

Suggested Projects: White House Innovation Fellows # 1 - Bring Content Analysis Online

I [suggest moving] US Government data and capabilities for analyzing communications into the public domain. This will create a cornucopia of new resources for American users - e.g., business opportunities for new software and Apps, faster understanding of changing global cultures and marketing opportunities, and insights into the domestic and global forces created by communications technology and social media.

The rigorous analysis of communications (Lasswell et al.) responded to mass communications media in the 1930s. Pioneers envisioned that “content analysis” tools could produce new measures that would become the equivalent, for all of the social sciences, of the new national income accounting for economics. My former MIT colleague, Ithiel de Sola Pool, edited a message-in-a-bottle volume from this generation in the late 1950s: The new mainframe technology was not ready for these bold applications, the methods were too expensive, they were limited to punch cards, human translators and coding of input, and memory/software that could do only simple frequency counts.

In 2012, the time is right. The intelligence community has developed resources that can be shared for civilian applications (www.opensource.gov) - e.g., large Reference Datasets for foreign languages and specific domains that allow sophisticated probability processing and “smart” search and translations - similar to what Google has achieved (Google News and its own search algorithms). We will need a civilian equivalent to SAS - a software analysis engine for text, visual, and other data - that entrepreneurs can link to the Reference Datasets and digital input. . . .

Lloyd S. Etheredge

June 5, 2012