# An Ocean of Data Requires Trans-disciplinary Education

*The typical person on earth generates about 1-2 Exabyte (1,000,000,000,000,000,000-bytes) per year. How much of this data is used by humans ever day? Not much. The vast majority of it is produced and stored (temporarily) in digital chips then reset (discarded) the second (or nanosecond) the equipment is needed to represent another piece of information (i.e. data in context, such as a yes/no answer on survey, amount of money in a credit transaction, or the amount of electricity consumed when turning on the kitchen light). In short, we are awash in a sea of bits. The new danger to civilization may come from this "deluge of data." 1 Unless science, business, and policymakers can gain control over it, the productivity gains we hope to achieve may drown in analysis paralysis.*

AUGUST 2009

NYU SCPS
School of Continuing & Professional Studies
At the center of the center of it all.

**Dorothy A. Durkin**
ASSOCIATE DEAN

A simple 'yes/no' decision by a human being can be represented in digital information terms as a 1 or a 0 that is - a "bit." A "byte" is 8-bits. A typical text word on a web page requires about 10-bits. The typical person on earth generates about 1-2 exabytes (1,000,000,000,000,000,000-bytes) per year.

How much of this data is used by humans ever day? Not much. The vast majority of it is produced and stored (temporarily) in digital chips then reset (discarded) the second (or nanosecond) the equipment is needed to represent another piece of information (i.e. data in context, such as a yes/no answer on survey, amount of money in a credit transaction, or the amount of electricity consumed when turning on the kitchen light). In short, we are awash in a sea of bits.

The new danger to civilization may come from this "deluge of data." 1 Unless science, business, and policymakers can gain control over it, the productivity gains we hope to achieve may drown in analysis paralysis.

Professional education plays an important role in saving us from this flood, but it will need to become adept at developing coherent programs across multiple disciplines. 2 'Data generating' technology is already pervasive. It is embedded in just about every device people use during the normal course of a day. To be useful in any given field the data has to be organized, collected and structured in the context of the applications designed for specific activities. Data generated by an fMRI system looks just like data generated by a voting machine until it is structured and expressed in arrangements that are meaningful to the users.

Data analytics3 is arguably the best overall header for describing the most general class of needs, but unless the analysis occurs in the context of a specific problem, it does not tell us very much. The visual and communication arts, psychology, human cognition, semiotics, anthropology, and linguistics all play a crucial role in helping humans interact with data in meaningful and purposeful ways.

The term "informatics" has become a useful way to bundle all the dimensions of the problem. For the sake of simplicity, consider terms such

[1] Andersen, Chris. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." **Wired Magazine, http://www.wired.com/science/discoveries/magazine/16-07/pb_theory**

[2] Referred to in some cases as "trans-disciplinary," to connote the idea that the knowledge required is unique to a given practice in an application area, like the data generated by elections, but used in the context of *reporting* election returns on television, rather than in the context of designing and engineering voting machines.

[3] Data analytics is also the most *commonly used term* found in the literature, but informatics seems to capture more of the dynamic scope of the actions involved.

as information architecture, data analysis, visualization, simulation, and virtualization to be special cases of informatics.

**The Multi-Disciplinary Imperative**

Stanford University provides the best description: "Informatics is the study of information, its structure, its communication, and its use. As society becomes increasingly information intensive, the needs to understand, create, and apply new methods for modeling, managing, and acquiring information has never been greater…"

In this definition, the hardcore techies create the tools that gather and contain the data with the understanding that IT tools and techniques are not sufficient without contextual understanding of the phenomena that generates the data or defines the interpretive setting.

For example, data modeling applies to practically any area of study from molecular to societal to galactic levels. Applied expertise is only useful in its specific arena, however. The data used for economic and atmospheric modeling share no commonalities that would allow people dealing with decisions specific to those fields to make sense of the same data in the same way. While weather has an economic impact, and economic decisions affect climate change, neither the weather forecaster nor the investment banker can make head or tails of the other's data models.

