

# 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）

## 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（小須田・二瓶）
- ② 実社会データ共有基盤の開発（相原）

# ソーシャルビッグデータグループ関連事業：主な活動①

小須田・二瓶

## ■ 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

- ソーシャル・ビッグデータグループでは社会データ構造化センターの発足当時からこれまでにかけて次のような活動が行われてきた。
  - 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
  - データリポジトリの試行的な開発とその課題に関する検討
- 本報告では、2020年度の活動報告として以下の内容について、その概要を報告する。
  - 各機関、学会、団体等における研究倫理、利益相反、個人情報保護、その他の各種権利に関する規定類の網羅的な調査及び構造化のためのスキーム作成
  - 研究データ管理、データ共有ポリシー策定のためのパブリッシュ段階における研究活動の推定およびチェックリスト化のための基礎的開発

# ソーシャルビックデータグループ関連事業：主な活動①

## 研究活動のデータ共有にまつわる動向（参考になったもの）

- エディンバラ大学においては、RDMのためのロードマップを公開、模索中。
- “Evaluate your RDM Offering”のような管理者 層向けの自己査定ツール

## 課題

- 練度の高い研究者は、研究活動を実施するにあたり、その注意事項、リスクを十分に把握している。
- 一方で、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**。
- また、それらは研究領域に依存するところもあり、管理者は専門的な人的資源に頼らざるを得ず、管理者としてのそうした人材の獲得は難しい。
- ここに生じる責任は、研究者にだけ課せられるものではない。
- 機関としてオープンサイエンスの推進を掲げたとしても、二の足が踏まれる。（ハードル）



- RDM（研究データ管理）が重要となる。
- エディンバラ大学「RDMのためのロードマップ」の *Training* として項目化されているように、局所的な研究リテラシーの向上（またはこれに代わる何か）が、研究データを取り巻くリスクの低減に寄与するものとする。

# ソーシャルビックデータグループ関連事業：主な活動①

## 課題に対する提案

- それぞれの研究領域でどのような規範が形成されているか？
  - 各機関、学会、団体等における研究倫理、利益相反、個人情報保護、その他の各種権利に関する規定類の網羅的な調査及び構造化のためのスキーム作成
- それぞれの研究活動が規範・ポリシーを逸脱するリスクを抱えていないか？
  - 研究データ管理、データ共有ポリシー策定のためのパブリッシュ段階における研究活動の推定およびチェックリスト化のための基礎的開発

# ソーシャルビックデータグループ関連事業：主な活動①

## ■ 網羅的な利益相反、個人情報保護、研究倫理に関する記述の調査

- **モチベーション**

- それぞれの機関で少なくとも何が求められているのか、遵守すべき内容の可視化を行いたい。
- RDMを含むポリシーのマトリックス

- **Webクローラーによる機械的な調査**

- (WIP)800+機関（大学法人、独立行政法人を含む）ならびに120+学会の公式ページを対象として、自動的な収集を行っている。
- PDFやWordファイルについても対象としている。
- ページごとにタグを付与し、利益相反、個人情報保護、研究倫理、その他に関する記述のあるページを列挙可能に。
- ホームページの運用に必要なプライバシーポリシーなどを区別して排除。
- (WIP)チェックリストなど有益な情報に関する調査を実施している。

- **記述内容の人手による精査**

- (WIP)規範記述のためのメタデータの検討を行っている。
- 研究者、管理者がアクセスしやすいようにコンテキストに沿ったインデックス、図書館で行われる「分類」のような作業が必要となるため、2021年度に実施。

# ソーシャルビックデータグループ関連事業：主な活動①

## ■ 網羅的な利益相反、個人情報保護、研究倫理に関する記述の調査

年度末に向けてと次年度

### • 進行中

- それぞれの記載の根底にある考え方は共通のはず。
- ただし研究領域ごとの特色がある。より領域によったもの。
- ポリシーに関する「ある、なし、表現パターン」のリスト化について年度内を目標に作成中。

