data Service Roadmap V3.2 2017-2020*, 2021年2月

[4] SPARC Europe, *Evaluate your RDM Offering*, <https://sparceurope.org/evaluate-your-rdm-offering/>, 2018

# 本年度の活動①（プラットフォーム構想）

共有する側(投稿者)と利用する側(利用者)  
双方に弊害が存在

投稿者側

投稿してよいか  
確認が研究者個人で必要

↓ 投稿しやすい環境へ

確認作業補助のための  
規定チェックの適用

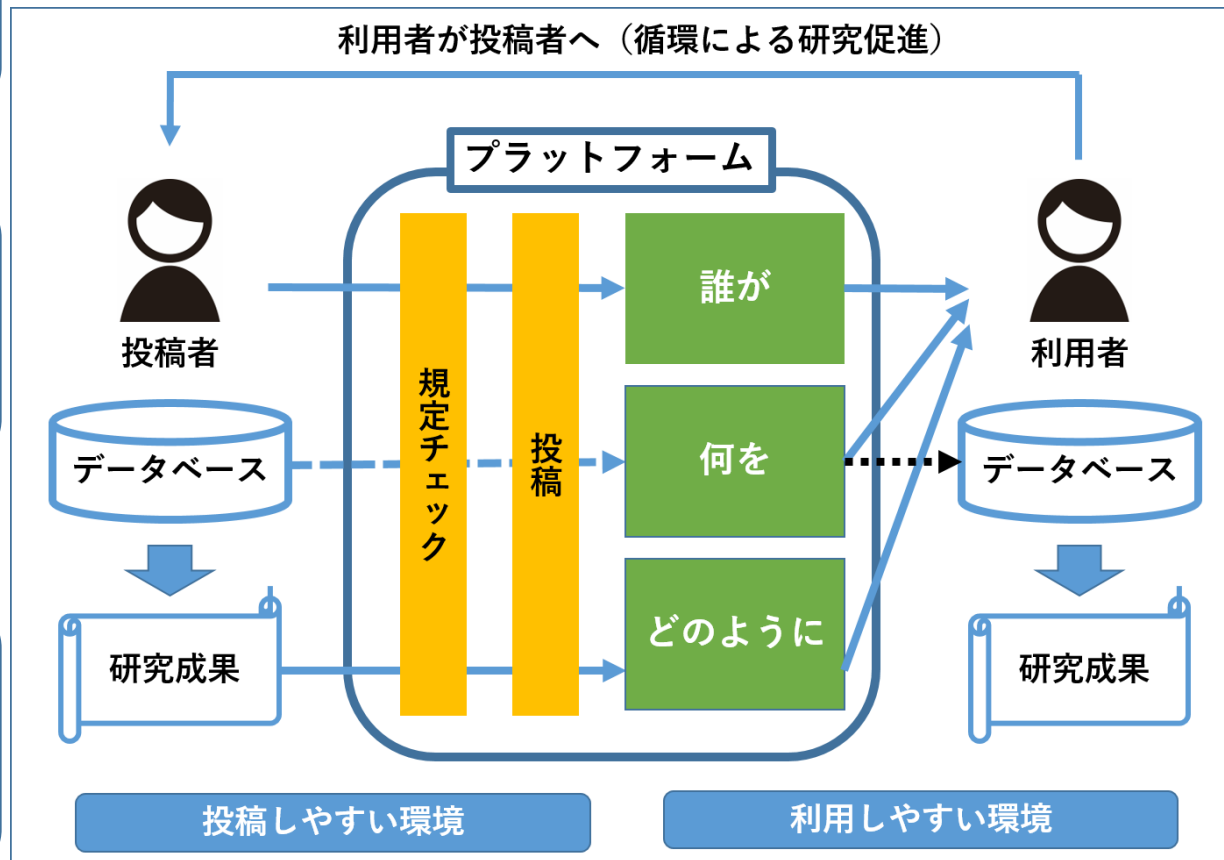
利用者側

データの活用方法等が  
分からない場合あり

↓ 利用しやすい環境へ

誰が何をどのように  
活用したか分かる  
管理方法の適用

利用者が新たな投稿者となる研究サイクルの実現



弊害を緩和するプラットフォームを

システム概要図

# 本年度の活動②（システム改修 投稿時）

csds.jp  クリップしたドキュメント 投稿する 小須田祐介

ドキュメントの投稿 ドキュメントをアップロードする > ドキュメントを確認する > 投稿 アップロード

**概要**

※全角70文字、半角英数200文字まで入力可能です。

※全角500文字、半角英数1500文字まで入力可能です。

**ドキュメントアップロード**

アップロードするPDFファイルをドラッグ&ドロップしてください

ファイルを選択

※PDF形式、3GBまでのファイルをアップロードできます。

**添付ドキュメント**

アップロードするファイルをドラッグ&ドロップしてください

ファイルを選択

※5つのファイルを計5GBまでアップロードできます。

**共著者情報**

※カンマ区切り著者名またはメールアドレスで入力可能です。メールアドレスは公開されません。

**公開設定**

全体に公開

**分野**

情報学基礎理論

キャンセル 次へ

## 投稿時

### 入力内容

- タイトル
- 概要
- ドキュメント
- 添付ファイル（データベース等）
- **共著者情報**
  - 入力された共著者に通知（登録ユーザのみ）
- **公開設定**
  - 全体公開
  - ログイン公開
  - 非公開
- **分野**



