Proposal Title: Convergence Accelerator Phase I (RAISE): Northwestern Open Access to Court Records Initiative
Institution: Northwestern University
Abstract Date: 08/02/19

The NSF Convergence Accelerator supports team-based, multidisciplinary efforts that address challenges of national importance and show potential for deliverables in the near future.

The broader impact and potential benefit of this Convergence Accelerator Phase I project is to facilitate easier access to the full range of court records, which should enable more effective systematic research and promote greater analysis on how the federal courts operate and are utilized. The US litigation system is the primary mechanism through which our laws are formally enforced. Understanding how effectively the litigation system operates is critical to maintaining public trust, which is one reason why the courts maintain detailed records of all federal litigation.

Development of the proposed Northwestern Open Access to Court Records Initiative (NOACRI) open resource brings together legal scholars, criminologist, sociologists, computer scientists, statisticians, and complexity scholars to build a unique open knowledge network that will enable the convergence of a broad range of diverse communities--including legal scholars, social scientists, economists, journalists, and public policy stakeholders to more systematically study the federal legal system. NOACRI proposes to enable both analytically savvy and inexperienced users to interrogate the court data assembled. The project proposes to create unparalleled access to both raw case data and data annotations, including data annotated by the project team and data that are community-annotated. This more accessible data should also enable development of machine-learning and artificial intelligence (AI) tools to study court data systematically.

This project proposes to expand the substantial scholarly potential of court records by bringing recent methodological advances in big data analytics to bear on the field of legal research. To date, researchers have, by necessity, tended to focus primarily on the text of judicial opinions. This project will create an open and free resource of public litigation data, linked to supplementary publicly available data, that will dramatically advance the quantitative understanding of the workings of the federal court system. Importantly, it seeks to make visible and measurable data on cases that are settled or dismissed. Settled and dismissed cases may constitute well over half of federal court activity but is rarely published on court websites or otherwise available for free, limiting systematic study. This research will also aid in the development of data federation standards and natural language querying approaches that will benefit researchers in other subject areas that have a similar need to extract systematic insights from text.

This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.