

# Andrew L. Halterman

Department of Political Science, Massachusetts Institute of Technology  
E53-406, 77 Massachusetts Avenue  
Cambridge, MA 02139  
ahalt@mit.edu | andrewhalterman.com | github.com/ahalterman

## EDUCATION

**Massachusetts Institute of Technology**, Cambridge, MA Sept. 2015–present  
PhD Candidate, Political Science  
Research areas: quantitative methods, natural language processing, security studies

**Amherst College**, Amherst, MA Sept. 2007–May 2011  
Bachelor of Arts, *magna cum laude*, Political Science

## PROFESSIONAL EXPERIENCE

Machine Learning Engineer *Kensho Technologies* May 2017–August 2017  
Washington, DC

Implemented data pipelines for transforming text into structured event data. Developed techniques for automated dictionary development. Deployed a custom text geoparsing system.

Technical Consultant *various* May 2017–present  
Washington, DC, Boston, MA

Provided technical advice and custom software for several clients including university research projects, a nonprofit, and a sports betting operation. Projects included automating predictive models, developing new pipelines for extracting meaning from text, and code to manage active learning annotation with multiple coders.

Analyst *Caerus Associates* July 2013–June 2015  
Washington, DC

Member of Caerus's data science team, working on automated extraction of political information from news sources, including securing new source material, building models and visualizations using event data, and improving the data production process. Built and applied open source software for extracting meaning from large datasets as part of a DARPA flagship big data program.

Research Intern *Center for Strategic and International Studies* Sept. 2012–May 2013  
Washington, DC

Fulbright Fellow *Kosovar Institute for Policy Research and Development* Sept. 2011–June 2012  
Prishtina, Kosovo

## SKILLS

*Data Analysis*: fluent in R and Python, large datasets (500 GB+), machine learning, Python and R package creation, visualization (ggplot2, ggvis, Shiny), reproducible research, NLP and data extraction from text, database creation and management

*Software*: Python, R, spaCy, Keras, Docker, SQL, MongoDB, Elasticsearch, Unix/Linux/shell, LaTeX, CoreNLP, git

*(Human) Languages*: German (full proficiency), Gheg Albanian (conversational)

## AWARDS and HONORS

NSF Graduate Research Fellowship  
Presidential Fellowship, MIT  
Fulbright Research Fellow, Kosovo, 2011–2012  
Amherst College Dean of Faculty Student Research Grant  
Amherst College Center for Community Engagement Student Research Grant  
Amherst College Alpha Delta Phi grant for thesis research  
Robert C. Byrd Scholarship  
National Merit Scholarship

## PEER-REVIEWED PUBLICATIONS

1. “Adaptive Scalable Pipelines for Political Event Data Generation,” The IEEE International Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD 2017). Boston, Massachusetts. 2017 (with Yan Liang, Phanindra Jalla, Solaimani Mohiuddin, Manar Landis, Jill A. Irvine, Christan Grant)
2. “Mordecai: Full Text Geoparsing and Event Geocoding,” *The Journal of Open Source Software*, vol. 2, no. 9 (2017)
3. “Generating Political Event Data in Near Real Time: Opportunities and Challenges,” (with John Beiler, Patrick Brandt, Erin Simpson, and Philip Schrodt), *Computational Social Science: Discovery and Prediction*, ed. R. Michael Alvarez, Cambridge University Press, 2016.
4. “Bounded Altruism: INGO Opportunities and Constraints during Humanitarian Crises and US Intervention,” (with Jill Irvine), *Journal of East European Politics*, August 2014.

## CONFERENCE and WORKING PAPERS

1. “Linking Events and Locations in Political Text,” accepted at the Text as Data conference, Seattle, WA, 2018.
2. “How Right Wing is Right Wing Populism? Evidence from the Manifesto Corpus,” presented at the Manifesto Corpus Conference, 1-2 February 2018 in Berlin (with Jill Irvine and Nicholas Halterman)
3. “New Techniques for Coding Political Events Across Languages,” IEEE 18th International Conference on Information Reuse and Integration. Salt Lake City, UT. 2018 (with Yan Liang, Christan Grant, Jill A. Irvine, Khaled Jabr).
4. “Creating an Automated Event Data System for Arabic Text,” Annual Meeting of the International Studies Association (ISA). San Francisco, CA. 2018 (with Jill A. Irvine, Christan Grant, Khaled Jabr, Yan Liang)
5. “Mining News Stories for Predictive Signals of State-Led Mass Killing,” presented at the International Studies Association conference, 2016 (with Benjamin Valentino and Jay Ulfelder)
6. “Forecasting Anti-Regime Mobilization Using Structural Variables and Event Data,” International Studies Association Annual Meeting (February 2015).
7. “A New, Near-Real-Time Event Dataset and the Role of Versioning,” European Network for Conflict Research (ENCoRe) Fall Conference, Uppsala, Sweden (October 2014). [[Paper](#), [Code](#)]
8. “Measuring Political Mobilization: Insights from Massive Machine-Coded Datasets,” (with Jill Irvine), International Studies Association Conference Paper (March 2014). [[Paper](#), [Code](#)]
9. “Bounded Altruism: INGO Opportunities and Constraints during Humanitarian Crises and US Intervention in Bosnia-Herzegovina and Kosovo,” “Economic Development and Political Transition in Kosovo,” American University in Kosovo (October 2012).

## PROFESSIONAL SERVICE

Reviewer for *Political Analysis*

Chair and organizer for an International Studies Association panel, “Forecasting about the Future: Novel Forecasting Techniques and New Predictions” (February 2015).

Chair and organizer for an International Studies Association roundtable, “Assessing Forecasts of (Rare) International Events” (February 2015).

ANDREW L. HALTERMAN

Founding member of the Open Event Data Alliance, a non-profit organization of scholars and researchers engaged in work on generating, using, and evaluating event data automatically extracted from text data. Responsible for all updates to the coding dictionaries.

Co-organizer of the GDELT DC Hackathon, hosted by Gallup in Washington, DC (December 2013).

**FUNDING**

NSF Graduate Research Fellowship

MIT Department of Political Science and Political Methodology Group funding for research and travel

Minerva Initiative project W911NF-13-0332.

Key named personnel on NSF RIDIR grant: “Extending Automated Event Data Coding Across Language, Location, and Source.” Other team members/PIs consist of Patrick Brandt (project lead), Benjamin Bagozzi, John Freeman, Jennifer Holmes, Jill Irvine, Javier Osorio, and Philip Schrod. SBE-SMA-1539302

Conference travel funding from Caerus Associates for the European Network for Conflict Research Fall 2014 Meeting, Uppsala, Sweden (October 2014).

Conference travel funding from the American University of Kosovo for “Economic Development and Political Transition in Kosovo,” Prishtina, Kosovo (October 2012).