### • 次年度

- 規範、ポリシーとして何が求められているのか、そうした要求に研究者また管理者として何が実行可能か？を検討したい。

# ソーシャルビックデータグループ関連事業：主な活動①

## ■研究活動の推定およびチェックリスト化のための基礎的開発

### ・モチベーション

- ・ 研究計画や研究論文などのテキストから、データ共有についてどのようなチェックが必要になるかを可視化し、網羅的なチェックの実施ではなく、管理者による必要にして十分なチェックの実施を実現したい。
- ・ 論文チェックツール<sup>[5]</sup>のように簡便な形式で、公表やリポジトリ登録にあたり確認すべき事項を通知したい。

### ・ 確認を設けたい表現の一例（日本語）

- ・ 「アンケートを取り、…」 研究プロセスに関する言及があるケース
- ・ 「～によるデータを利用して、…」 第三者のデータを活用するプロセスに関する記述があるケース
- ・ 「深層学習をもちいた～における…の獲得」 研究プロセスにおいてデータの利用が頻繁に行われる研究テーマおよびその分類であるケース

### ・ チェックを行いたい内容の一例

- ・ 十分なインフォームドコンセントはなされたか、利用するデータセットは第三者利用が可能なデータであるか
- ・ オプトインがなされ、オプトアウトが可能な状態か、エビデンス資料は存在するか
- ・ 被験者が不利益を被る研究ではないか、データ提供元との利害関係はないか



# ソーシャルビックデータグループ関連事業：主な活動①

## ■研究活動の推定およびチェックリスト化のための基礎的開発

年度末に向けてと次年度

- 進行中

- モデル的な改修と多様なチェック項目に対して検知が可能なように訓練。
- チェック用の簡易的なWeb UIの構築と、教師データをアップロードする機能の追加。
- テスト用のデータセットを作成して、精度評価の実施。

- 次年度

- 研究者と管理者がポリシーの遵守状況や活動プロセス上のリスクについて、対話を行いやすいインターフェースを検討したい。

# ソーシャルビッグデータグループ関連事業：主な活動②

相原

## ■ 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を行う
- 事例と実証：地方自治体等との活動
  - 流動データ
    - クラウドソーシングによる公共交通やシェアリング等の活動状況の把握（札幌市）
    - コロナ禍で、市民対象のデータ収集の実証が難しい状況
  - 道路状態・交通状況データ
    - 道路状態および交通状況の把握のためのデータ収集と活用
    - 島根県からの受託研究と合わせて実施

# 実社会データ共有基盤のシステム整備

## ■交通系流動データ

- データ収集
  - Bluetooth ビーコンの ID を、一般利用者向けに配布する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にて、バスや市中に展開する移動体の位置情報を収集する。
  - デファクト標準の形式（GTFSリアルタイム）に準拠したデータの入出力
  - 課題：
    - アクセス制御
- 対象： バス、シェアサイクル