Ever-increasing specialization characterized new job creation throughout the Industrial Age, on into the early stages of the Information Age. The Digital Age is putting a new spin on the process. Instead of "knowing more and more about less and less," the new Noah (the people who will save us from this deluge) will have to be broadly conversant with the application fields they work in: healthcare, education, public policy, banking, real estate, marketing, media, and so on.

The preeminent role of the "geek" and "propeller head" in the IT landscape is receding. They are being outsourced and relegated to anonymous roles in massive networking operations and data centers that constitute, The Cloud. The emerging demand is for people who can use the data to help make sense of what is happening in education, healthcare, public administration, business, construction, real estate, media, tourism and every other type of organization that makes decisions in dynamic environments.

## How Some Institutions Fill in the Information Landscape

The study of informatics at NYU is primarily restricted to programs at the School of Nursing and the School of Medicine at NYU Langone Medical Center and in an MS Bioinformatics program at NYU Poly. Other prominent institutions offering undergraduate and graduate programs in informatics include UCLA, Indiana University, Columbia, Yale University and Stanford University. (Any internet search on the "informatics" will produce scores of returns, though the vast majority is part of, or closely linked to, medical practice, biomedical applications, nursing, and so on.) One proprietary enterprise, The Informatics Group, claims collaboration with some 50 colleges and universities around the world.[4]

Most of the informatics curricula at these schools draw heavily on medical, pharmacology, nursing, and epidemiology. European institutions such as University of Sussex[5] and the School of Informatics at University of Edinburgh offer more varied curricula with opportunities for students to focus on other areas of applications, including International Development and Computer Science.

Indiana University's School of Informatics offers both bachelors and masters degrees. The bachelors program augments its IT core with "tracks" in biology, business, chemistry, cognitive science, communications and culture, economics, fine arts, geography, linguistics, mathematics, psychology, public and environmental affairs, and telecommunications. This is one of the more forward looking programs uncovered in the course of this research. The bulk of its masters program aligns with biomedical, medicine, and healthcare sectors, however it also offers two intriguing tracks of informatics study, Music Informatics and Security Informatics.

University of Michigan also uses the "tracks" approach for its informatics programs, but follows a more conservative course.

- Computational Information Track: design, manipulate, assess and evaluate usable computing solutions

[4] Based in the Philippines, the company is listed on the Singapore Stock Exchange, and claims to work with some 50 colleges and universities around the world offering corporate training and professional certification, in location-based and online programs. The bulk of the courses appear to be "vendor" specific training in off-the-shelf software and architectures.

[5] Offers an internationally renowned center for research in Computer Science, including Artificial Intelligence, Cognitive Science, Evolutionary Electronics, Human Computer Interaction, Interactive Learning Environments, Multimedia, Neural Computation, Robotics, Software Engineering and Virtual Environments.

- Information Analysis Track: collect, manage, analyze, and visualize massive datasets
- Life Science Informatics Track: apply computation and statistics to problems in life science and bio-medical research
- Social Computing Track: build, evaluate, and refine new and emerging social software applications

Michigan provides diverse curricula at both the undergrad and graduate levels but its integration at the employment and career level heels closely to the public health, health sciences, and healthcare sectors. Michigan supports those fields with a variety of continuing education and online professional programs for informatics.

The reason for this strong association is readily apparent in the employment landscape and job picture.[6]

- **Hospitals and Other Healthcare Providers**: Areas where improved data analytics and informatics design are crucial.
    - Patient medical records
    - Diagnosis
    - Evidence-based management for infection prevention and disease control
    - Hazardous waste management
    - Robotics design
    - Facilities and energy management
    - Collaborative and consultative practices

- **Pharmaceutical Companies**:
    - Health informatics professionals aid in the gathering and analysis of the information from trials
    - Informatics plays a crucial role in integrating data generated by prescription patterns and treatment outcomes providing a much more complete perspective on drug interactions and use
    - Modeling software aids science and engineering processes
    - Marketing data becomes more accurate and informative with better quality information for physicians, pharmacists, and individual patients

- **Medical Software Companies**:
    - Back office management process (billing, scheduling, insurance, etc.) can help lower medical costs

[6] See sample job listing in the appendix.

