GOVLAB

THE POTENTIAL OF SOCIAL MEDIA INTELLIGENCE TO IMPROVE PEOPLE'S LIVES

Social Media Data for Good

Executive Summary

September 24, 2017

By Stefaan G. Verhulst and Andrew Young

GOVLAB DATA COLLABORATIVES

THE POTENTIAL OF SOCIAL MEDIA INTELLIGENCE TO IMPROVE PEOPLE'S LIVES

Social Media Data for Good

September 24, 2017

By Stefaan G. Verhulst and Andrew Young*

COVER IMAGE: YOLANDA SUN

* Stefaan G. Verhulst is Co-Founder and Chief Research and Development Officer of The Governance Lab
Andrew Young is Knowledge Director at The Governance Lab

THE POTENTIAL OF SOCIAL MEDIA INTELLIGENCE TO IMPROVE PEOPLE'S LIVES

By Stefaan G. Verhulst and Andrew Young

Executive Summary

The twenty-first century will be challenging on many fronts. From historically catastrophic natural disasters resulting from climate change to inequality to refugee and terrorism crises, it is clear that we need not only new solutions, but new insights and methods of arriving at solutions. Data, and the intelligence gained from it through advances in data science, is increasingly being seen as part of the answer. This report explores the premise that data—and in particular the vast stores of data and the unique analytical expertise held by social media companies—may indeed provide for ***a new type of intelligence that could help develop solutions to today's challenges.***

In this report, developed with support from Facebook, we focus on an approach to extract public value from social media data that we believe holds the greatest potential: data collaboratives. Data collaboratives are an emerging form of public-private partnership in which actors from different sectors exchange information to create new public value. Such collaborative arrangements, for example between social media companies and humanitarian organizations or civil society actors, can be seen as possible templates for leveraging privately held data towards the attainment of public goals.

THE PROMISE OF DATA COLLABORATIVES

Existing research on data collaboratives is sparse, but a number of recent examples show how social media data can be leveraged for public good. These include Facebook's sharing of population maps with humanitarian organizations following natural disasters; predicting adverse drug reactions through social media data analysis in Spain; and the city of Boston's use of crowdsourced data from Waze to improve transportation planning. These examples and 9 additional cases are discussed in the full report.

By assessing these examples, we identify **five key value propositions** behind the use of social media data for public goals:

1. SITUATIONAL AWARENESS AND RESPONSE

Data held by social media companies can help NGOs, humanitarian organizations and others better understand demographic trends, public sentiment, and the geographic distribution of various phenomena. In doing so, data contributes to improved situational awareness and response.

Case Studies:

- Facebook Disaster Maps
- Tracking Anti-Vaccination Sentiment in Eastern European Social Media Networks
- Facebook Population Density Maps

2. KNOWLEDGE CREATION AND TRANSFER

Widely dispersed datasets can be combined and analyzed to create new knowledge, in the process ensuring that those responsible for solving problems have the most useful information at hand.

Case Studies:

- Yelp Dataset Challenge
- MIT Laboratory for Social Machines' Electome Project
- LinkedIn's Economic Graph Research Program

3. PUBLIC SERVICE DESIGN AND DELIVERY

Data Collaboratives can increase access to previously inaccessible datasets, thereby enabling more accurate modelling of public service design and helping to guide service delivery in a targeted, evidence-based manner.

Case Studies:

- Facebook Future of Business Survey
- Easing Urban Congestion Using Waze Traffic Data
- Facebook Insights for Impact Zika

4. PREDICTION AND FORECASTING

New predictive capabilities enabled by access to social media datasets can help institutions be more proactive, putting in place mechanisms based on sound evidence that mitigate problems or avert crises before they occur.

Case Studies:

- Tracking the Flu Using Tweets
- Predicting Floods with Social Media Metatags
- Predicting Adverse Drug Events by Mining Health Social Media Streams and Forums

5. IMPACT ASSESSMENT AND EVALUATION

Access to social media datasets can help institutions monitor and evaluate the real-world impacts of policies. This helps design better products or services, and enables a process of iteration and constant improvement.

