

# **Progress in U.S. Government Information Technology**

Compiled and Edited by Michael Erbschloe



© 2019 Michael Erbschloe

# Table of Contents

Section	Page Number
About the Editor	3
Big Data Initiatives	4
Quantum Computing	10
Blockchain Technology	20
Cloud Computing	29
The Internet of Things	38
Department of Defense Additive Manufacturing	49
Telemedicine and Telehealth	56
Radio Frequency Identification (RFID)	73
Facial Recognition Technology	79
Artificial intelligence (AI)	86
IT to Take Government to the People	94
Machine Learning	100
Open Source Initiatives	105
Cybersecurity Progress in the U.S. Government	117
The Federal Cybersecurity Reskilling Academy	133
U.S. Government Efforts to Manage E-Waste	135

# About the Editor

Michael Erbschloe has worked for over 30 years performing analysis of the economics of information technology, public policy relating to technology, and utilizing technology in reengineering organization processes. He has authored several books on social and management issues of information technology that were published by McGraw Hill and other major publishers. He has also taught at several universities and developed technology-related curriculum. His career has focused on several interrelated areas:

- Technology strategy, analysis, and forecasting
- Teaching and curriculum development
- Writing books and articles
- Publishing and editing
- Public policy analysis and program evaluation

Books by Michael Erbschloe

Extremist Propaganda in Social Media: A Threat to Homeland Security (CRC Press)

Threat Level Red: Cybersecurity Research Programs of the U.S. Government (CRC Press)

Social Media Warfare: Equal Weapons for All (CRC Press)

Walling Out the Insiders: Controlling Access to Improve Organizational Security (CRC Press)

Physical Security for IT (Elsevier Science)

Trojans, Worms, and Spyware (Butterworth-Heinemann)

Implementing Homeland Security in Enterprise IT (Digital Press)

Guide to Disaster Recovery (Course Technology)

Socially Responsible IT Management (Digital Press)

Information Warfare: How to Survive Cyber Attacks (McGraw Hill)

The Executive's Guide to Privacy Management (McGraw Hill)

Net Privacy: A Guide to Developing & Implementing an e-business Privacy Plan (McGraw Hill)

# Big Data Initiatives

Many companies are sponsoring Big Data-related competitions, and providing funding for university research. Universities are creating new courses and entire courses of study to prepare the next generation of data scientists. Organizations like Data without Borders have helped by providing pro bono data collection, analysis, and visualization. There have also been U.S. Federal Government programs that address the challenges of, and tap the opportunities afforded by, the big data revolution to advance agency missions and further scientific discovery and innovation. This paper presents a small sampling of the activities the government.

The Department of Defense (DOD) was investing \$250 million annually across Military Departments in a series of programs to harness and utilize massive data in new ways and bring together sensing, perception and decision support to make truly autonomous systems that can maneuver and make decisions on their own. This has included improving situational awareness to help warfighters and analysts provide increased support to operations. DOD has been seeking a 100-fold increase in the ability of analysts to extract information from texts in any language, and a similar increase in the number of objects, activities, and events that an analyst can observe.

The Defense Advanced Research Projects Agency has pursued the Anomaly Detection at Multiple Scales (ADAMS) program addresses the problem of anomaly-detection and characterization in massive data sets. In this context, anomalies in data are intended to cue collection of additional, actionable information in a wide variety of real- world contexts. The initial

ADAMS application domain is insider-threat detection, in which anomalous actions by an individual are detected against a background of routine network activity. The Cyber-Insider Threat (CINDER) program seeks to develop novel approaches to detect activities consistent with cyber espionage in military computer networks while the Insight program addresses key shortfalls in current intelligence, surveillance and reconnaissance systems and aims to develop a resource-management system to automatically identify threat networks and irregular warfare operations through the analysis of information from imaging and non-imaging sensors and other sources.

The DARPA Mission-oriented Resilient Clouds program aims to address security challenges inherent in cloud computing by developing technologies to detect, diagnose and respond to attacks, effectively building a “community health system” for the cloud. The program also aims to develop technologies to enable cloud applications and infrastructure to continue functioning while under attack. The loss of individual hosts and tasks within the cloud ensemble would be allowable as long as overall mission effectiveness was preserved.