    - Front office and patient wellness systems help control costs
    - Work with consultants and experts in business processes with skills in HR, finance, accounting, CRM, and related disciplines
    - Multiple database integrations

- **Consulting Companies**: The medical sector is a major client for business and management consulting companies. Areas of focus are diverse, from HR practices, to waste management.
    - Consultants familiar with healthcare environments usually advise and assist healthcare organizations with development and implementation of information systems that provide knowledge management solutions.
    - Consultants play a role in vendor selection, assessment, and procurement offering expert counsel to hospitals and medical practices unfamiliar with the standards and practices common to other businesses.

- **Public Health Organizations**: Information on populations and communities is collected and analyzed. The design and implementation of surveillance systems, and disease reporting systems, are carried out by public health organizations.
    - Infectious disease prevention & control
    - Evidence-based treatment
    - Work with consultants and experts in business processes with skills in HR, finance, accounting, CRM, and related disciplines
    - Multiple database integrations

- **Government and Non-governmental Organizations and Agencies**: These agencies carry out health planning and analysis of information at several levels of government and collaborate with other international bodies and organizations. In the process, tons of information is analyzed, stores, and made readily available for retrieval.
    - Healthcare delivery is a major focus of international aid and development. NGOs play an important role both in delivery and in patient education.
    - Social Performance Assessment tracks actions to outcomes and costs.

- **Insurance Companies**: Work here involves the analysis of health insurance claims and health records
    - Regulation and oversight of insurance practices is set for a major overhaul in the United States.
    - Information systems integration will be crucial for obtaining cost reductions and control.
- **Medical Education and Research**: Schools of medicine are focal points for leading edge medical technology and training. They work closely with pharmaceutical and medical engineering companies moving research to market in the form of new products and services.
    - Vast amounts of data travel between research medical centers, private sector companies, and regulatory agencies, each with their own perspective on the data and role to play in assuring safety and effectiveness.

The development of informatics for use in various aspects of healthcare is likely to remain near the top of the national agenda for years to come. The government and the public place a lot of hope on informatics to bring new efficiencies that will save billions of healthcare dollars (public and private) while improving the delivery and quality of the service.

The main action will focus on the makeup and distribution of patient health records, including diagnostics and tests results. Data security and encryption will take on a completely new role as medical records become the key component for sense-making at every level of the healthcare system.

The informatics quest will be able to satisfy all of these requirements from a common data infrastructure that is at once transparent and protects patient privacy. Not only must these advances take place in the near future, but also they must provide substantial cost savings and improve the quality of care. For informatics, this is the Holy Grail.

Healthcare and biomedical research provide the beachhead for the development of “informatics,” though they are not the high-water mark. Just as defense spending fueled much of American industrial and scientific development after WWII and during the Cold War, national healthcare may provide the energy that enables every other industry sector to grapple with the flood of data produced by its daily interactions. Most of these enterprises over the next several decades will devote massive new resources to mining the gold in the data.

## The 'Green' Glass is half full

The other major economic driver will be the "green sector," which includes renewable energy, conservation, transportation, environmental management, housing, real-estate, construction, manufacturing, and just about every other enterprise, whether public or private. There will be new investment, new employment, and sustainable industries. Regulatory policy will make it increasingly difficult to distinguish between public and private goods. The devil is in the details, and details will be in the data.