# 本年度の活動③（機能開発 投稿チェック）

ドキュメントの投稿 ドキュメントをアップロードする > ドキュメントを確認する > 投稿

このページで入力された内容が公開されることはありません。

### 研究活動の審査状況について

基本審査の状況をご回答ください。

利益相反の確認

研究論理の確認

その他の審査

### ドキュメント内容の分析

ドキュメント内容の分析結果は以下のものになります。

確信度	ドキュメント中の表現	説明
0.84	...アンケート...	プライバシー情報を含む可能性があります。 at 1.

予測されるチェックリスト（研究分野により異なります）

投稿の際は、すべてのチェックリストに対してご回答ください。

予測されるチェックリスト  全てのチェックリスト

プライバシーは保護されていますか。  
 はい  いいえ

あなたがアップロードしようとしている資料に以下の関連するキーワードがありましたので、ご確認ください。

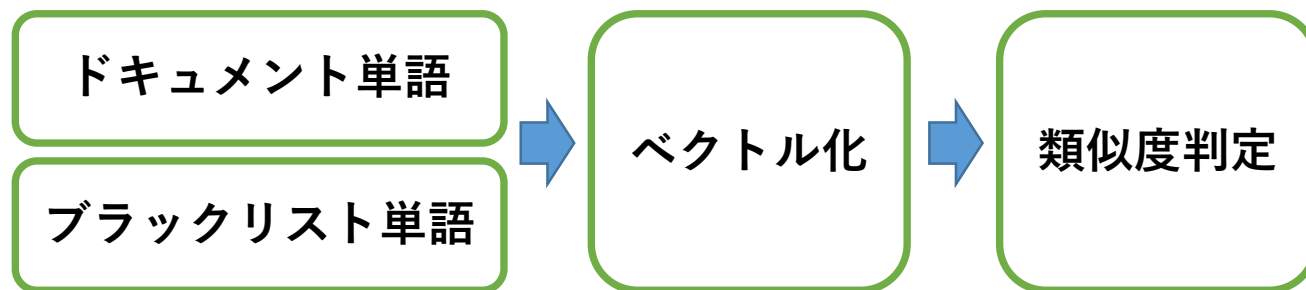
	確認必要確率
...アンケート...	0.84
...調査にて データを取...	0.7

## 投稿チェック

- 基本審査の確認
- ドキュメント内容の分析結果
- 予測されるチェックリスト
  - ドキュメント内容の分析結果の反映
  - 注意喚起の要素
- 全てのチェックリスト
  - 現状の予測ではリスク低減不足
  - 全てのチェックが必要

## 投稿チェックシステム

- ワードベースでの分析



# 本年度の活動③（システム開発 管理者操作）

## 管理

ユーザー
ドキュメント
ヘルプ
プライバシーポリシー
利用規約
リリース
ドキュメント審査関連
サーバー設定

### ユーザ

- ・追加、無効化、リセット

### ドキュメント

- ・公開、非公開設定、削除

### ヘルプ、プライバシーポリシー、利用規約、リリース

- ・各ページの文言設定

### ドキュメント審査関連

- ・規定チェックページの設定

### サーバー設定

- ・ユーザ設定トークンの送り元メールアドレス等

### ユーザーの招待

ユーザーの招待を行います。

メールアドレス（必須）

氏名

所属

招待

### ← ユーザ招待

### ↓ ユーザ無効化、リセット

3 小須田祐介 社会データ構造化センター test test

無効化 リセット

### ↓ ドキュメント非公開、削除

1 Admin ソーシャル・ビッグデータを用いたインバウンド対策としての小規模宿泊施設における経営支援

2021-12-24 未

非公開

削除

# 本年度の活動④（システム改修 その他）

## その他

- サービス全体のSSL対応（HTTPSプロトコルの強制）
- 各種UI,UXの変更（情報修正等各種機能アップデートに伴う修正）

## 閲覧

どのように

The screenshot shows a document viewer interface. The main content area is a blue rectangle with the text "資料や論文" (Materials and Papers) in white. Below this is a dark blue footer with navigation icons and the text "1 of 31". To the right of the main area is a sidebar with a search bar and results. The search bar contains the text "何を" (What). Below it, a red box highlights the result "データ.csv". Below that, another red box highlights the result "誰が" (Who) with a user icon and the name "小須田祐介". Below the sidebar, there are two lines of text, "test", each with a small blue icon to its left. At the bottom left of the screenshot, there is a date "2022-01-21" and a small blue icon with the text "限定公開" (Limited Public).

