

## DATA CORPS

*Tackling local, national, international challenges by getting our hands dirty with data!*

**VISION:** The *Data Corps* is envisioned as an effort to help unleash the power of data at the *local, state, national, and international* levels in the service of science and society by providing practical experiences, teaching new skills, and offering teaching opportunities in data science to U.S. data scientists and data science students.

**MISSION:** The mission of the *Data Corps* is to:

- (i) Enable U.S. data scientists and data science students to obtain practical experience with data-intensive applications;
- (ii) Promote a better understanding of the power of data, and the role that data can play in addressing issues at the local, regional, national, and international levels;
- (iii) Teach data literacy and provide basic training in data science to the existing workforce in communities, organizations, and institutions at the local, state, national, and international levels.

*Data Corps* builds upon and enhances American leadership in data science and helps to increase American competitiveness in the data-driven global landscape by enabling sound data science practices across communities.

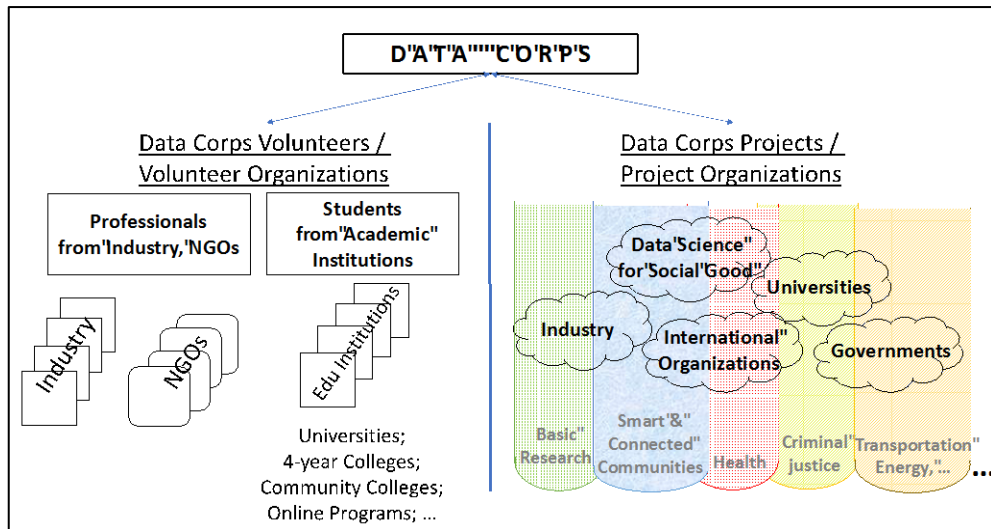
**BACKGROUND:** In practice, using data involves work. For example, one must understand the dataset context; the data may have to be cleaned and filtered; they may need to be “wrangled” into different formats; data collection and curation methods may require improvement; the data may need to be analyzed and visualized using alternative methods; and, finally, the findings need to be communicated to the intended audience to inform decision-making. There are many examples in practice where data collected are left unused or underused<sup>1</sup>. In other cases, even rudimentary approaches to data collection—let alone data analysis—have not taken root among organizations. *Data Corps* would engage with participating organizations/institutions across rural, inner city, state, national, international communities to enable increased use of their data assets, while providing practical, hands-on experience to *Data Corps* volunteers. Proficiency in data collection, data curation, data management, data analysis, and data-driven decision-making is increasingly becoming a major differentiator between efficient, rich, highly functioning communities versus less efficient, poorer communities. Bridging the data-to-knowledge gap will enable communities at all levels to make better use of their data assets. *Data Corps* will focus especially on underserved communities.

Interactions between *Data Corps* volunteers and the communities they serve will be mutually beneficial—enabling data literacy and transfer of data science skills to local communities, on the one hand, while providing insights, practical experience, and teaching opportunities to participating data scientists, on the other.

---

<sup>1</sup> **Example:** Telephone call metadata (e.g., time, location, forwarding information, etc.) are collected for all 9-1-1 emergency calls at 1000’s of Public Safety Answering Points (PSAPs) across the country. There is no standardization in data collection. The data are primarily used for administrative purposes, even though some basic data cleaning would enable visualization of the data to provide a holistic view of the 9-1-1 system and situational awareness to emergency responders. Spatiotemporal analysis of such data could be used to build predictive models for 9-1-1 calls, by combining these data with a variety of other data in each region, including traffic, weather, current events, etc. (NSF Proposal IIS-0429448 on *Spatiotemporal Analysis of 9-1-1 Call Stream Data*).