Case Studies:

- Sport England's #ThisGirlCan
- Using Twitter Data to Analyze Public Sentiment on Fuel Subsidy Policy Reform in El Salvador
- Using Twitter to Measure Global Engagement on Climate Change

RISKS

RESPONSIBLY SHARING SOCIAL MEDIA DATA

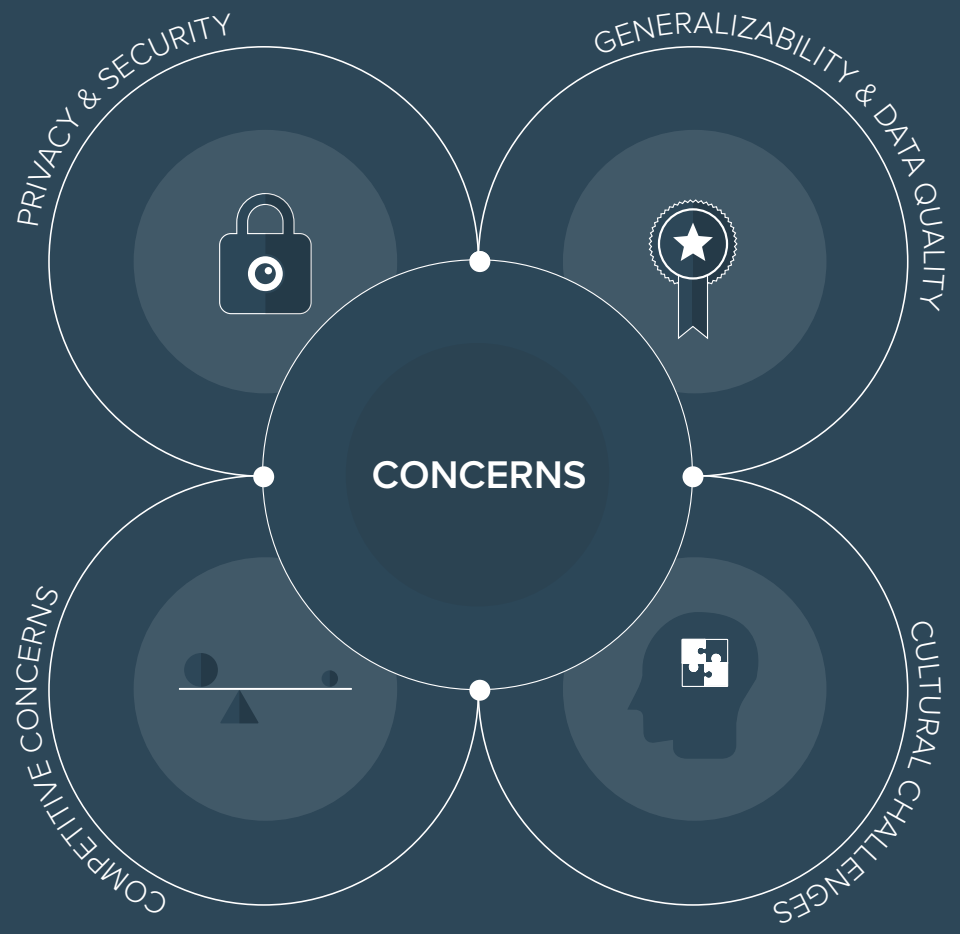

FIGURE 1

Despite the potential of data collaboratives, companies and public organizations often have concerns about sharing data. Many of these concerns are legitimate: data sharing is not without risks and challenges. We identify four key risks and challenges, and discuss ways to mitigate them. These include:

PRIVACY AND SECURITY

The most common concern expressed by individuals and companies involves the concern that sharing information may result in disclosing personally or demographically identifiable information, which may create privacy and/or security violations. Such concerns are not only natural, but very important: data sharing must not result in any dilution of protections for individuals, many of whom might not even be aware that the data was collected about them in the first place.

COMPETITIVE CONCERNS

Companies are often concerned that sharing data—usually without charge—will threaten their commercial interests or affect their competitive advantage. While such concerns are important to address, our research into the field of cross-sector data-sharing[1] suggests that this view is based on a false, zero-sum understanding of data collaboration and its potential. There are often methods of balancing competitive risk with data sharing for public good – such as aggregating data or sharing insights from datasets rather than the raw data.

GENERALIZABILITY, DATA BIAS, AND QUALITY

A key concern when using social media data involves the level of representativeness or data bias. Social media data is often gathered from a particular demographic subset, possibly ignoring so-called "data invisibles"—individuals, often from vulnerable communities, who are unrepresented in private or public datasets. For these reasons, caution needs to be exercised in extrapolating general observations from such data.

BARRIERS TO A CULTURE OF DATA SHARING AND COLLABORATION

Our exploration indicates that one of the chief obstacles to more widespread data sharing—and one that may underlie other concerns—is a lack of understanding about the benefits of sharing social media data, and a lack of comfort and familiarity with such strategies. Social media companies today operate in a rapidly changing environment where notions of collaboration, sharing, and mutual benefit are far more widely accepted as philanthropic and commercial propositions. Embedding these values in business operations is one of the central challenges—and greatest opportunities—facing social media companies today.