- 同一プラットフォーム上で、**誰が、何を、どのように**活用し、研究したかを確認することが可能

# 現在進行中と今後の課題

## ■ 共同利用データの管理方法ならびにプラットフォームの整備

- 進行中
  - テストによるフィードバックから仕様修正
    - 規定チェックの分析部分の調整
    - 規定チェック部分に該当なしという選択肢を作る等
- 今後の課題
  - 規定チェックシステムの改良
    - ワードベースの分析にとどまっている
    - 例えば、調査というワードが入っているが、プライバシーに関係のないことだとしても、分析結果に出力されてしまう。
    - リスク低減ということで、様々なものが反応することは良いことかもしれないが、チェック項目が複数ある以上情報が煩雑になる可能性が高い→要修正
    - ワードベースから文章ベースに変更等を考えている

# ソーシャルビックデータグループ関連事業：主な活動②

## ■ 実社会データ共有基盤の開発

- 街の流動を反映したデータの共有基盤
  - 社会活動状況を反映するデータを動体から取得し、解析、共有するための基盤開発

## ■ データ収集機能

- 交通系
  - Bluetooth ビーコンを搭載した移動体（バス、タクシー、レンタサイクル等）を、スマートフォンアプリにて検知し、位置情報とともにスマートフォンの回線にて送信するクラウドソーシングの方法論
- 道路状態・交通状況データ
  - 既存のドライブレコーダーアプリに路面性状検知の機能を追加

## ■ データ互換性

- GTFSリアルタイムや各種道路関連データとの相互互換を踏まえ、デファクトを含めた標準形式に対応したデータの入出力

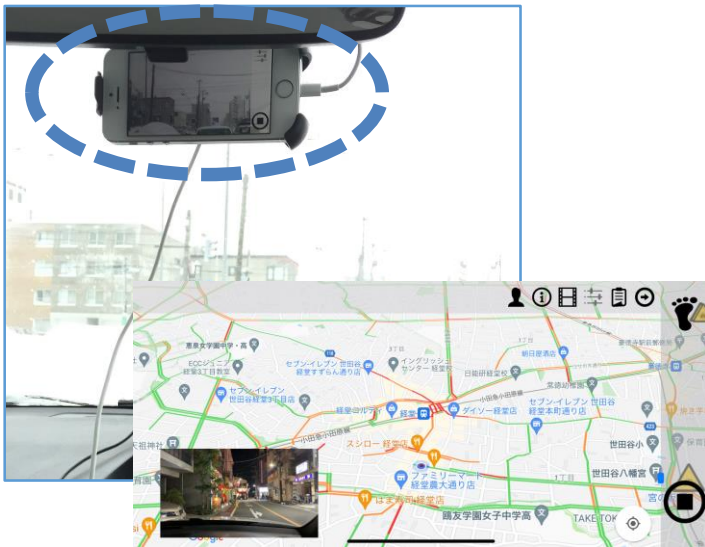
# 2021年度：実社会データ共有基盤の開発

- 実社会データ共有基盤の収集データの利活用実証を通じた整備
- 路面状態・交通状況データ活用の社会実装を通じた実証（島根県）
  - 業務への実装を前提に、実証を進める。
  - 実車を用いたデータ収集の検証と、収集データの蓄積から活用まで

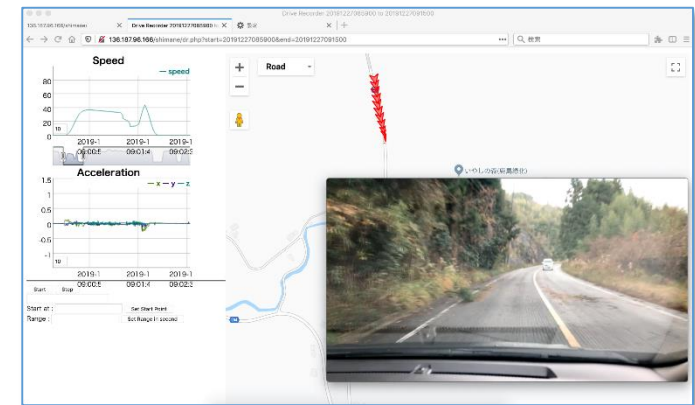
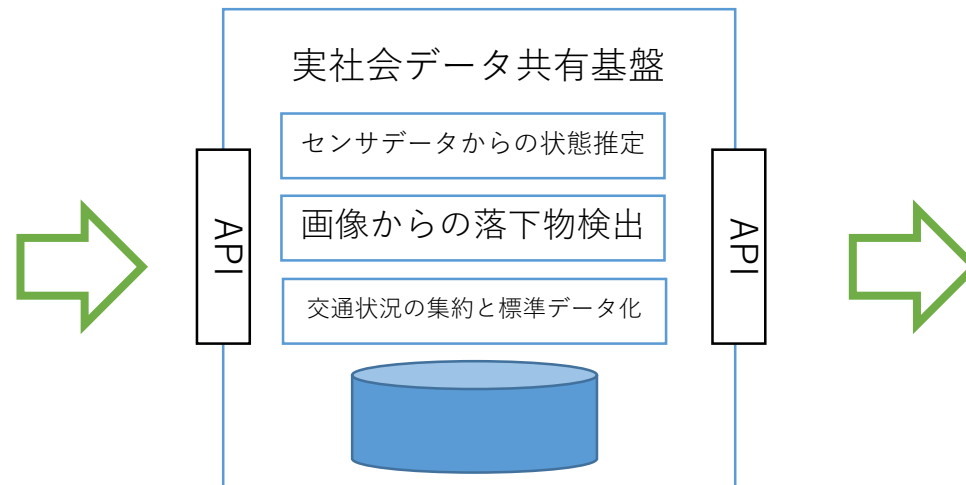
# 実証：道路状況把握のための路面データ収集と活用（島根県）

