

# 7 つながる実世界データの利活用

## —課題解決指向のオープンデータ 統合基盤—



是津 耕司 (情報通信研究機構)

### オープンデータの潮流

今日、さまざまな組織や機関、個人が収集したデータを公共財として自由に利活用できるようにするオープンデータの取り組みが世界的に広まっている。代表的な例として、Data.gov<sup>☆1</sup>では、米国政府機関が保有する科学技術、交通、小売、通信、雇用、金融、貿易、エネルギーなど多岐にわたる数十万件の公共データを公開しており、同様のデータ公開が世界40カ国以上に広がっている。また、国際科学会議(ICSU)の世界科学データシステム(World Data System, WDS)<sup>☆2</sup>は、前身である世界データセンター(WDC)が設立された1950年代から、“Full and Open Access”の理念のもと科学データの公開を行ってきており、現在世界中の100を超える拠点からデータが公開されている。さらに、気象や大気品質、自然災害などさまざまなセンサネットワークで観測されたデータが、NASAやNOAA、気象庁を始め各機関のWebサイトから時々刻々と配信され、TwitterやFacebook等のソーシャルメディアからは人々の動向を反映した情報がリアルタイムに発信されている。近年、実世界のヒト・モノをつなぐインターネット(Internet of Things, IoT)が脚光を浴びる中、こうした実世界の動きを捉えたさまざまな種類・分野の実世界ビッグデータの公開と流通は、今後ますます増えていくと考えられる。

ビッグデータの世界では、3V (Volume, Variety, Velocity) が重要であると言われ、これまでは主に圧倒的なデータ量 (Volume) をリアルタイム

性高く (Velocity) 扱えるようにすることで、従来のアプリケーション性能を大幅に改善するスケールメリットが重要視されてきた。しかし、オープンデータの広がりにより、全体を把握しきれないほど多種多様な (Variety) データが利用できる環境が整いつつある中、これらを横断的に組み合わせ“コネクションメリット”を発揮する技術へのニーズが高まってきている。本稿では、こうしたオープンデータの横断的利活用に関する取り組みの現状と、それらを支える基盤技術について概観する。

### マッシュアップデータの検索

Google Map や Open Street Map など、Web上で利用可能なデジタル地図の普及に伴い、さまざまな位置情報付きデータを地図上で組み合わせるデータマッシュアップが盛んに行われており、近年では環境問題や災害対策、保健衛生など実世界の課題解決への利用が進んでいる。たとえば、大気汚染の観測データと産業施設の分布データを重ね合わせ、各種産業が環境に及ぼす影響を調査したり<sup>☆3</sup>、緊急事態や災害に関する情報や、疫病やインフルエンザなどの保健衛生に関する情報<sup>☆4</sup>を世界中から収集し地図上でリアルタイムにモニタリングするサービスなどが公開されている (図-1)。

こうしたデータマッシュアップにより、さまざま

☆1 Data.gov, <https://www.data.gov/>

☆2 WDS, <https://www.icsu-wds.org/>

☆3 EPA Air Emission Sources, <http://www.epa.gov/air/emissions/where.htm>

☆4 Health Map, <http://www.healthmap.org/en/>

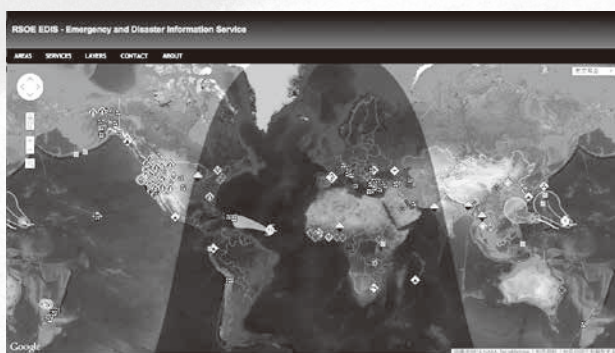


図-1 ROSE Emergency and Disaster Information Service <sup>☆5</sup>。1,000を超える情報源から発信された世界中の緊急事態や災害に関する情報および気象情報をリアルタイムに収集し表示

