Current draft: 18Oct2019.1156pst

Journalism and Vote2020

Threats against the integrity of our elections challenge American democracy.

Today's partisan environment could destroy America's trust in our election systems in 2020.

Can we verify the integrity of our elections in a way that most Americans would trust? Specifically, can American journalism build a national common understanding of how to monitor the integrity of our democracy?

On election day, news organizations now rely on voting counts from 6,459 major jurisdictions, and several thousand subjurisdictions. All together, there are over 9,000 voting jurisdictions that count votes.

This is a proposal to build a parallel data system, joining thousands of students and faculty across America in the national data science curriculum into partnership with news organizations, to build a new way to monitor election integrity, understand the data, and share analysis of the elections in those jurisdictions in an open and transparent way.

Many efforts to detect and prevent attacks on our voting systems are underway, at a national, state, and local level.
The US Election Assistance Commission (EAC) has distributed almost $4 billion dollars to hundreds of jurisdictions; the Department of Homeland Security monitors and mentors hundreds of jurisdiction on hardening their systems; the National Security Agency, and the US CyberCommand are running training exercises to defend our election systems from threats of interference.

But American news organizations have a unique, independent capability and a special journalistic responsibility to publically report on the integrity of elections, whether government-run or privately-run. All news organizations are designing their systems now to report on the 2020 election.

Here is the important convergence: As American news organizations build their capabilities to report on the election, they can build, at the same time, elements of what America needs to *protect* its election systems--a new, vast network of local sources, local data, and local analysis able to detect anomalies and irregularities.

As election systems have come to rely almost entirely on information technologies, new

capabilities are inextricably joined with new vulnerabilities. Estimates range into the billions of dollars to update or replace voting systems in America's nine thousand voting jurisdictions to reach desirable levels of security. Primary elections in 2020, and the general election in 2020 will run on antiquated and vulnerable equipment. Time and money have run out. There is no practical way to change the voting systems for the majority of these jurisdictions.

But we can build reporting systems, news systems, data analysis systems to significantly increase our trust in election results. These systems will require new initiatives by American news organizations, joined with new capabilities of data acquisition and analysis developed in the rapidly expanding world of data science. The expertise to do this lies in American companies, American universities, and American institutions devoted to national security.

This site outlines a workable approach to join these communities in a common effort for America's elections.

Public media

- Major US papers
- Television news
- Twitter/Facebook/Reddit

Human incentives

Trials and demonstrations

Building

How to build the system

How can ethical journalism help an American citizen understand how our election systems work, how our systems can be improved, and how we can decide to trust the results?

Our best way to answer that question is to observe these systems at work, continuously, throughout the process, assess performance, and report the results.

This site proposes several ways to build a national continuous news network to examine our critical national infrastructures.

As we develop methods to examine our critical election systems, those same methods may be applied to America's other critical infrastructures, those controlling power, water, information, transportation and financial systems.

Call it *continuous, open, visible, scalable data journalism*.

We propose to use our most critical democracy infrastructure, our election system and the data it generates, as the foundation for fact-based, local news in every community in the country.

Where do we find people in every jurisdiction to build the data and perform the local jurisdiction analysis?

Today, thousands of data science faculty members are hunting for relevant data sets and relevant analytic frameworks to assign to their students as homework. Build a Jupyter Notebook for analyzing data from any voting jurisdiction, and any academic institution in that jurisdiction will use it.

How many students become potential partners? Hundreds of thousands of students are taking the new Data Science courses today.

Today's partisan environment, amplified by unethical journalism, may destroy our trust in our election results. Can we build a common understanding of how to verify the integrity of our democracy?

Can post-election audits of the results of an election establish to the satisfaction of all that the results were fairly decided, and reflect the actual vote?

And how do we build a consensus on the integrity, the good faith and honesty of the millions of Americans who will participate in carrying out the processes and procedures of the election? Our registrars, our election boards, our technical staff, our volunteers in fire stations and post offices and schools across the country?

#Challenges

Will 9,000 US voting jurisdictions be able to safeguard their voting systems in 2020? Do they know how?

- Is there money enough and time enough to upgrade or replace the most vulnerable elements of today's election and voting systems? in the next five months before the primaries? or in the next twelve months in the runup to November, 2020?
- What are those vulnerable elements, and how do I know if they are used where I vote? Or used in swing states or closely contested elections that I care about?
- What do we need to do to put in place new auditing systems to increase our trust in the accuracy of election results?
- Can we harden our systems against tampering, and detect tampering when it occurs? Which elements are easy to harden, and which elements are difficult? What would it cost to improve systems for a 100,000-person city election, or for a 747,000-person Congressional Distict election?
- What role should Federal agencies play in US elections? in local elections? What roles should US CyberCommand, NIST, DARPA, DHS, the EAC, the Department of Defense, and the Department of Justice play in strengthening our election systems?
- What role can State Secretaries of State, or local Registrars or Election Commissions play in strengthening our election systems?
- Who is doing a good job? Who could do better?