## ■ データ収集から基盤への蓄積までの実用性検証

- 道路管理業務における道路パトロールにおいて、スマートフォンなどの普及デバイスを活用した半自動でのデータ収集と、データからのイベント抽出を実施
  - 落下物や損傷の把握
  - 路面性状の推計
- 実業務への適用を前提に、収集データからの路面性状推定の実用性の検証を実施



車載スマートフォンアプリ



データ活用例  
(e.g. 路面状態の変化の把握)

# 道路状況把握のための路面データ収集と活用（島根県）

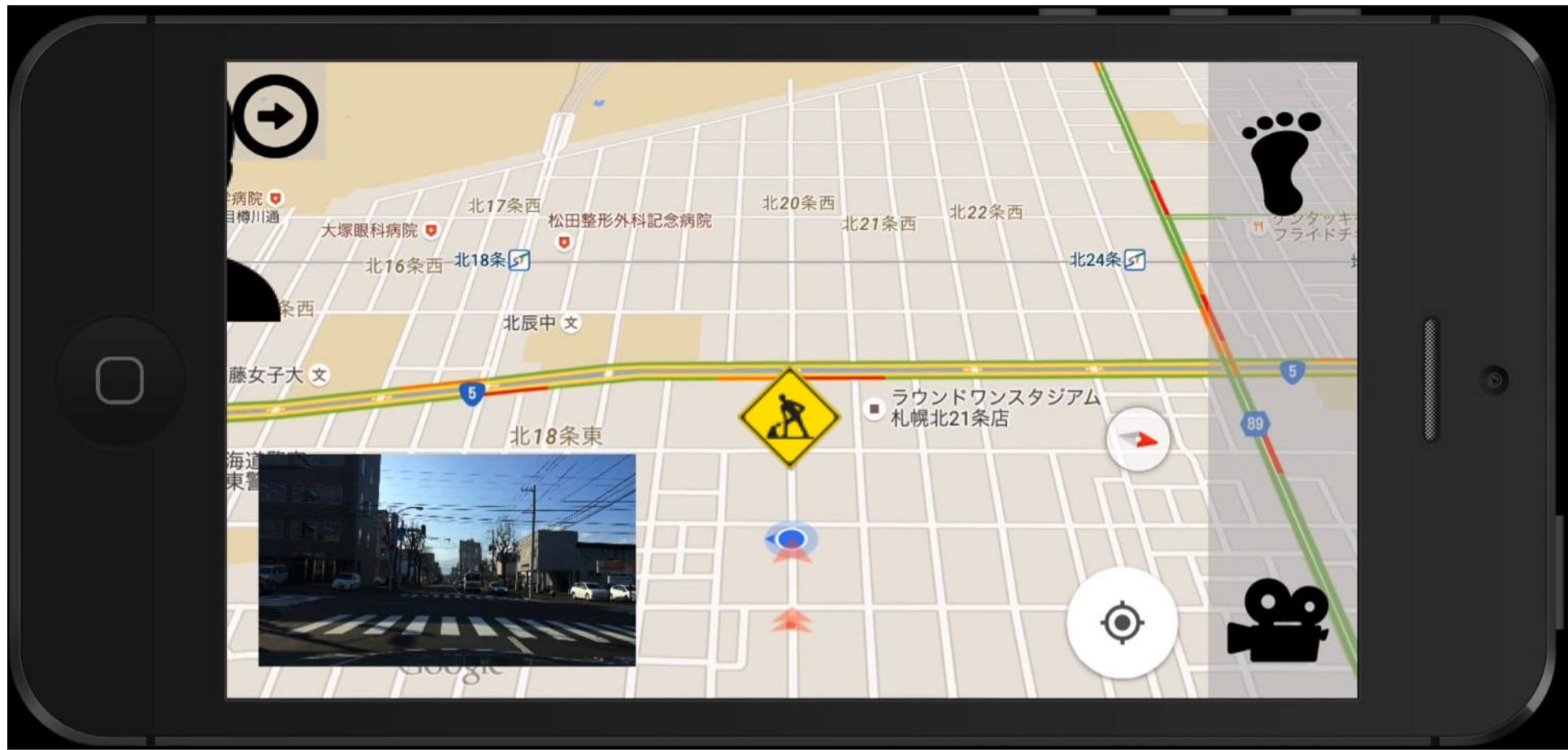
## ■ データ収集と活用に関する実用性検証（FY2021）

- 道路管理業務における道路パトロールにおいて、スマートフォンなどの普及デバイスを活用した半自動でのデータ収集と、データからのイベント抽出を実施。
  - 路面性状の推計
  - 実車・実作業における利用
- 業務への導入を念頭に、データ収集からの蓄積・活用までを実施。