な分野のオープンデータを横断的に組み合わせ課題解決に役立てるためには、膨大なデータの中から地図上で重ね合わせて有意な発見につながるデータをいかに効率よく発見するかが大きな課題となる。しかし、Data.govやWDSなど、従来のオープンデータポータルは、メタデータによるカタログ検索しか提供しておらず、キーワードにヒットする膨大な検索結果リストの中から組合せ可能なデータを発見することは至難の業である。こうした問題を解決すべく、情報通信研究機構(NICT)で開発している相関検索システムCross-DB Search<sup>1)</sup>では、データが生成された場所や日時、データの観測対象に関する記述をメタデータから抽出し、データ間の地理空間的、時間的、テーマ的な類似性に基づく複合相関検索を行うことで、クエリにヒットするデータの“周辺データ”、すなわち類似した場所、時間において関連する対象を観測したデータを、芋づる式に発見できるようにしている。図-2の例では、WDSを対象に2001年以降に欧州で観測された降雨(precipitation)に関するデータを検索すると、クエリにヒットする降雨データだけでなく、その周辺で同時期に観測された環境汚染に関するデータも発見することができる。事実、異常な降雨が川を流れ湾の環境に影響を及ぼすことが論文等で確認されており<sup>2)</sup>、この検索結果は2009年から2010年にかけてアルプスで発生した異常多雨がライン川やポー川を経て、オランダ沿岸やイタリア沿岸に環境被害をもたらしたことを示している。こうした検索

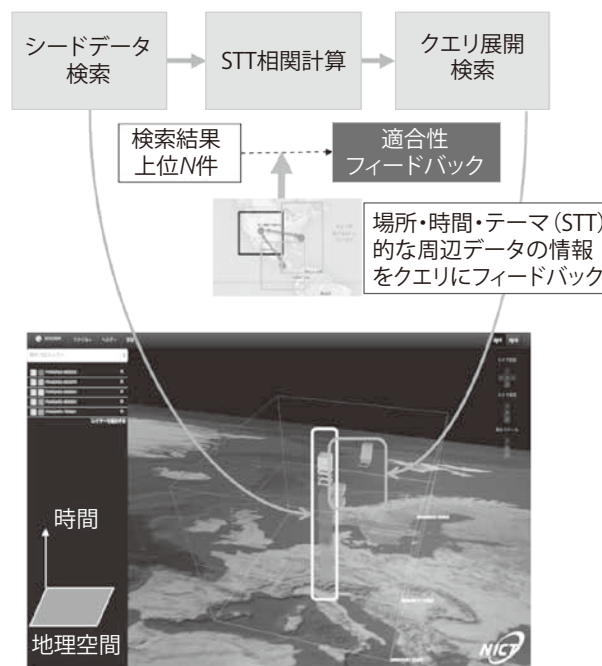


図-2 Cross-DB Searchによる相関検索。2009年のヨーロッパの異常多雨を示す観測データ(中央枠内)の周辺から、時間・場所・概念的に相関の高いプランクトン有毒化(右枠内)や植物の化学物質蓄積の観測データを発見し、異常多雨による環境被害の調査に利用

は、環境問題など、原因や影響が十分に解明されておらず科学的モデルの作成が困難な場合に、実世界で空間・時間・テーマ的に相関が高いデータを発見し分析の手掛かりを見つけることに役立つ。

さらに最近では、既存のオープンデータをマッシュアップするばかりでなく、ユーザがデータ収集や問題解決に直接参加する双方向サービスも広がり始めている。たとえば、Open Sense<sup>☆6</sup>(スイス)では、市民がモバイルセンサを持って大気汚染の観測に参加する見返りに、市内の大気汚染を避けた経路を案内するサービスを提供している。また、米国のPhilly 311<sup>☆7</sup>では、普段は市民からの道路や施設の破損、ごみの不法投棄などの苦情をスマートフォンから写真付きで受け付けるが、災害時には市民から被害情報を収集して地図上に表示し、行政が素早く復旧対応できるようにしている。こうしたサービスは、市民の力を使って効率的な行政サービスを実

☆5 <http://hisz.rsos.hu/alertmap/>

☆6 Open Sense, <http://www.opensense.ethz.ch/>

☆7 Philly 311, <http://www.phila.gov/311/>

現しようとするガバメント 2.0 の取り組みとして注目されている。

## 複合イベントの発見

アーカイブデータの分析だけでなく、IoT を介して環境の変化や人・モノの動きに関するデータをリアルタイム性高く収集し、横断的に分析することで、各種のスマートサービスをより高度化する試みも進められている。スマートシティでは、環境や交通、治安など都市の中で起こるさまざまな状況を幅広くセンシングし、公共サービスの効率化や非常時への迅速な対応を実現することを目指しており、たとえばバルセロナ市やシカゴ市では、交通量センサや環境センサ、IP カメラなどを Wi-Fi ネットワークで接続し、交通や治安など市民の生活の質を向上させるさまざまなスマートサービスを総合的に提供する試みが行われている。また、こうして集められたデータを市民や地域の企業・NPO に提供しサービスの改善に協力してもらう活動も進められており、たとえばサンフランシスコ市の DataSF<sup>☆8</sup> では、これまでに 200 以上のデータを公開し、60 以上のアプリケーションを市民に提供している。