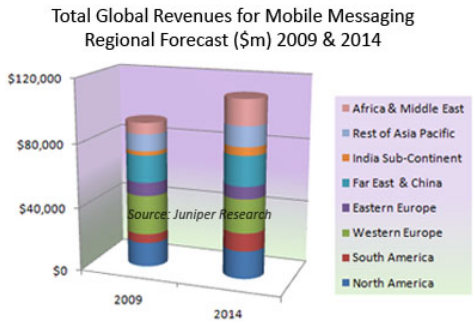
Interlocking 'Green' Components

- Energy extraction and renewable development
- Energy development, generation and management, distribution
- Manufacturing (all sectors)
- Transportation
- Environmental Management
- Construction
- Small business (retail, installation, maintenance)
- Regulation
- Finance
- Business operations

'Green' is more than energy. It embodies the new economics of sustainability that looks not just at the cost of products and services in terms of money, but in terms of impact on social performance. When do we stop tracking the cost of economic decisions such as selecting energy sources for development, deciding on packaging for products, or determining land use policies, housing policies, infrastructures subsidies, and defense requirements? The choices people make for themselves and their communities have effects far beyond their personal boundaries.

## Mobile & Marketing

Even our pets have digital implants these days to help us identify and retrieve them if they get lost. Combined with radio tracking, satellite imaging, and GPS mapping systems, naturalists use similar technology to monitor animal migratory patterns, population levels, and assess the health of small niches and large ecosystems. Now that much of the world's population is equipping itself with mobile phones (with enabled GPS), studying human social behavior will be much easier and more profitable.

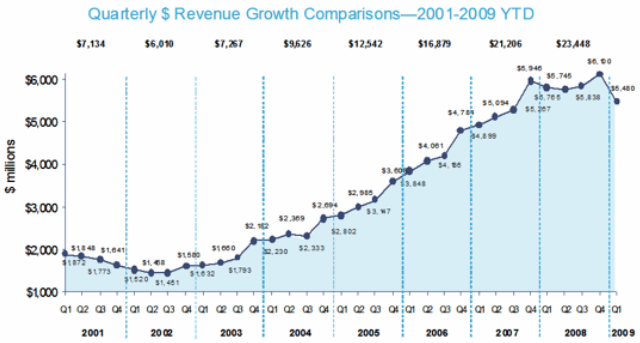



## Tracking and Targeting

Similar technology is deployed in private automobiles, trucks, boats, and shipping containers. GPS and satellite communications track cargo movements - moment to moment. Informatics turn the data into interactive

displays (such as CNN's 'Magic Wall' tracks election returns), calculate route alternatives (avoid left turns to save time, money, fuel, and reduce accidents and insurance costs), warn workers of unexpected changes to conditions, and suggest route changes and alternative distributions patterns to save time and money. Long haul shippers of commodities and perishable products use integrated GPS and information from commodity exchange databases to alter the destination of ships at sea to take advantage of price fluctuations in different markets.

*The sudden drop in 2009 quarterly online revenue sends shockwave through media world.*




Online advertising and media have fueled the growth of the Web, while conventional mass media has suffered. Newspapers and magazines have been particularly hard hit by this growth. *Wired Magazine* commentator, Chris Anderson, has lifted spirits, however, claiming the public's response to web apps (the small software objects downloaded to smartphones enabling users to interact with content in specific ways) proves that consumers *will* pay for content that "does something" for them personally. He bases his argument on the *Wall Street Journal's* success with charging for content, and Apple's recent sale of its one-billionth iPhone app. Even at one dollar per app, that is a lot of money.

With behavioral targeting and the use of "dynamic content" engines (i.e., content management systems with a "behavioral tracking layer") publishers will not only target content to specific users, but provide the target with a reader that uses the content in a personally customized way. How would you like to interact with your favorite reading material? Bet on sports favorites. Identify designers of the fashions worn by celebrities in photographs. Compose shopping lists from recipe's you read. Shop online for travel or accommodations at a resort or destination you read about in books and articles just by pointing at it. If you subscribe, the 'vendor' might receive the personal profile you provide which includes your preferences, preferred payment method, and a secure identity key that guarantees the safety of your information as you specify.