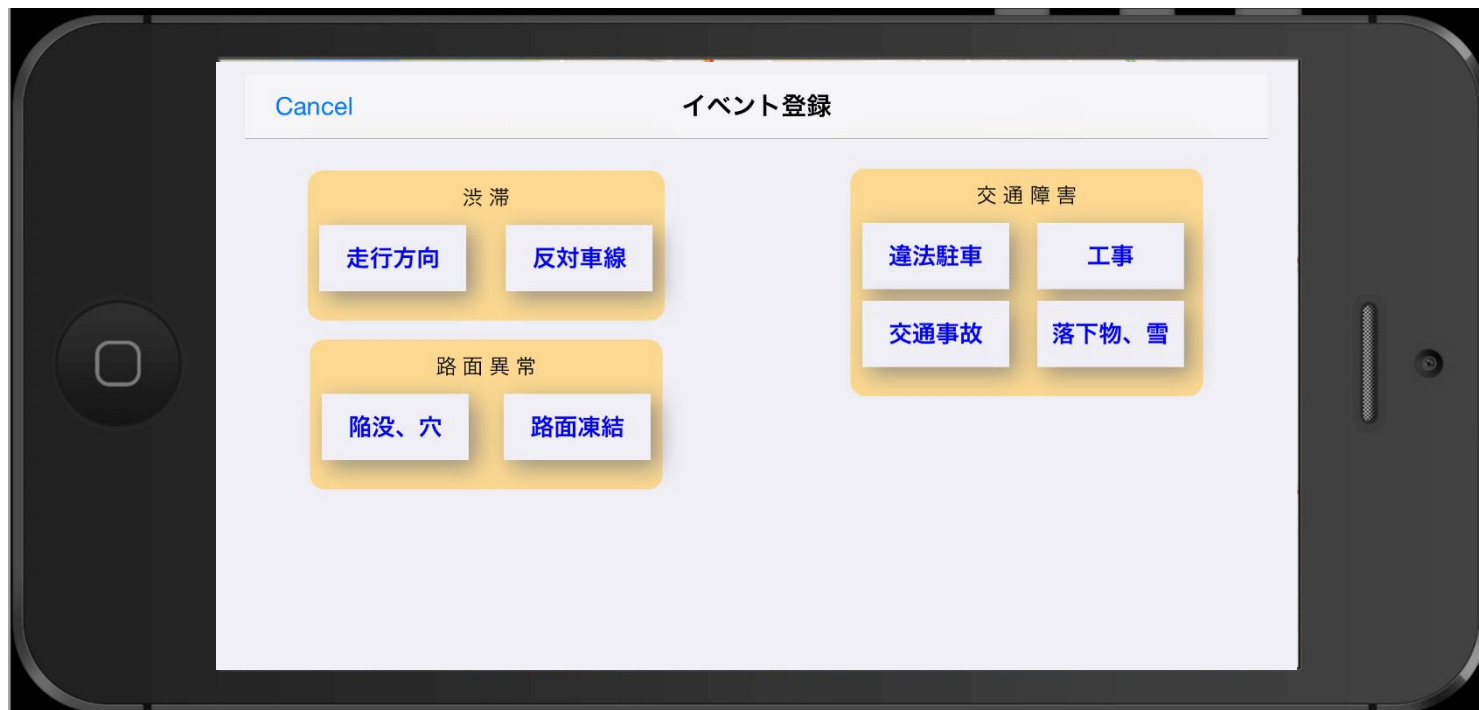
# データ収集に利用するツール

- 国立情報学研究所 相原研究室で開発されてきたDrive around-the-corner. アプリサービス (Drive ATC) に、一部機能拡張をしたものを利用する。
- Drive ATC は、運転中に自動車内に固定されたスマートフォンにて使用される。
- Drive ATC アプリは、ドライブレコーダー機能を有しており、運転中の移動、挙動と、走行映像を自動で記録することができる。
- 取得データ
  - 位置情報 (緯度、経度、測位精度、移動方向、速さ)
  - 加速度情報 (xyz 3軸、回転3軸)
  - 映像
  - 利用者が入力したイベント情報
  - 映像以外のデータは、携帯回線を通じてリアルタイムにサーバ上に送信され、蓄積される。映像は、規定時間の長さの映像ファイルとしてスマートフォン本体内に保存されるが、携帯回線を通じて自動送信することも可能である。



Drive  
around  
the  
corner.

