



Opening government data sets enhances research and innovation

By Aaron Dunn

A research team led by the National Renewable Energy Laboratory recently used the capabilities developed by the US federal government's Materials Genome Initiative (MGI) to identify 27,000 compounds as candidate materials for nontoxic alternative solar panel materials using only computational tools. The team then narrowed this set of materials down to the best six candidates for synthesis, of which two showed significant promise as solar materials. Such a data-driven approach to materials design could dramatically accelerate the discovery of materials with advantageous properties.

Innovators are also increasingly looking to leverage open data sets. QuesTek Innovations has partnered with the MGI and is using open data in applications such as thermoelectric materials. The company uses integrated computational materials engineering to quickly design and deploy next-generation materials, including flight critical components on SpaceX rockets and in landing gear on US Navy and Air Force aircraft.

US federally sponsored research projects such as the MGI rely on a backbone of large data sets that are often partially or completely owned by federal agencies. Researchers have a strong interest in these data sets, including their accessibility, machine-readability, format, and continued growth as more data are added to them. Likewise, the government's treasure trove of open data sets that are available for use by researchers and businesses alike can be a key element of innovation and job creation.

The world of government data is dizzyingly large and diverse, so it helps to put the issue into some historical context. Federal agencies collect data on a wide variety of topics, some well known

and others behind the scenes. Examples of government data sets include federal spending, commodity prices, weather and climate, the census, disease rates and geographical distributions, and sunlight distribution. Until 2013, there was no unifying open data policy across government agencies. Although data archiving is required by the Federal Records Act, no standards for data format, open access, or inventory existed. In 2013, President Obama signed an executive order titled, "Making Open and Machine Readable the New Default for Government Information," which directed the creation of a government-wide open data policy "to advance the management of Government information as an asset." It also called for best practices for agency adoption of the policy and tracking its implementation. Agency implementation of the new policy has been uneven, and, currently, the newly created data.gov portal contains only select data sets from government agencies, but significant progress has been made.

Fortunately, US lawmakers from both political parties have recognized the importance of open data and are committed to ensuring maximum taxpayer access to government data of all types, while maintaining privacy and other legal constraints. Two bipartisan bills now being debated in Congress would establish government-wide standards for publication of data sets that are useful to the academic, business, and innovation communities. The OPEN Government Data Act would codify many of the provisions of the 2013 executive order and ensure that each government agency creates a data inventory. The Preserving Data in Government Act would ensure that data sets published in an open, machine-readable fashion remain permanently

available to the public. Together, these bills aim to maximize the positive impact of government data on research, innovation, and the economy.

Despite positive developments in fostering greater openness and machine-readability in government data, lack of awareness at the congressional level and inconsistent engagement at the academic level make backsliding a concern. Researchers who do not currently take advantage of big data-oriented tools such as the MGI should consider how they can benefit from them, and those in charge of big data initiatives should work to expand awareness of these capabilities. Scientific societies should work to develop data standards so that data sharing both within the scientific community and between the government and scientists becomes simpler, more automatic, and requires minimal effort on the part of researchers.

Perhaps most importantly, scientific communities that rely on open government data, including the materials research community, should consider contacting their members of Congress in the United States or their representative in their respective governments around the world about the value of open data, and share stories of how open government data has enabled research and helped innovation. In the United States, Congressional Visit Day, organized by universities and professional societies such as the Materials Research Society, presents a good opportunity to discuss these issues with members of Congress and their staff. Recent events such as the March for Science have shown an increasing willingness of the scientific community to engage in the policymaking process (see the July 2017 issue of *MRS Bulletin*, doi:10.1557/mrs.2017.152)—this engagement should include an emphasis on the value of open data.

Aaron Dunn served as the 2016–2017 MRS/TMS Congressional Fellow.