The DARPA Video and Image Retrieval and Analysis Tool (VIRAT) program aims to develop a system to provide military imagery analysts with the capability to exploit the vast amount of overhead video content being collected. If successful, VIRAT will enable analysts to establish alerts for activities and events of interest as they occur. VIRAT also seeks to develop tools that would enable analysts to rapidly retrieve, with high precision and recall, video content from extremely large video libraries.

The DHS Center of Excellence on Visualization and Data Analytics (CVADA), has supported research efforts on large,

heterogeneous data that First Responders could use to address issues ranging from manmade or natural disasters to terrorist incidents; law enforcement to border security concerns; and explosives to cyber threats.

The DOE Office of Advanced Scientific Computing Research (ASCR) provides leadership to the data management, visualization and data analytics communities including digital preservation and community access. Programs within the suite include widely used data management technologies such as the Kepler scientific workflow system; and Storage Resource Management standard; a variety of data storage management technologies, such as BeSTman, the Bulk Data Mover and the Adaptable IO System (ADIOS); FastBit data indexing technology (used by Yahoo!); and two major scientific visualization tools, ParaView and VisIt.

Mathematics for Analysis of Petascale Data addresses the mathematical challenges of extracting insights from huge scientific datasets and finding key features and understanding the relationships between those features. Research areas include machine learning, real-time analysis of streaming data, stochastic nonlinear data-reduction techniques and scalable statistical analysis techniques applicable to a broad range of DOE applications including sensor data from the electric grid, cosmology and climate data.

The Office of Basic Energy Sciences (BES) Scientific User Facilities have supported a number of efforts aimed at assisting users with data management and analysis of big data, which can be as big as terabytes (10<sup>12</sup> bytes) of data per day from a single experiment. For example, the Accelerating Data Acquisition, Reduction and Analysis (ADARA) project addresses the data workflow needs of the Spallation Neutron Source (SNS) data

system to provide real-time analysis for experimental control; and the Coherent X-ray Imaging Data Bank has been created to maximize data availability and more efficient use of synchrotron light sources.

The Office of Fusion Energy Sciences (FES) has supported the Scientific Discovery through Advanced Computing (SciDAC) partnership between FES and the office of Advanced Scientific Computing Research (ASCR) addresses big data challenges associated with computational and experimental research in fusion energy science. The data management technologies developed by the ASCR – FES partnerships include high performance input/output systems, advanced scientific workflow and provenance frameworks, and visualization techniques addressing the unique fusion needs, which have attracted the attention of European integrated modeling efforts and ITER, an international nuclear fusion research and engineering project.

The Office of Scientific and Technical Information (OSTI), the only U.S. Federal agency member of DataCite (a global consortium of leading scientific and technical information organizations) plays a key role in shaping the policies and technical implementations of the practice of data citation, which enables efficient reuse and verification of data so that the impact of data may be tracked, and a scholarly structure that recognizes and rewards data producers may be established.

The Consortium for Healthcare Informatics Research (CHIR) has worked to develop Natural Language Processing (NLP) tools in order to unlock vast amounts of information that are currently stored in VA as text data. Meanwhile AViVA is the VA's next generation employment human resources system that will separate the database from the business applications and from the browser-based user interface. Analytical tools are

already being built upon this foundation for research and ultimately support of decisions at the patient encounter.

The Center for Disease Control & Prevention (CDC) has pursued BioSense 2.0 which was the first system to take into account the feasibility of regional and national coordination for public health situation awareness through an interoperable network of systems, built on existing state and local capabilities. BioSense 2.0 removes many of the costs associated with monolithic physical architecture, while still making the distributed aspects of the system transparent to end users, as well as making data accessible for appropriate analyses and reporting.

The Center for Medicare & Medicaid Services (CMS) has pursued a data warehouse based on Hadoop is being developed to support analytic and reporting requirements from Medicare and Medicaid programs. A major goal is to develop a supportable, sustainable, and scalable design that accommodates accumulated data at the Warehouse level. Also challenging is developing a solution complements existing technologies.