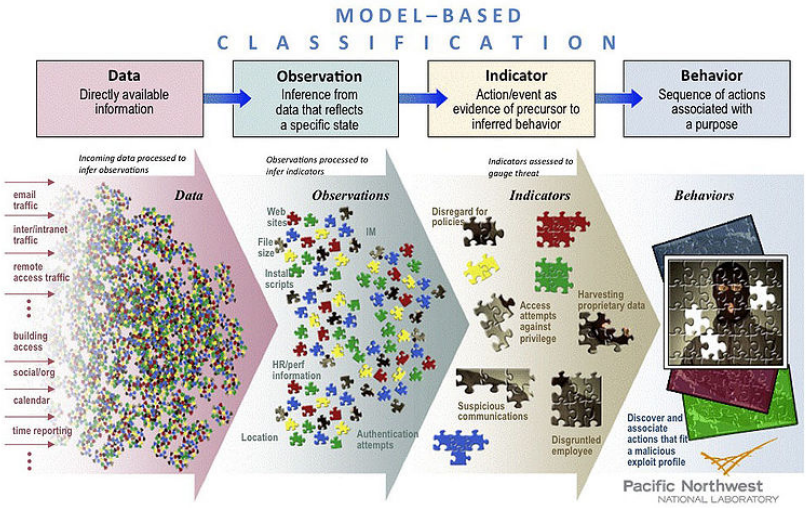
The energy of the Blogosphere, Apple Reader Apps, Amazon's Kindle, Sony's Reader Digital Book, Twitter, and its flock of clones flies in the face of those who think the publishing business will perish. The paper trail may have ended, and publishing may be down — but it is not out, not by a long shot.

So-called "dynamic content engines" already reformat content and images to fit the great variety of display formats that are used in mobile phones, PDAs, Smartphones, iPods, and other personal devices. Informatics specialists will also conceive and design the interactive environments that will appeal to users.

**Behavioral Targeting and 'The Big Sleep'**

Already the typical Internet user is generating about 47 gigabytes of data a day depending on how much time he/she spends online, on the phone (mobile or landline), using a Blackberry, credits cards, EZ-Pass, ATMs and so on. That data in the hands of a properly schooled and trained informaticist tells a great deal about your interactions, preferences, habits, purchasing behavior, personal network, and so on. It is called your 'digital footprint' (here is a link where you can download and calculate your own digital footprint).

*Behaviorists can model online activity to gauge the user's intent. That is useful for protective and security agencies, and invaluable to marketers.*




Behavioral Targeting (BT) is already the topic of public controversy. Various government regulatory bodies including the Federal Trade Commission, the FCC, to Homeland Security and sundry law enforcement agencies have designs on implementing and/or controlling the use of informatics. Consumers, civil liberties activists, and ordinary citizens who

are deeply concerned about losing the last vestiges of anything that feels like privacy are hard at work trying to severely restrict its use, or outlaw it all together. The Center for Democracy and Technology, Consumer Action, Consumer Federation of America, Electronic Frontier Foundation, Privacy Activism, Public Information Research, Privacy Journal, Privacy Rights Clearinghouse, and World Privacy Forum asked the Federal Trade Commission to provide needed consumer protections in the behavioral advertising sector. The groups have asked the FTC to implement a "Do Not Track List" intended to protect consumers from having their online activities unknowingly tracked, stored, and used by marketers and advertising networks."7 Given the density of media, legal, and marketing communities in the New York Metro area, BT should provide sufficient heat to fuel a number of professional conferences.

**Current Occupation Outlook**

The employment and career advancement outlook for people with degrees and certifications in informatics (or its related fields) is very robust. A search of job posting boards produced thousands of returns (see samples in appendix).

Common job and occupation titles include:

| | |
|---|---|
| Bioinformaticist | Software developer |
| Data mapper | Health informaticist |
| Webmaster | Information architect |
| IT consultant | Technical writer |
| Multimedia specialist | Systems administrator |
| Chemical informaticist | Usability analyst |
| Database designer | User experience analyst |
| Digital artist | Interaction designer |
| HCI interface designer | User experience analyst |
| Network manager | |

As is customary in new fields, they begin with a base set of titles borrowed from other occupations and rapidly evolve their own vernacular (as "web designer" was to Internet technology). The list above was gleaned from a variety of university career sites and online job boards.