## ■道路状態・交通状況データ

- データ収集
  - 既存のドライブレコーダーアプリへの機能追加・カスタマイズ

# 道路状況把握のための路面データ収集と活用（島根県）

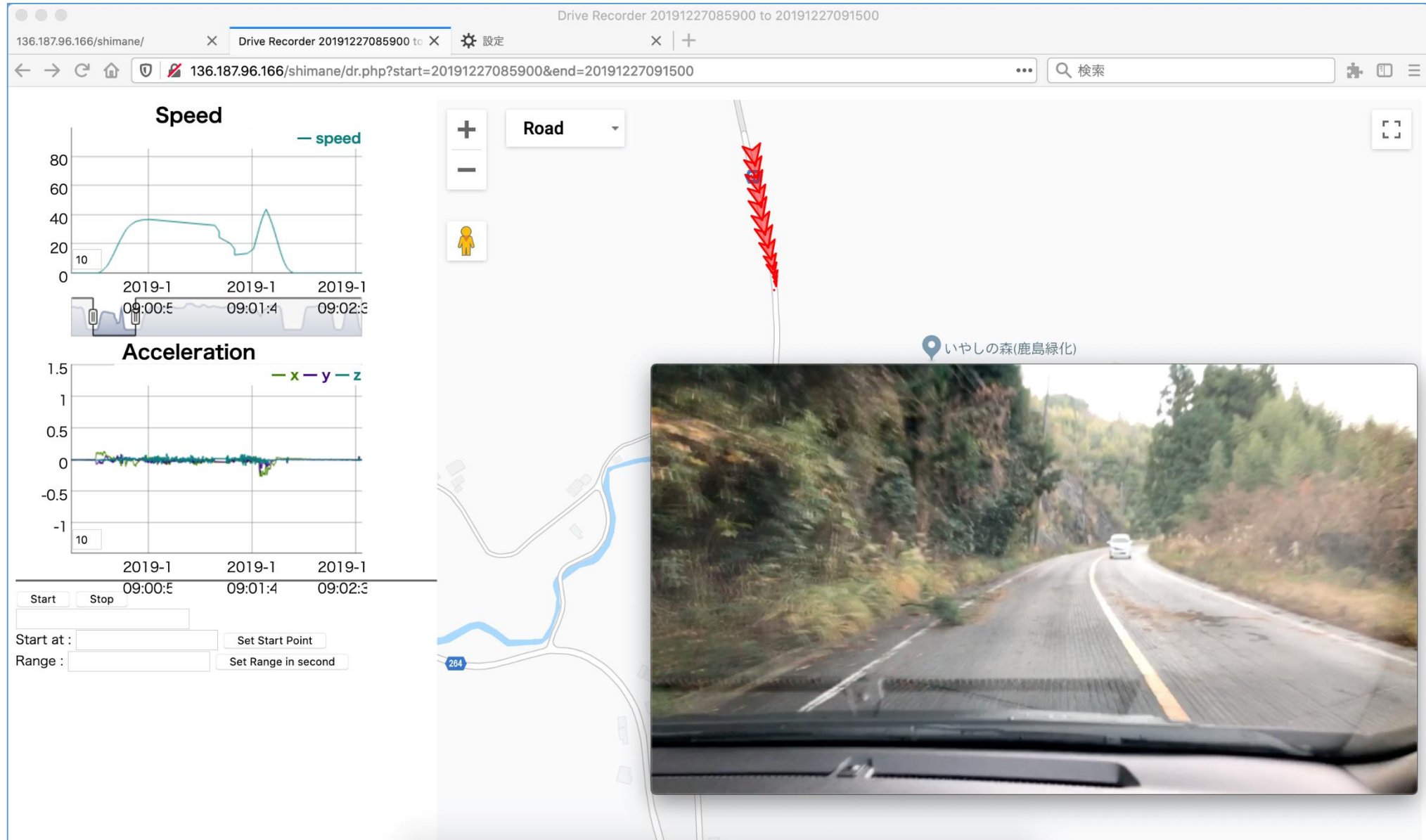
## ■ 収集データの実用性検証（FY2020）

- 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
- FY2021 以降の業務への適用を念頭に、収集データからの路面性状推定の実用性の検証を実施。
  - FY2020 の路面性状調査事業での結果を用い、収集データからの推定精度等から、簡易的な方法でどこまで従来方法を補間、置き換えできるかを検証する。

# 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）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。

# 収集データの例



# 2021年度計画

相原

## ■ 実社会データ共有基盤の開発（相原）

- 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
- 流動データの取り組み（札幌市）
  - 状況を見て先方と相談の上進める。
- 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）
  - FY2019に実施した映像からの落下物抽出等と合わせ、業務への実装を前提に、実証を進める。

# 学会発表等（論文 1 件，国際会議 1 件，国内会議 0 件） 2020年1月～

- 論文

- Kenro Aihara, Atsuhiro Takasu: Development of One-Stop Smart City Application by Interdisciplinary Data Linkage, Distributed, Ambient and Pervasive Interactions, LNCS, Vol. 12203, pp.379-390 (2020).

- 国際会議

- Kenro Aihara, Atsuhiro Takasu, Integrating inter-field data into space-time to grasp and analyze activities in town, Prodeedings of the International Conference on Human-Computer Interaction (HCI2021) to appear (2021).

- 国内会議

- その他

- Kenro Aihara, Hajime Imura, Crowdsourcing for Smart Cities that Realizes the Situation of Cities and Information Sharing, Handbook of Smart Cities, Springer, to appear (2021).