The FDA Virtual Laboratory Environment (VLE) was designed to combine existing resources and capabilities to enable a virtual laboratory data network, advanced analytical and statistical tools and capabilities, crowd sourcing of analytics to predict and promote public health, document management support, tele-presence capability to enable worldwide collaboration, and basically make any location a virtual laboratory with advanced capabilities in a matter of hours.

The National Archives & Records Administration (NARA) has worked to develop a Cyberinfrastructure for a Billion Electronic Records (CI-BER) through a joint agency sponsored testbed



notable for its application of a multi-agency sponsored cyber infrastructure and the National Archives' diverse 87+ million file collection of digital records and information now active at the Renaissance Computing Institute. This testbed will evaluate technologies and approaches to support sustainable access to ultra-large data collections.

NASA's Advanced Information Systems Technology (AIST) awards seek to reduce the risk and cost of evolving NASA information systems to support future Earth observation missions and to transform observations into Earth information as envisioned by NASA's Climate Centric Architecture. Some AIST programs seek to mature Big Data capabilities to reduce the risk, cost, size and development time of Earth Science Division space-based and ground-based information systems and increase the accessibility and utility of science data.

NASA's Earth Science Data and Information System (ESDIS) project, active for over 15 years has worked to process, archive, and distribute Earth science satellite data and data from airborne and field campaigns. With attention to user satisfaction, it strives to ensure that scientists and the public have access to data to enable the study of Earth from space to advance Earth system science to meet the challenges of climate and environmental change.

The Global Earth Observation System of Systems (GEOSS) is a collaborative, international effort to share and integrate Earth observation data. NASA has joined forces with the U.S. Environmental Protection Agency (EPA), National Oceanic and Atmospheric Administration (NOAA), other agencies and nations to integrate satellite and ground-based monitoring and modeling systems to evaluate environmental conditions and predict outcomes of events such as forest fires, population

growth and other developments that are natural and man-made. In the near-term, with academia, researchers will integrate a complex variety of air quality information to better understand and address the impact of air quality on the environment and human health.

The National Cancer Institute (NCI) Cancer Imaging Archive (TCIA) is an image data-sharing service that facilitates open science in the field of medical imaging. TCIA aims to improve the use of imaging in today's cancer research and practice by increasing the efficiency and reproducibility of imaging cancer detection and diagnosis, leveraging imaging to provide an objective assessment of therapeutic response, and ultimately enabling the development of imaging resources that will lead to improved clinical decision support.

National Institute of Biomedical Imaging and Bioengineering (NIBIB) has supported the Development and Launch of an Interoperable and Curated Nanomaterial Registry, led by the NIBIB institute, seeks to establish a nanomaterial registry, whose primary function is to provide consistent and curated information on the biological and environmental interactions of well-characterized nanomaterials, as well as links to associated publications, modeling tools, computational results and manufacturing guidance. The registry facilitates building standards and consistent information on manufacturing and characterizing nanomaterials, as well as their biological interactions.

NIH Biomedical Information Science and Technology Initiative (BISTI) Consortium for over a decade has joined the institutes and centers at NIH to promote the nation's research in Biomedical Informatics and Computational Biology (BICB), promoted a number of program announcements and funded

more than a billion dollars in research. In addition, the collaboration has promoted activities within NIH such as the adoption of modern data and software sharing practices so that the fruits of research are properly disseminated to the research community.

The Neuroscience Information Framework (NIF) is a dynamic inventory of Web-based neuroscience resources: data, materials, and tools accessible via any computer connected to the Internet. An initiative of the NIH Blueprint for Neuroscience Research, NIF advances neuroscience research by enabling discovery and access to public research data and tools worldwide through an open source, networked environment.

The NIH Human Connectome Project is an ambitious effort to map the neural pathways that underlie human brain function and to share data about the structural and functional connectivity of the human brain. The project will lead to major advances in our understanding of what makes us uniquely human and will set the stage for future studies of abnormal brain circuits in many neurological and psychiatric disorders.