# イベント投稿機能



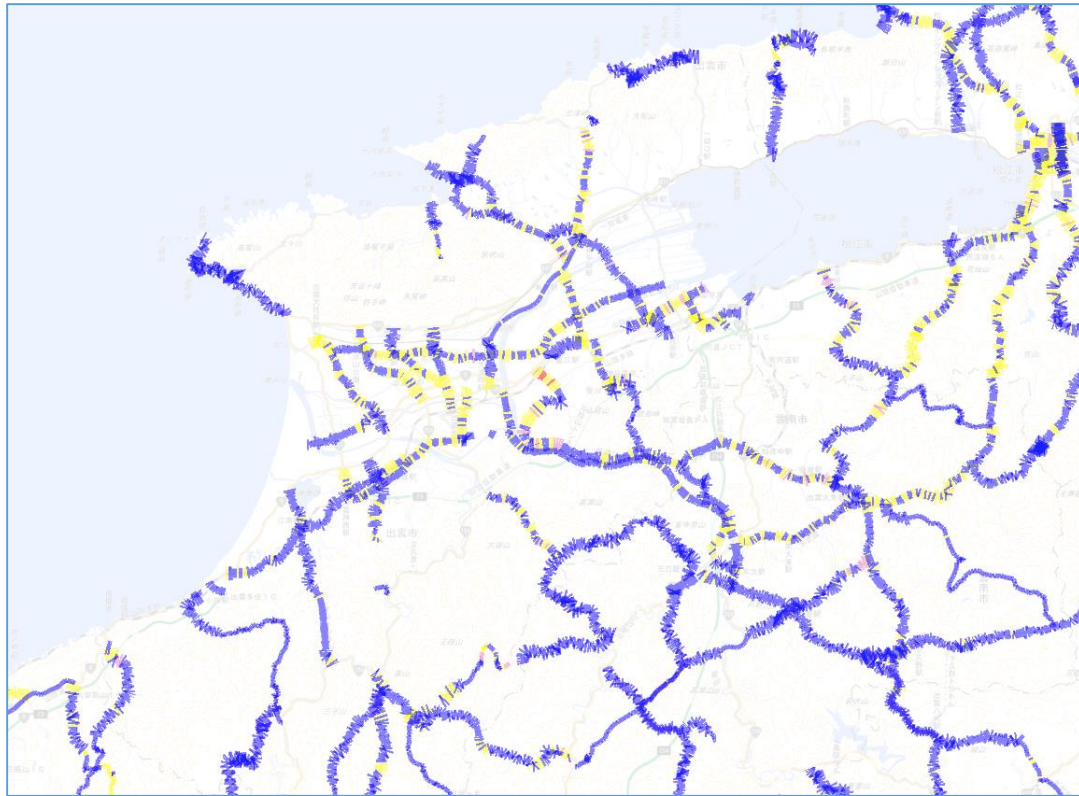
- 信号待ちや渋滞などでの停止中などに、簡単な操作（3タップ）で「今何が起きているのか」を投稿することができる。
- 投稿されたイベントは、利用者間で共有される。

# データ収集の実施

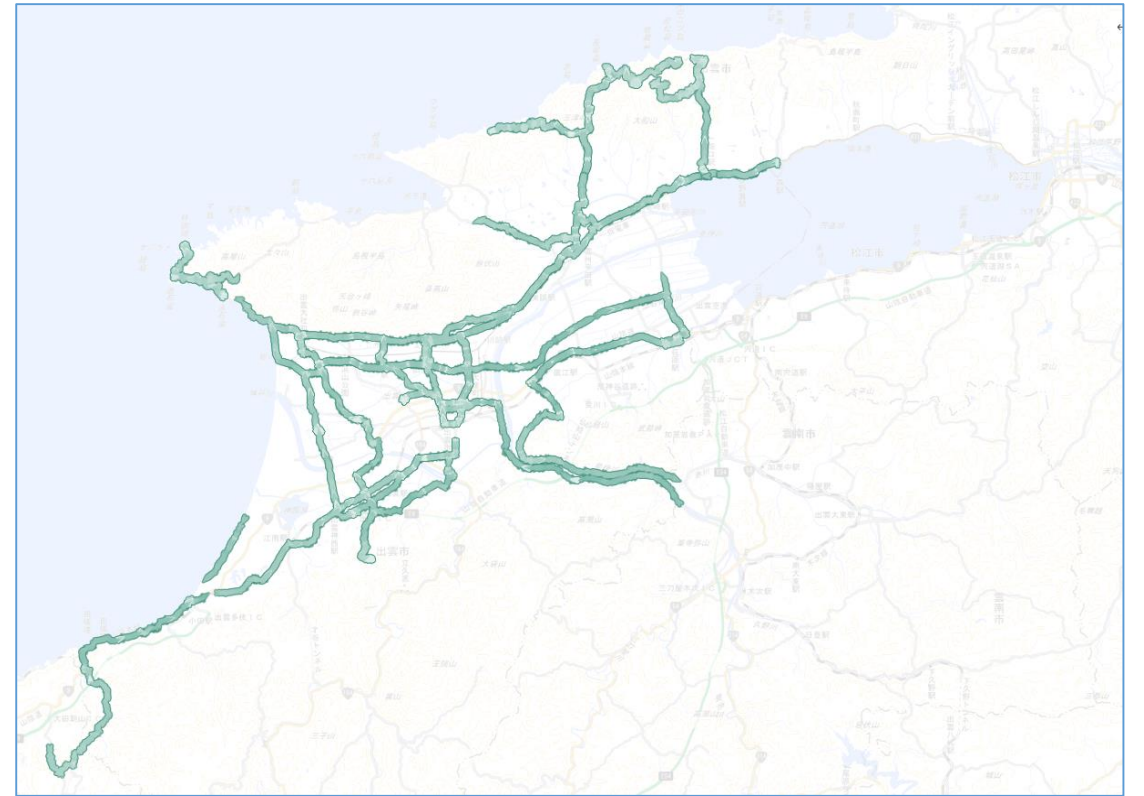
- 2022-01-24～2022-02-15
- 自動車にスマートフォンを固定し、Drive ATC 改良版（IRI 推計・収集機能を実装）によるデータ取得を実施
- 取得データ
  - Drive ATC（～2022-02-13）
    - 映像データ（間欠型）：212.8 GB
    - センサーデータ（移動＋加速度）：23,056 千サンプル（約 14 GB）