How can a news organization with a few hundred employees create data gathering partners in 9,000 jurisdictions?

Impossible a few years ago. Easy, today, with the explosion of Data Science courses and Jupyter Notebooks.

Data Journalism

Building journalism data science templates for the 50,000 students now taking data science courses

Thousands of data science faculty members are hunting for relevant data sets and relevant analytic frameworks to assign to their students as homework. Build a Jupyter Notebook for analyzing data from any voting jurisdiction, and any academic institution in that jurisdiction will use it.

This is the answer to the problem of recruiting data sources and data analyses from 9,000 voting jurisdictions.

See EdX Data Science courses from Berkeley, Harvard, MIT

Sign up for the UC Berkeley Data Science 8x course. Free. Watch the videos of the first week's classes.
UC Berkeley Data Science 8x: 65,000 now enrolled

#Goals

Goals

1. Improve professional journalism resources in covering all aspects of Election2020
2. Improve journalism training in high schools, higher education, and corporate reporting
    - Numeracy
    - Data Science
    - Visualization
    - Connection of cause and effect
    - Ethical responsibility

Integrity

Integrity of American journalism

The integrity of American journalism may be the most important force determining how American voters view the validity of the highly contested 2020 elections.

And the integrity of American journalism is under attack.

This is an opportunity to build a new level of belief in journalistic integrity, using our ability to examine facts generated by our election systems, evaluate them in real time, report on them, and analyze them to validate or challenge election results.

How well we do this may affect how American democracy persists | develops | survives | continues.

By building the tools of careful data examination, we demonstrate how to build believable journalism. And, with believable journalism monitoring our elections, we demonstrate how to build belief in the integrity of our elections, and, based on that success, we demonstrate the integrity of our news organizations.

Building continuous reporting in every jurisdiction

Over the next twelve months, careful examination of the exercise of democracy in every jurisdiction will deeply affect how America trusts the result of the vote. Our democracy depends upon this trust.

This examination must be continuous, not episodic. Local, not regional or national. Distributed, not centralized.

This examination should extend to every jurisdiction at every level, not just a State Election Board or a Secretary of State.

This examination must enable fair comparisons between jurisdictions, and establish accountability for failures in integrity.

This examination must be carried out at the local jurisdictional level. The integrity of that examination must be accepted by those in that jurisdiction. A key step to accomplish that is to involve as many people in that jurisdiction as possible in the fact-based process of observing

its daily workings.

Local data-based journalism requires a partnership between traditional news organizations and local data reporting and analysis groups. Today, we have the tools to build this partnership. Today, the explosion of data science courses in every community college, college, and university creates tens of thousands of potential partners--an *ElectionCorps*--of capable local observers.

Every election has a story that begins long before the day of the vote. We must build those stories, in every jurisdiction. We must examine how the people and the institutions charged with the public trust to carry out fair elections actually carry out their responsibilities, particularly when technical and human vulnerabilities are under attack.

#NBC actions

NBC actions

Inventory NBC Resources

- New York
- Bureaus
- O&O
- Affiliates
- Producers
- Reporters
- Analytics and Statistics Department: Existing Jupyter Notebooks; present and planned map and interactive displays for anchors, guests.
- Web site engineering: design of data visualizations, maps: move to Jupyter Notebooks, using new J3-like capabilities.

Assess time investment in working with other organizations

- Discuss alliance with NYT Future Journalism group

Build web site incorporating example Jupyter Notebooks;

This week, set up tutorial sessions or plenary presentations of NBC's new Continuous Data Journalism

- Present at all journalism conferences cited below. Provide tools to all conference attendees.
- Pitch Knight Foundation, Pierre Omidyar for funds

Create national movement to cover local voting conditions

Develop local reporting and analysis in every jurisdiction, partnering with affilites and O&O's, and with local educational institutions teaching data science.

- 9,000 voting jurisdictions;
- 37,000 special districts: water, power, irrigation, storm ;
- 20,000 school districts, public and private;
- 20,000 municipalities

Build data journalism foundation course on EdX

Collaboration between NBC, AP, and other news organizations, and:

1. UC Berkeley Graduate School of Journalism, Columbia Journalism School, Missouri School of Journalism,
2. National journalism associations
3. Retired journalists, statisticians
4. National engineering societies: IEEE, ACM, AWWS
5. UC Berkeley Institute of Data Science; MIT Data Science; Stanford Data Science;

Use traditional beat reporter structure at national and affiliate level, but add more local stories, and recruit new local sources and data to update them regularly.