**STRUCTURE:** The *Data Corps* provides a (lightweight) organizational structure to (i) help establish and disseminate best practices; (ii) maintain lists of volunteer organizations and project organizations; and, (iii) provide a brokering function between these sets of organizations, to match *Data Corps* volunteers to projects in the field (see figure). The *Data Corps* could enhance existing internship and study abroad programs at academic institutions by providing a data science “track”/focus. Programs like



the National Science Foundation’s Big Data Regional Innovation Hubs and Spokes and Smart & Connected Communities, which already have a community-engagement component, and similar programs in other agencies, would be ripe targets for the *Data Corps*.

**INAUGURAL WORKSHOP**

The *Inaugural Workshop on Data Corps* will be held in December 2017 at Georgetown University, and will include representation from the full range of stakeholders:

- Organizations that provide the volunteers, such as academic institutions (including, minority-serving institutions and community colleges), NGOs, industry;
- Organizations that can offer projects, e.g., local/state/federal government agencies, industry, NGOs, academic institutions, international organizations;
- Organizations engaged in related programs, e.g., Data Science for Social Good; The Opportunity Project, Department of Commerce; NTIS/18F; DataKind; rescue.org
- Professional organizations, e.g., ACM, IEEE, ASA, etc.
- Organizations interested in funding the activity, e.g., private foundations, government agencies, industry, international organizations such as ITU and The World Bank, etc.

**Workshop Organization**

The inaugural workshop will be organized by Prof. Michael Bailey, Georgetown University, working with an Organizing Committee. High-level objectives of the workshop (not necessarily in order):

- Obtain an overview of any existing, related activities across the full range of stakeholder communities, including overview of:
  - Student programs and study abroad programs at universities, colleges;
  - Organizations operating knowledge exchange programs in the US and/or internationally;
- Identify needs and requirements from the perspectives of individual Data Corp volunteers and the organizations they may belong to, and from the perspective of organizations that offer/host Data Corps projects and project experiences.
- Identify the desired characteristics of a Data Corps, and the role(s) it could play in the volunteers/projects ecosystem.
- Discuss what outreach is needed to broadly engage all relevant stakeholders;

- Identify first steps, and next steps, for a Data Corps pilot program;
  - This may include additional workshops in US, and abroad?

**Workshop Organizing Committee (under construction)**

- Professor Michael Bailey (Chair), Interim Dean, McCourt School of Public Policy, Colonel William J. Walsh Professor, Department of Government, Georgetown University
- Joanna Chan (Co-chair), AAAS Big Data Fellow, National Science Foundation
- Rene Baston, Executive Director, NSF Northeast Big Data Regional Innovation Hub, Columbia University
- Melissa Cragin, Executive Director, NSF Midwest Big Data Regional Innovation Hub, University of Illinois
- Gary Fowlie, Head, International Telecommunication Union Liaison Office to the United Nations
- Graham Harrison, Program Manager, Office of International Science and Engineering, National Science Foundation & The World Bank Group
- Prasanna Lal Das, Lead KM Officer (Data), Trade & Competitiveness, The World Bank Group
- Tania McVeety, Manager, Global University Programs, Corporate Technology, IBM
- Donna Medeiros, Director of Research, Data, Analytics, International Rescue Committee
- Trevor Monroe, Big Data Initiative Lead, The World Bank Group
- Renata Rawlings-Goss, Co-Executive Director, NSF South Big Data Regional Innovation Hub, Georgia Tech
- Lea Shanley, Co-Executive Director, NSF South Big Data Regional Innovation Hub, UNC Chapel Hill
- Sarah Stone, Deputy Director, NSF West Big Data Regional Innovation Hub, University of Washington
- Drew Zachary, Project Lead, The Opportunity Project, Department of Commerce

**Contacts:**

For further information please contact:

Joanna Chan, AAAS Big Data Fellow, CISE Directorate, National Science Foundation, [jchan@nsf.gov](mailto:jchan@nsf.gov)