# データ収集路線

2020（令和2）年度調査での対象としている路線と比較した場合、2台の車両のみでも、地域内の路線を密度高く取得していることが確認できる。

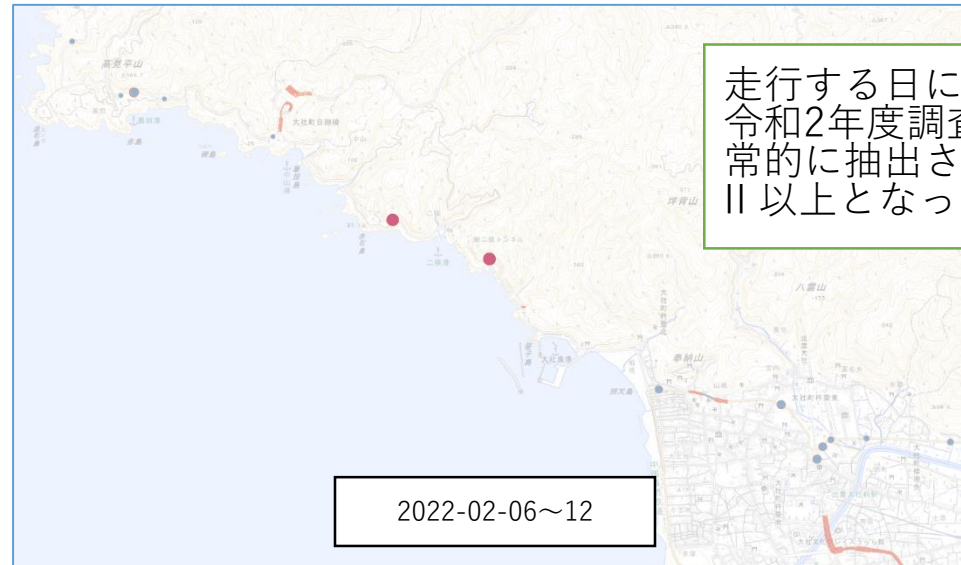
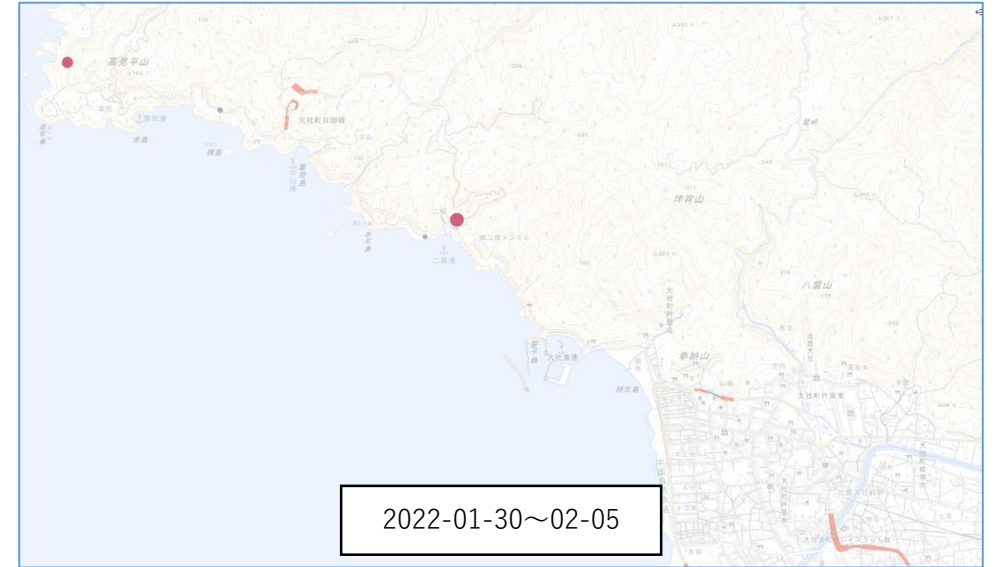
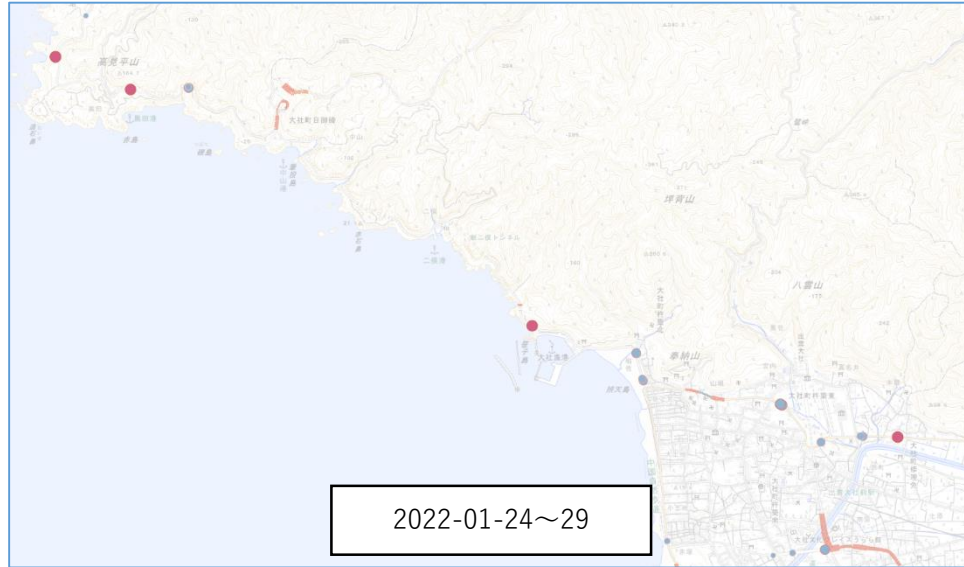