TOWARD DATA RESPONSIBILITY

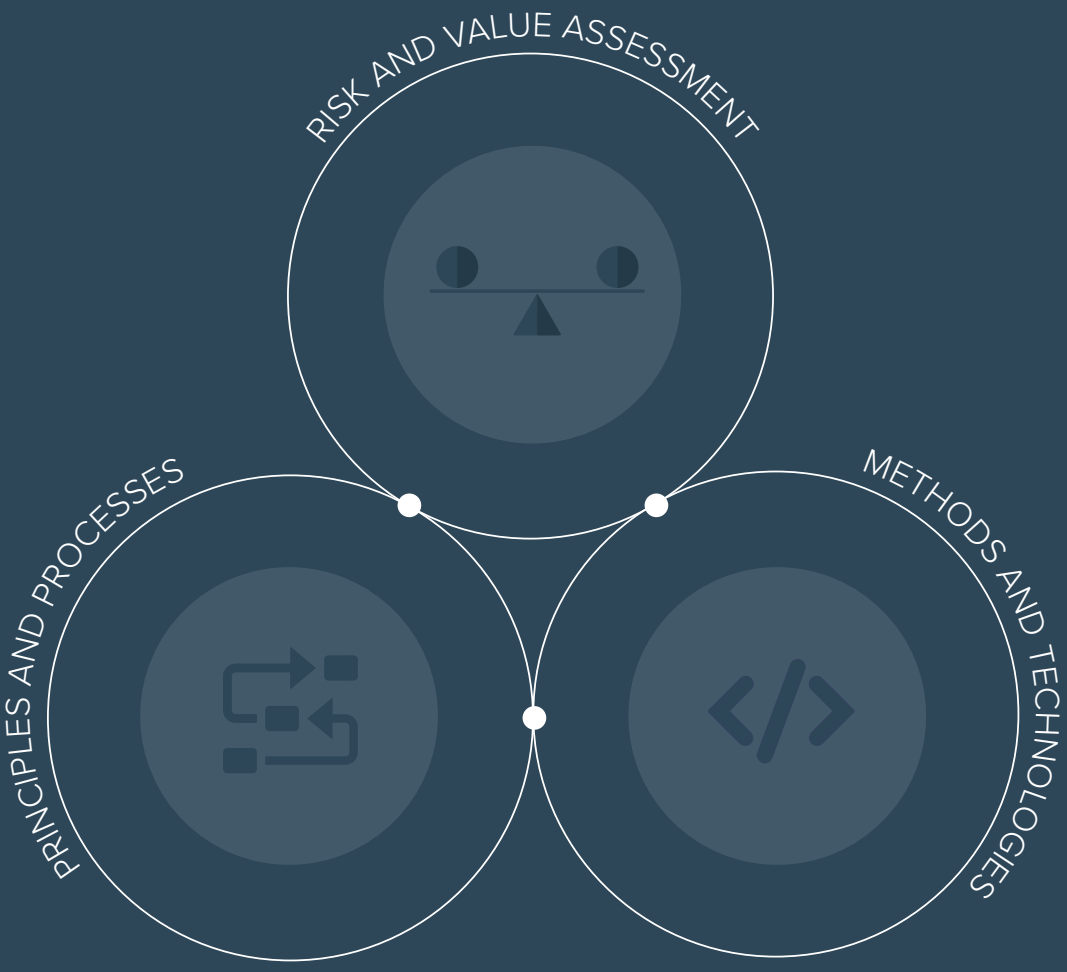

FIGURE 2

Some social media corporations and organizations, including Facebook, the International Committee of the Red Cross, and United Nations Global Pulse, have started to explore frameworks to share data responsibly. For such efforts to succeed, they will likely need to include the following four elements to assure concerns are addressed meaningfully and legitimately:

RISK AND VALUE ASSESSMENTS

Risks—including privacy, ethical and commercial concerns—exist across the social media data lifecycle, and include inaccurate, non-representative data entry during collection; insufficient, outdated, or inflexible security provisions during processing; incompatible cultural or institutional norms or expectations during sharing; aggregation or correlation of incomparable datasets during analysis; and controversial or incongruous data usage. Assessing such risks, and the risks of not sharing data with the potential for transformative impact, is an important task for all data collaborative efforts.

NEW METHODS AND TECHNOLOGIES

Data responsibility frameworks will also need to evolve beyond written policies to create tools and technologies that enable flexibility and context-specific implementation. Data responsibility decision trees, such as the Center for Democracy & Technology's Digital Decision Tool, can be used to translate principles into a series of questions. Consideration of and experimentation with differential privacy could also help to limit privacy risks. Similarly, a transparency report showing with whom data is being shared and toward what public benefit could help allay concerns about government misuse of private-sector data assets.

