
## The OrgPedia Open Organizational Data Project

Funded by the Alfred P. Sloan Foundation, the OrgPedia project is developing a free, not-for-profit online directory of data about domestic and international, public and private companies.

The OrgPedia prototype makes available and downloadable via the web a rich array of information about each firm, including:

- *"Business card" information* about name, location, and any entity identifier (ie. ID numbers used by different agencies as well as private numbering schemes like Open Symbology and, eventually, the FINRA financial services LEI)
- *Ownership information* about who owns and manages the firm and whom it owns and operates.
- *Other open government data*, including securities and patent filings, environmental and workplace safety records .

Using a linked data approach, OrgPedia builds the directory from three sources:

- Government databases of information collected from firms as part of transparent, regulatory processes.
- Private but open databases of information (ie. New York Times financial data).
- Public contributions to fill in gaps in the business card directory. Public contributions are clearly distinguished from authenticated government data.

Like Wikipedia, Orgpedia uses common collaboration features, such as tagging and rating, to provide a check on data quality by enabling companies to flag and correct inaccuracies and report mistakes back to government authorities. A developer's API enables the information to be easily exported to the websites and tools of other organizations, wishing to use this data for their own projects.

Unlike Wikipedia, OrgPedia is not a single repository of information but, rather, a clearinghouse with information linking back to myriad sources. OrgPedia, therefore, promotes, rather than competes with, business intelligence companies who offer analytics above and beyond OrgPedia’s goal of providing a directory of firms in the public interest.

OrgPedia builds upon parallel efforts within the financial services community to develop a global, open, legal entity identifier. We anticipate that a broader array of agencies from patent to environmental regulators will want to take advantage of the new universal LEI. Firms, too, will want to use the LEI to obtain information about those with whom they do business. OrgPedia provides the extensible infrastructure and public interest governance to offer a directory of more firms, both public and private; to link the new LEI to existing, legacy LEI systems already in use; and to show additional data vital to every regulator, such as location and address.

By providing an independent source for authenticated data with features to correct data quality problems and fill in gaps, OrgPedia has the potential to bring down the costs of compliance for both companies and regulators by giving them a reliable source of corporate information in a single place. The hope is also to empower consumers and businesses seeking better data about the corporations with whom they do business and in whom they invest. OrgPedia will also be useful to economists and journalists doing research. Researchers can build their own customized view of the information in the system using a standards-based query interface. OrgPedia also provides a source for data-driven policymaking by enabling regulators to compare data across different agencies and databases.

Currently a prototype with thousands of live records, OrgPedia aims to expand its scope through collaboration with government and business. As part of the project, we have also been exploring how to better serve users with particular privileges (ie. government users entitled to see EIN identifiers) or having access to paid identifiers (such as DUNS numbers) and developing a strategy (legal, technical, policy, financial) for ensuring future growth of the platform and the community that will steward it.

Principal Investigators:

James Hendler is the Tetherless World Professor of Computer and Cognitive Science, and the Assistant Dean for Information Technology and Web Science, at Rensselaer. He is also a faculty affiliate of the Experimental Multimedia Performing Arts Center (EMPAC), serves as a Director of the UK's charitable Web Science Trust and is a visiting Professor at the Institute of Creative Technology at DeMontfort University in Leicester, UK. Hendler has authored about 200 technical papers in the areas of Semantic Web, artificial intelligence, agent-based computing and high performance processing. One of the inventors of the "Semantic Web," Hendler was the recipient of a 1995 Fulbright Foundation Fellowship, is a member of the US Air Force Science Advisory Board, and is a Fellow of the American Association for Artificial Intelligence, the British Computer Society and the IEEE. He is also the former Chief Scientist of the Information Systems Office at the US Defense Advanced Research Projects Agency (DARPA) and was awarded a US Air Force Exceptional Civilian Service Medal in 2002. He is the Editor-in-Chief emeritus of IEEE Intelligent Systems and is the first computer scientist to serve on the Board of Reviewing Editors for Science. In 2010, Hendler was named one of the 20 most innovative professors in America by Playboy magazine and was selected as an "Internet Web Expert" by the US government.

Beth Simone Noveck served in the White House as the first United States Deputy Chief Technology Officer and founder and director of the White House Open Government Initiative (2009-2011), where she was responsible for developing and coordinating President Obama's Administration policy on transparency, participation, and collaboration. UK Prime Minister David Cameron recently appointed her Senior Advisor for Open Government. She served on the Obama-Biden Transition Team and was a volunteer advisor to the Obama for America campaign on issues of technology, innovation, and government reform. A law professor at New York Law School, Dr. Noveck is a leading expert on institutional innovation. She organized the recent Club de Madrid annual meeting, the convening of former Presidents and Prime Ministers. Among other projects, she designed and built the U.S. government's first expert network. With a new grant from the Aspen Foundation, she is developing a strategy for assessing costs and benefits of strategies for implementing data transparency. She is also designing a research network on the impact of new technology on democratic institutions for the MacArthur Foundation. A graduate of Harvard University and Yale Law School, she was named one of the "100 Most Creative People in Business" by Fast Company magazine, " Top 25 Game Changers" by Politico and one of the "Top Women in Technology" by Huffington Post. Her book about building an expert network for the United States Patent Office, Wiki Government (Brookings Institution Press 2009), appeared this year in Arabic and Chinese and in an audio edition and will also be translated into Russian. She is also co-editor of The State of Play: Law, Games and Virtual Worlds (NYU Press 2006). She will be speaking at TEDGlobal in June and her new book, The Networked State will appear in 2013.