参考：令和2年度 調査対象路線



2022年 実走実績

# 取得データ例 (1)



走行する日によっても若干の違いが見取れる。令和2年度調査では検出されていない箇所が定期的に抽出されていることから、路面が健全度II以上となっている可能性が示唆される。

一定以上の IRI 計測値が観測された区間を ○ で示す。  
○が大きいほど IRI 値が大きい。2 未満は青系、2 以上が赤色となっている。

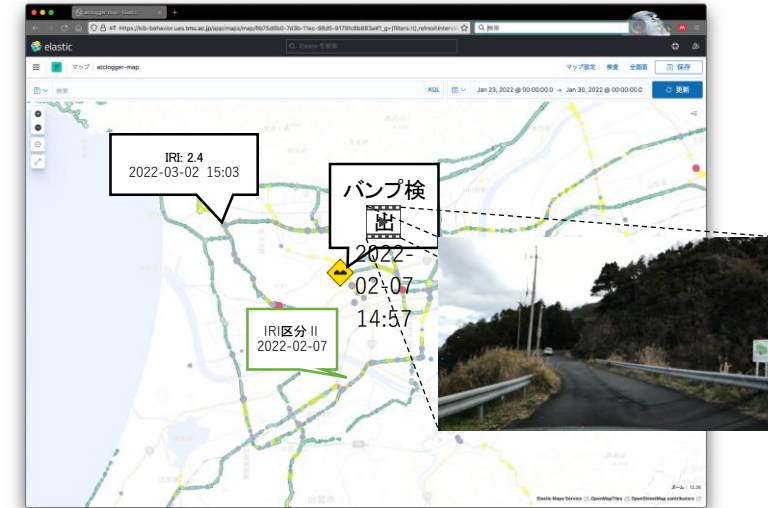
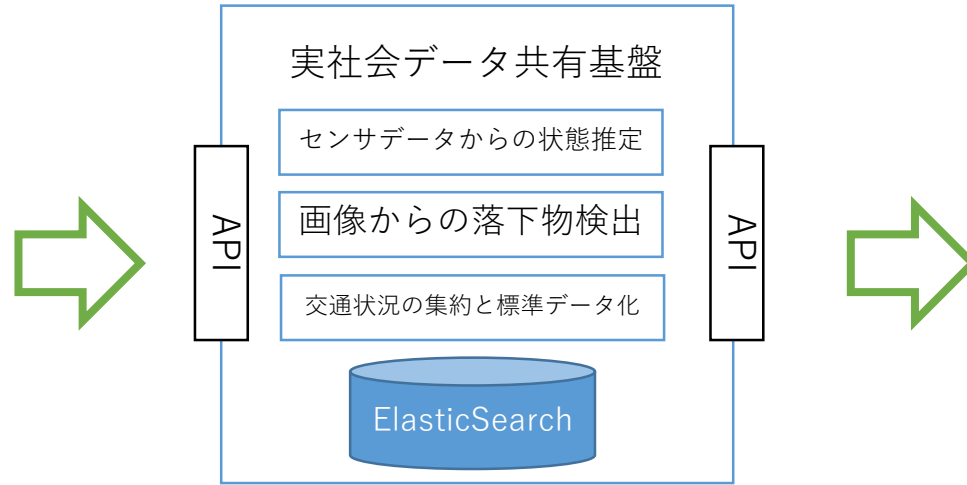
令和2年度調査では、片側車線を1回計測した結果となっているが、アプリでの計測は両方向を繰り返し計測できるところに強みがある。下記例では、令和2年度調査では東行き方向で健全度Ⅱ判定となっている箇所について、アプリでは、同方向（東行き）ではIRI値が1.5未満となっている一方、逆方向（西行き）で相対的に大きな値が検出される結果となった。アプリでの計測が時間的、空間的により高い分解能で行えることが分かる。



# 実社会データ共有基盤



車載スマートフォンアプリ



状況把握・分析用ダッシュボード例  
(e.g. 路面状態の変化の把握)

公開のための集計・プライバシー保護等の処理  
(主に画像について)



FY2022

- 試験運用への実装を検討
- データ数、利用車数、期間の拡大
- プラットフォームとの連携と公開のための処理