- Pick representative districts, find representative people, cover over time. Recurrent updates, maybe bi-weekly. Build inventory of compelling people. Later, bring them together physically, to exchange stories, once audience is built for their stories.
- Provide overall template evaluating how a district is doing (technology, costs,performance, training, bang for the buck, competence)
- Build reporting infrastructure nation-wide in every district using NBC standard evaluation template:
    - High school newspapers and video reporters (now streaming on Youtube) equipped with NBC templates
    - Higher ed (community colleges, colleges, universities) newspapers building data (Jupyter Notebooks) for every voting district
- Affiliate, O&O teams holding regional summits to build NBC grassroots coverage
    - Use local coverage of local election machinery and institution to build local history of elections:who did what when.
    - Partner with high school and community college faculty to teach course on data journalism, using election data as core assignment material
        - Use NBC local templates as foundations for a sequence of homework assignments in data science classes
    - Create local NBC annual prize event (Data Journalism Emmy) for student, non-traditional coverage of local elections
        - Use NBC on-air talent to create a three-hour EdX course on how to build election-related, or infrastructure-related podcast and videoblog
    - Partner with popular technology vehicles: Popular Science, Popular Mechanics, hundreds of websites and podcasts, to do the local investigatory work.
    - Create NBC Investigatory Data ethical standards, publish them. Similar to Fact Checking guidelines, but for technical subjects of infrastructure.

- Use the hundreds of Ethics in Data courses blossoming on the thousands of campuses now teaching Data Science.

Build new NBC Data Journalism for the 21st Century initiative using national journalism conferences

Organization	Members	Convention Dates
Journalism Education Association (JEA)		
JEA/NSPA National High School Journalism		21-24 Nov 2019
National Scholastic Press Association (NSPA)		November 21-24, Marriott, DC
Journalism Funders Conference		29-30 October 2019 San Francisco: HenryJ.Kaiser Family Foundation185 Berry Street, Suite 2000San Francisco, CA 94107
National College Media Convention		31 Oct - 3 Nov 2019: Grand Hyatt, DC
Nordic Data Journalism Conference		15 Nov 2019
ACP/CMBAM National College Journalism Convention		Feb. 27-29, 2020 · Hyatt Regency · San Francisco

Reporter's questions to build baseline data for a jurisdiction for Election 2020

Goal: build a Reporter Template to capture the basic outline of a registration and voting system in a jurisdiction, with figures of merit.

Whether it's a high school student body election, a school board election, a city council election, a congressional election, or a Presidential election, common elements enable comparisons and rankings.

How can a reporter build a picture quickly, for any jurisdiction?

Core question	Basic Answer	Figure of Merit	Example
Who could vote?	Geography; age; citizenship; status (felon,service member)	Eligible/Total population; Registered/Eligible; Voted/Registered ; VEP is best: excludes non-citizen(8.3%, or 21M), felons(3.1M), in-prison(1.4M), parole(.5M), probation(2.1M)	(VAP) Voting Age Population -> (CVAP) Citizen Voting Age Population -> (VEP)Voting Eligible Population (exclude in-prison ,felons, parole, probation,non-citizens;add overseas) : VAP.2016: 250.1M ->VEP.2016: 230.6M; VAP.2018 255.1M -> VEP.2018 235.7M
Who was registered to vote?			REG.2018: 211 million
How many applied for registration between 2016 and 2018?			80 M
How many			

registrations were denied?			
How many registered voters were purged from the list between elections?			Purge.2018 : 17 M
Do you have Election Day Registration (EDR) or Same Day Reg (SDR)? How many used it?			EDR.2018 800,0000 ; 50% of states
Do you have online registration?			75% of states
Who did vote?	Geography; age; citizenship; status (felon,service member)	Voted/Eligible; Percent by age; by party; by sex; by renter/owner; by race; by occupation Higher percentage is better	Total Vote / VEP; Presidential.2016: 138.6 M ; Congressional.2018 : 118.6 M / VEP.2018 = 50.3% ; Percent of registered: 118.6/211 = 56%
Who did not vote?		DidNotVote/Eligible; Percent by age; by party; by sex; by renter/owner; by race; by occupation	
How many vote in person?			
How many			

vote by mail or absentee?			
How many Uniformed and Overseas Citizen Absentee (UOCAVA) voted?			UOCAVA.2018: 350,000
How many vote early?			
How many polling places?			2018: 200,000 polling places
How many poll workers?			2018: 600,000 poll workers
How many pieces of voting equipment deployed?			2018: 300,000 (EAC must be wrong, that's only 2 scanners or voting machines per polling place)
How are boundaries decided?	How we measure fairness, balance. Who designed boundaries. How. Who accepted boundaries. Who can change them.	Percent by party; by race; by sex; by turnout	