これらのデータは、個々の目的に応じて実世界の事象（イベント）を断片的にセンシングしたものであるが、それらを組み合わせて実世界のさまざまな状況を総合的に分析する複合イベント処理が、スマートサービスを実現する上での重要な技術として位置づけられている。従来は、事前に検出したい複合イベントを決め、それに合わせて準備したセンサからデータを高速に収集し処理する技術が重要視されてきたが、オープンデータが普及し多種多様なセンサデータがリアルタイムに公開されるようになると、複合化可能なデータの組合せをリアルタイム性高く発見し統合する技術の必要性が高まっている。たとえば、ゲリラ豪雨や竜巻など突発的に発生する異常気象の周辺で、環境センサや交通センサ、ソーシャ

ルセンサ（SNS）などから時々刻々と発信されるデータの中から相関の高いものを発見し、多岐にわたる被害状況を把握するようなケースである。

異なる情報源からデータを収集し統合する技術としては、データウェアハウスが主流であるが、近年、センサから発信されるデータストリームを対象としたデータウェアハウスの研究が盛んに行われている。従来の蓄積データを対象としたデータ結合や多次元データ分析（OLAP）を、イン・メモリ処理、増分更新、クエリや導出テーブルの一貫性管理などにより、データストリームに対してリアルタイムかつ連続的に実行できるようにする技術が開発されている<sup>3)</sup>。その一方で、物理センサから発信される数値データのストリームだけでなく、監視カメラ映像や、ソーシャルメディアを流れるテキストや写真など、異種・異分野のストリームデータを対象とした複合イベント処理の研究も進められている。NICT で開発しているイベントデータウェアハウス<sup>4)</sup> では、気象などの物理センサから SNS などのソーシャルセンサまで、異種・異分野のデータストリームの変化の相関を分析し複合化可能なデータの組合せスケーラブルに発見できるようにし、ゲリラ豪雨や大気汚染（PM2.5）の被害分析などに応用している（図-3）。こうした技術は、さまざまなモノ・ヒトが接続する IoT のデータを横断的に統合し、さまざまなスマートサービスで利活用する基盤技術として期待される。

さらに、より複雑な複合イベント発見では、既存の情報源から提供されるデータだけでは不十分な場合も多く、新たなデータをオンデマンドに収集しなければならない。その解決策として、人々が携帯するモバイルデバイスをセンシング機器として用いることで、既存のセンサネットワークによるデータ収集を補強する参加型センシング（participatory sensing）が注目されている。近年では、クラウドソーシングによるタスク管理を駆使し参加型センシングを効率的に行う crowd sourced sensing<sup>5)</sup> に関する研究も盛んに行われている。