PRINCIPLES AND PROCESSES

Having a set of principles that can determine if and how social media data can be shared is important for accountability and effectiveness. While frameworks exist, data responsibility principles will need to be updated or newly designed to address emergent risks. Initiatives like the Signal Code and the Handbook on Data Protection in Humanitarian Action are important, but are limited to specific contexts and sectors. More broadly applicable and mutually agreed upon principles could lead to greater uptake and impact.

Principles and assessment frameworks have no value if there is no process in place to apply them or oversee their implementation. Including governance processes is a key element to achieve data responsibility. Such processes should be transparent and participatory, while being flexible and responsive to different needs and contexts. For example, to accommodate collaborative research using social media data, Facebook designed an innovative review process that involved in-house training, different stages of review, and the application of evaluation criteria to determine whether to go ahead.

REALIZING THE POTENTIAL OF SOCIAL MEDIA INTELLIGENCE THROUGH DATA COLLABORATION

From our discussion, we derive a series of recommendations for developing data collaboratives, which we have grouped into four broad categories. Taken together, these form a roadmap for companies or organizations considering using data for public ends.

STEWARDS

Social media corporations should consider themselves, and act as, the standard bearers for a new corporate paradigm of *data stewardship*. This coinage represents a move away from the concept of data as something to be owned and towards stewardship of data as a public good.

Social media companies should pioneer the role or position of Data Stewards within their organizations. Such individuals (or offices) would be tasked with matching and coordinating the demand and supply of social media data in a trusted way. These new positions could provide models for other companies considering the use of private data for public goals.

Among their other roles, **Data Stewards could help develop new coordinating mechanisms to unlock corporations' supply of social media data sets with potential public interest value.** Such mechanisms must include a due process to respond to data requests; a system for filtering or prioritizing certain kinds of information; and a method to ensure that the data being released matches public needs and demands.

EVIDENCE

A more detailed repository of case studies should be established to document impact and practice. Such a repository, which could build on the foundation offered by the case studies presented here, would highlight best practice in value propositions, technical arrangements, and legal frameworks for data collaboratives and give strategies for measuring impact.

Existing metrics and systems of measurement may not be sufficient to capture the true impact and value of data collaboratives. **Companies and organizations involved in data sharing can together help develop better metrics of value and impact**, and indeed, entirely new definitions of success.

We also **need a more granular understanding of how available data matches information needs**. Such an understanding can be achieved through data audit methodologies and tools.

METHODS

In order to scale the use of data collaboratives and realize the potential of the underlying data, **we need a better understanding and systematization of the methods for their deployment**. Organizations and researchers should mine existing data collaborative experiments, such as those to be included in the repository described in the previous section, for examples of successful practice.

Lessons and observations from this can be translated and shared as a toolkit or roadmap for corporations considering sharing data. They would form the basis of a published (and ideally, openly available) "Best Practices" kit that would constantly be updated and improved, based on the latest evidence.

We also need a better understanding of new techniques used by Data Collaboratives to analyze and seek insight from data. The field is built on a number of emergent techniques—including natural language processing, neural networks, computational social science, network science, sentiment analysis, data-mining and machine-learning—each of which needs further study in the context of data sharing .

Many nonprofit and civil society organizations lack the necessary expertise to apply these techniques within Data Collaboratives. **Corporate data sharing initiatives need also to consider sharing their expertise in handling data**—for example, through training initiatives, educational programs, informal peer- and practitioner-mentoring mentoring setups, and the development of affordable, user-friendly tools.

MOVEMENT

Data sharing cannot be thought of in isolation from other activities or environments. Much **more thought and energy needs to be directed toward developing a network for collaboration, supporting ecologies and a platform for sharing**. What's required is a Data Collaboratives movement.

To that end, **efforts should be made to bring together various actors from the social media data community**—including those currently acting as data stewards—at dedicated convenings to share lessons learned; identify pain points; and develop common solutions, procedures, and practices.

Corporations could help facilitating such convenings by providing a venue (virtual or physical) where data providers and users can build knowledge in an emerging field and co-create ideas and insights.

Creating a supporting ecology means interacting with and nurturing all actors in that ecology. As mentioned above, the practice of **data sharing needs to move beyond data per se to include a wider set of skills, expertise and knowledge areas**.

As the data collaboratives and social media intelligence movement continues to take shape, **engagement with the populations whose attributes and behaviors often make up the data held by social media companies will be key**.

In addition to engaging with other skills, there also **needs to be broader engagement with other actors whose decisions can play an important—perhaps even determining—role in the success of data collaboratives**. Such actors could include policymakers, regulators, and potential funders, both for-profit and nonprofit (philanthropic).

...we look to the potential of leveraging data to provide a new type of intelligence that could help develop global solutions to global challenges.

THE POTENTIAL OF SOCIAL MEDIA INTELLIGENCE TO IMPROVE PEOPLE'S LIVES

Social Media Data for Good

facebook