How do voters register?	Process to apply; process to determine if applicant meets criteria; how are applicants notified? What is process of appeal?	Time; cost; availability or barriers to registration; percent rejected	
How are ineligible registrants purged?	Number of dead per district per year, Number of changed address or name; decided by what process? How to appeal?	Percent per year; error rate;	
How are purged registrants notified?	How soon after decision? How informed of appeal process?		
How are registration records made public?	Open access	Cost; time;	

Core question	Basic Answer	Figure of Merit
What voter data is available to campaigns? to the public? to the press?	All collected data fields are available; free; within 1 day of official DB update; online	cost; timeliness; accuracy
What are all the collected data fields? What is the data schema for registrants? What unique identifier allows individual comparison between voter and other official databases?		
Who decided what data to collect? How can that decision be changed?		
Who can change a voter registration record? What audit trail of changes is maintained? How may inaccurate changes be challenged? by the voter? by others?		

Core question	Basic Answer	Figure of Merit	Example
Do you have post-election audit?			PostAudit.2018: 80% of states
How does it work?			
Do you use Risk Limiting Audits?			
Do you know what they are?			

- Behavioral modeling
- Social media targeting
- Advertising models
- Money flow
- Information flow
- Building journalistic competence
    - Numeracy

- Levels of involvement
- High school Journalism
- College journalism
- Journalism schools

Reporters

Reporter	Organization	Story
Joseph Marks	[Washington Post]()	
Tonya Riley	[Washington Post]()	
Jacob Ward	[NBC]()	
[Bill Clark]()	[Wired]()	
[Andrew Harrer]()	[Bloomberg]()	
[Steven Stover]()	Army Times	Avengercon III
[Olivia Gazis]()	CBS	NSA role
Jordan Wilkie	Carolina Public Press	NC Board of Elections: 18Oct19
Karolina Buczek	[WLEX-18-NBC: Louisville, KY]()	Kentucky State election security: 16Oct19

Design a framework for continuous coverage stories, allowing continuous updates

How to generate continuous coverage of the state of America's election infrastructure?

Same challenge for news organizations as in covering electric or water infrastructure:

- Ubiquituous: 9,000 election districts; 15,000 water districts in California;
- Highly skewed: a few very big ones, a huge majority very small ones
- Ostensibly apolitical, but fundamentally political; viz Flint
- Boring, hard to explain why a power system failed (except PG&E) or a water system failed (Flint); but, instant understanding of impact on lives when failures occur. Leads to stories on how to prevent failures in the future.But, these stories are rarely followed up; news organizations move on to the next big thing. That's what *news* means.
- So, move beyond short news cycle to *continuous news cycle.*
- *This is the central idea of continuous data journalism*. Automated, continuous coverage, with crowd-sourced local expertise using the same evaluative templates.
- *Think of this as building a logistics or industrial maintenance system, but for facts.*

Who runs our elections?

- Daily profile of a registrar: rural, urban, swing district, historically Jim Crow district, previously-sanctioned district.

- Follow five registrars for the next year, checking in weekly to hear their progress in maintaining and improving the integrity of their office

    - What challenges they face
    - What they have to do to get adequate resources
    - How they recruit and train their core staff, and prepare for rapid recruitment for recounts and election day
    - How they navigate the regulatory and legal environment. Exemplary case: huge fight this week in Kentucky, captured by affiliate WLEX: NBC coverage of meltdown fight

between Secretary of State and incompetent Election Administrator

- Washington Post
    - Joseph Marks: WashPost

Interview templates

- Establish baseline facts
- Assess current state of voting system:
  - budget,
  - political management,
  - history,
  - performance metrics
- Registrar story
- Secretary of State story

#What do we have to work with?

520,000 elected officials in the US know how they were elected, who ran the elections, and how well the election was conducted. Millions of defeated candidates also have observations about their local elections.

600,000 poll workers, and tens of thousands of civil servants working in America's 9,000 jurisdictions have observations.

Tens of thousands of students of data science, now on almost every higher education campus, and rapidly expanding in high schools and online courses.

Every news organization preparing for coverage of the 2020 election.

This site explores the possibility of creating a new, decentralized network of data sources and data analysis to monitor our elections, joining national news organizations, local news organizations, and social media reporting entities with tens of thousands of data scientists, nationwide, in a common effort to acquire and analyze election system data.

This site explores how to shift from episodic news coverage to continuous coverage, utilizing new tools of data acquisition and data analysis. From discrete to continuous.