The Worldwide Protein Data Bank (wwPDB), a repository for the collection, archiving and free distribution of high quality macromolecular structural data to the scientific community on a timely basis, represents the preeminent source of experimentally determined macromolecular structure information for research and teaching in biology, biological chemistry, and medicine. The U.S. component of the project (RCSB PDB) is jointly funded by five Institutes of NIH, DOE/BER and NSF, as well as participants in the UK and Japan. The single databank now contains experimental structures and related annotation for 80,000 macromolecular structures. The Web site receives 211,000 unique visitors per month from 140 different countries.

Around 1 terabyte of data are transferred each month from the website.

The Biomedical Informatics Research Network (BIRN), a national initiative to advance biomedical research through data sharing and collaboration, provides a user-driven, software-based framework for research teams to share significant quantities of data – rapidly, securely and privately – across geographic distance and/or incompatible computing systems, serving diverse research communities.

The National Archive of Computerized Data on Aging (NACDA) program advances research on aging by helping researchers to profit from the under-exploited potential of a broad range of datasets. NACD preserves and makes available the largest library of electronic data on aging in the United States.

The Collaborative Research in Computational Neuroscience (CRCNS) is a joint NIH-NSF program to support collaborative research projects between computational scientists and neuroscientists that will advance the understanding of nervous system structure and function, mechanisms underlying nervous system disorders and computational strategies used by the nervous system. In recent years, the German Federal Ministry of Education and Research has also joined the program and supported research in Germany.

Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA) is a new joint solicitation between NSF and NIH that aims to advance the core scientific and technological means of managing, analyzing, visualizing and extracting useful information from large, diverse, distributed and heterogeneous data sets. Specifically, it will

support the development and evaluation of technologies and tools for data collection and management, data analytics, and/or e-science collaborations, which will enable breakthrough discoveries and innovation in science, engineering, and medicine - laying the foundations for U.S. competitiveness for many decades to come.

Cyber Infrastructure Framework for 21st Century Science and Engineering (CIF21) develops, consolidates, coordinates, and leverages a set of advanced cyber infrastructure programs and efforts across NSF to create meaningful cyber infrastructure, as well as develop a level of integration and interoperability of data and tools to support science and education.

CIF21 Track for IGERT. NSF has shared with its community plans to establish a new CIF21 track as part of its Integrative Graduate Education and Research Traineeship (IGERT) program. This track aims to educate and support a new generation of researchers able to address fundamental Big Data challenges concerning core techniques and technologies, problems, and cyber infrastructure across disciplines.

Data Citation, which provides transparency and increased opportunities for the use and analysis of data sets, was encouraged in a dear colleague letter initiated by NSF's Geosciences directorate, demonstrating NSF's commitment to responsible stewardship and sustainability of data resulting from federally funded research.

Data and Software Preservation for Open Science (DASPOS) is a first attempt to establish a formal collaboration of physicists from experiments at the LHC and Fermilab/Tevatron with experts in digital curation, heterogeneous high-throughput storage systems, large-scale computing systems, and grid access

and infrastructure. The intent is to define and execute a compact set of well-defined, entrant-scale activities on which to base a large-scale, long-term program, as well as an index of commonality among various scientific disciplines.

Digging into Data Challenge addresses how big data changes the research landscape for the humanities and social sciences, in which new, computationally-based research methods are needed to search, analyze, and understand massive databases of materials such as digitized books and newspapers, and transactional data from web searches, sensors and cell phone records. Administered by the National Endowment for the Humanities, this Challenge is funded by multiple U.S. and international organizations.

The USGS John Wesley Powell Center for Analysis and Synthesis announced eight new research projects for transforming big data sets and big ideas about earth science theories into scientific discoveries. At the Center, scientists collaborate to perform state-of-the-art synthesis to leverage comprehensive, long-term data

# Quantum Computing

In July 2016, the National Science and Technology Council of the Executive Office of the President, in a report titled "Advancing Quantum Information Science: National Challenges and Opportunities", described Quantum Information Science (QIS) as a foundational science, with envisioned applications (that) include sensing and metrology, communications, simulation, and high-performance computing. The report also pointed out specifically that Quantum communication, the ability to transmit information encoded in quantum states of light or matter, is currently an active area of development. The report also states that in the longer term, quantum networks will connect distributed quantum sensors to allow long-distance transmission of quantum information.