☆8 DataSF, <http://datasf.org/>

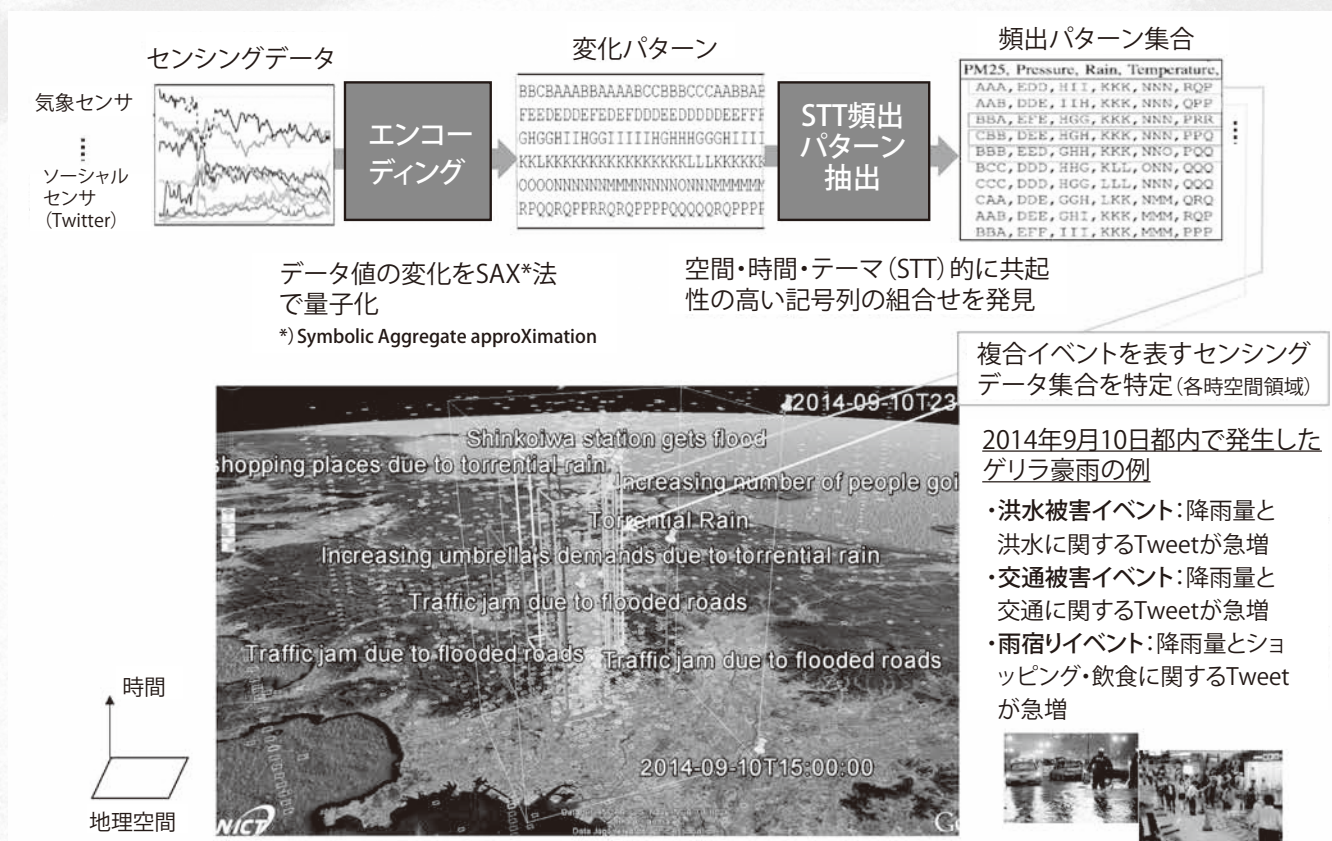


図-3 イベントデータウェアハウスによる複合イベント発見。気象センサの値が急に变化した地域・時間帯で同じく变化したソーシャルセンサのデータ（トピックごとのSNS出現頻度を計測）を特定し、さまざまな気象被害を表す複合イベントを発見

## 将来展望

実世界の課題解決のために、さまざまなオープンデータを横断的に利活用し、分野の壁を超えリアルとサイバーをつなぐことでコネクションメリットを創出するための取り組みはいまだ黎明期にあり、基盤技術も応用も試行錯誤の状態である。次の段階へと進化させていくには、複雑に影響し合う実世界のさまざまなイベントの間の関係性を、オープンデータを使って集団的に分析すべく、オープンサイエンス（あるいはシチズンサイエンス）<sup>6)</sup>のプロセスと連携し、コミュニティが中心となって原因の究明と課題解決を一体的に進めることが重要になる。さらに、環境や社会、経済活動にまで至るグローバルなIoTネットワークを構築し、その上に実世界のヒト・モノ・コトに関する情報を横断的に統合するプラットフォームを形成し、大規模なスマートサービス連携を実現するICT基盤の展開が今後期待される。

### 参考文献

- 1) Takeuchi, S., et. al. : Spatio-Temporal Pseudo Relevance Feedback for Large-Scale and Heterogeneous Scientific Repositories, In Proc. IEEE BigData Congress 2014, pp.669-676 (2014).
- 2) Weise, A. M., et. al. : The Link between Precipitation, River Runoff, and Blooms of the Toxic Dinoflagellate *Alexandrium tamarense* in the St. Lawrence, Journal of Fisheries and Aquatic Science, 59(3), pp.464-473 (2002).
- 3) Golab, F., et. al. : Stream Warehousing with DataDepot, In Proc. ACM SIGMOD 2009, pp.847-854 (2009).
- 4) Dao, M.-S., et. al. : A Real-time Complex Event Discovery Platform for Cyber-Physical-Social Systems, In Proc. ACM ICMR 2014, pp.201-208 (2014).
- 5) Boulos, K., et. al. : Crowdsourcing, Citizen Sensing and Sensor Web Technologies for Public and Environmental Health Surveillance and Crisis Management, Int'l J. Health Geographics, 10:67 (2011).
- 6) 林 和弘：オープンサイエンスをめぐる新しい潮流（その5）オープンな情報流通が促進するシチズンサイエンス（市民科学）の可能性，科学技術動向，150号，pp.21-25 (2015)。

(2015年5月22日受付)

是津 耕司（正会員） zettsu@nict.go.jp

1992年日本IBM、2003年通信総合研究所（現情報通信研究機構）を経て、2011年より同ユニバーサルコミュニケーション研究所情報利活用基盤研究室室長、博士(情報学)、キール大学招聘研究員(2009)、京都大学連携准教授(2008～2012)など、データベース、情報検索、ソフトウェア工学の研究に従事。