[7] Center for Democracy & Technology. http://www.cdt.org/press/20071031press.php

## Salary Range and Benefits for Informaticist

According to the career office of the School of Informatics at Indiana University, "An Informaticist with a bachelor's degree might start at around $40,000 to $50,000 per year."[8] Besides relatively high salaries, benefits are also usually very good because of the need to lure people to take IT jobs. What may make the field especially attractive to adult learners is the value of "specific experience" in an applications area. Whether it's marketing, media, engineering, design, or another discipline, knowledge of the context in which the data is to be used is crucial."

The Career Builder search that produced the set of finding in the appendix was filtered for jobs postings "requiring a 4 year college degree" and providing salaries ranging from "$50K to $120,000." Postings were arbitrarily selected to reflect a range of positions from entry level to upper management in fields from biomedicine to interactive design.

## Qualities and Qualifications Present in an Informatics Graduate

The following is excerpted from the University of Sussex' website, explaining the qualities and properties expected from a graduate of its informatics programs.

- Developed technical, numerate and analytical skills and the ability to think creatively, logically and quantitatively
- Knowledge of software design, distributed systems, multimedia systems, networks and the internet – and of the technical and social challenges that they pose
- Developed abilities in testing, documentation and evaluation of systems and the ability to recognize the capabilities, limitations and risks of computer-based solutions
- Experience examining the design of intelligent computer systems and an understanding of intelligence in both machines and humans

## Viable program Structures

The 'application track' strategy, already in place at a number of universities, makes a good deal of sense given the current state of the art. NYU SCPS with its emphasis on specific professional sectors can provide the "applications-specific" expertise needed. The broad range of the schools reaches into communities of adult learners already holding

[8] University of Indiana

bachelors or advanced degrees, suggesting that an “additional” skill and credential in informatics (or data analytics) could add value for the early, transitioning, or mid-career professional.

The demand for human interaction with data through some visual or kinetic mode suggests using tracks directed at designers, animators, graphics artists, musicians, and other ‘creatives’ who want to add specialized job appeal to their creative gifts.

Of course, the inverse is equally appropriate. Long term Department of Labor projections for computer programming jobs show a four percent decline in the New York area by 2016.9 The decline in demand for Audio-Video Collection Specialist is projected at 21 percent for the same period suggesting an impending career area migration. In contrast to the instabilities evident in specific occupational categories, the overall trends in the data design and analytical occupations show increasing demands, ranging from six percent to forty-two percent for Networks Systems and Data Communications Analysts in the NY metro. [10]

Similar arguments can be made for Data Analytics or Informatics as a ‘first professional degree’ possibly leading to accelerate programs that combine a bachelor’s degree with an MPS. A stand-alone MS degree in informatics would be reasonable once the field or its nomenclature gains wider acceptance in fields beyond biomedical applications. In the meantime, graduate certificates based on well-designed “informatics” or “data analytics” core could provide the sort of hyphenated certification that adds value to an existing master’s degree.

In any event, much of the hope for economic growth over the next decades rides on the utilizations of data to create value and innovation. Significant investments in a new national system of healthcare, sustainable energy resources and a clean, healthier global environment will attract and capture the imagination and attention of the world’s human talent. Professional education has an opportunity to shape this new wave.

[9] US Department of Labor http://www.acinet.org/acinet/oview4.asp?next=oview4&Level=BAplus&optstatus=&jobfam=&id=1%2C&nodeid=6&soccode=&stfips=00&ShowAll=

[10] The DOL Career One-Stop site, displays trend results so that a specific occupation in a specific metropolitan (or larger) area can be compared to broader, national frameworks. In this case the NYC-Westchester-Wayne, NJ ‘metro’ generally lags in percent of employment growth against national averages. Salary levels in the metro, however, tend to be substantially higher suggesting that New York’s role as a magnet for talent will persist.