Quantum information science combines two of the great scientific and technological revolutions of the 20th century: quantum mechanics on the one hand, and computer-based information science on the other. One of the fundamentally important research areas involved in quantum information science is quantum communications, which deals with the exchange of information encoded in quantum states of matter or quantum bits (known as qubits) between both nearby and distant quantum systems.

Quantum computing is based on quantum bits or qubits. Unlike traditional computers, in which bits must have a value of either zero or one, a qubit can represent a zero, a one, or both values simultaneously. Representing information in qubits allows the information to be processed in ways that have no equivalent in classical computing, taking advantage of phenomena such as quantum tunneling and quantum entanglement. As such, quantum computers may theoretically be able to solve certain problems in a few days that would take millions of years on a classical computer.

Qubits are the quantum analogue to the classical computer bits “0” and “1.” Engineering materials that can function as qubits is technically challenging. Using supercomputers, scientists from the University of Chicago and Argonne National Laboratory predicted possible new qubits built out of strained aluminum nitride. Moreover, the scientists showed that certain newly developed qubits in silicon carbide have unusually long lifetimes.

Quantum computers could break common cryptography techniques, search huge datasets, and simulate quantum systems in a fraction of the time it would take today’s computers. However, engineers first need to harness the properties of quantum bits. Engineering new qubits with less difficult methods could lower one of the significant barriers to scaling quantum computers from small prototypes into larger-scale technologies.

One of the leading methods for creating qubits involves exploiting specific structural atomic defects in diamonds. Using diamonds is both technically challenging and expensive. Now researchers from the University of Chicago and Argonne National Laboratory have suggested an analogous defect in aluminum nitride, which could reduce the difficulty and ultimate cost of manufacturing materials for quantum computing applications. Using the Edison and Mira supercomputers at DOE’s National Energy Research Scientific Computing Center and Argonne National Laboratory respectively, the researchers found that by applying strain to aluminum nitride, they can create structural defects in the material that may be harnessed as qubits similar to those seen in diamonds. They performed their calculations using different levels of theory and the Quantum Espresso and WEST codes, the latter developed at the University of Chicago. The codes allowed them to accurately predict the position of the defect levels in the band-gap of semiconductors. The researchers also closely collaborated with experimentalists to understand and improve the performance of qubits in industrial materials. Recently, they showed that newly developed qubits in silicon carbide have much longer coherence



times than that of the more well-established defect qubits in diamond. Their results pointed to industrially important polyatomic crystals as promising hosts for coherent qubits for scalable quantum devices.

Peter Shor's 1994 breakthrough discovery of a polynomial time quantum algorithm for integer factorization sparked great interest in discovering additional quantum algorithms and developing hardware on which to run them. The subsequent research efforts yielded quantum algorithms offering speedups for widely varying problems, and several promising hardware platforms for quantum computation. These platforms include analog systems (usually cold atoms) used for simulating quantum lattice models from condensed-matter and high-energy physics, quantum annealers for combinatorial optimization, boson samplers, and small-scale noisy prototypes of digital gate-model quantum computers.

In the longer term, the emergence of scalable, fault-tolerant, digital quantum computers offers a new direction for progress in high performance computing as conventional technologies reach their fundamental limitations. Quantum speedups have been discovered for a number of areas of DOE interest, including simulations for chemistry, nuclear and particle physics, and materials science, as well as data analysis and machine learning. In addition, quantum speedups have been discovered for basic primitives of applied mathematics such as linear algebra, integration, optimization, and graph theory. These demonstrate the potential of quantum computers to yield better-scaling methods (in some cases exponentially better) for performing a wide variety of scientific computing tasks. Practical realization of this potential will depend not only on advances in quantum computing hardware but also advances in optimizing languages and compilers to translate these abstract algorithms into concrete sequences of realizable quantum gates, and simulators to test and verify these sequences. The development of such software has recently seen rapid progress, which can be expected to continue given sufficient support.

## Thank You for previewing this eBook

You can read the full version of this eBook in different formats:

- HTML (Free /Available to everyone)
- PDF / TXT (Available to V.I.P. members. Free Standard members can access up to 5 PDF/TXT eBooks per month each month)
- Epub & Mobipocket (Exclusive to V.I.P. members)

To download this full book, simply select the format you desire below

