eScience for the eDisciplines Development and eCollaboration

Dr. Jože Gričar, Professor Emeritus, Joze.Gricar@UM.si
Dr. Adolf Šostar, Professor Emeritus, Adolf.Sostar@UM.si

February 17, 2021

The role of the professors emeriti in interdisciplinarity

Professors emeriti can play a special and important role in promoting and implementing the interdisciplinary approach in research and development. First, they derive from very different scientific disciplines, the cooperation of which is essential in solving a particular problem. Second, they are relaxed (liberated) by attachments to their original discipline. Third, they have no restrictions of the home university in making contacts and collaborating with colleagues in the network they create. An example of the role of the professors emeriti is expressed by the Emeritus College, University of British Columbia: interdisciplinarity and community outreach (The Academic Role of the College).

Interdisciplinary eCollaboration

In December 2020, the Interdisciplinary eCollaboration Think Tank was established. The purpose of this network is to promote learning from others/ sharing the ideas with others in regard to problems identified and solutions possible.

The members are actively involved in several actions in their respective environments, where they are learning that various disciplines have to collaborate more closely by exploiting eTechnologies. In numerous areas, for example, climate change, aging society, pandemic, healthy food procurement, human rights and defence of democracy, and related domains, an accelerated usage of the Internet is needed within all disciplines, and between the disciplines – eDisciplines. In particular, with the overwhelming penetration of artificial intelligence, big data and robotics, it is necessary to question how these technologies may interact with human compassion and empathy, stimulating the role of social sciences and humanities in conjunction with ICT.

Information Systems and Information Technology professionals have an important role to play in helping bridge disciplines to enable efficient and effective collaboration in addition to highlighting technology application opportunities, threats, and benefits.

eScience

In the Netherlands, the eScience Center was established in 2011 as the national expertise center for the development and application of research software (eScience development and experiences in The Netherlands).

Collaborating with researchers from all academic disciplines, the center extends the breadth and depth of research opportunities by exploiting the latest insights from

computer and data science, as well as making optimal use of hardware, software, and data infrastructures.

It does so through problem-driven research projects where eScience research engineers, employed by the eScience Center, collaborate with researchers in all disciplines at Dutch academic institutions.

Project software is generalized and made available for reuse for other disciplines and goals. The center has three main technological competences: efficient computing, optimized data handling, and data analytics.

Furthermore, on the national level it coordinates and contributes to science policies on computing, data, and applications thereof. With its two main assets, a staff of highly educated and multi-disciplinary eScience Research Engineers and an open online directory of research software tools and knowledge, it successfully contributes to the Dutch scientific landscape and enhances and accelerates all research in The Netherlands and beyond.

Questions:

Which problems of your personal interest and engagement are the best candidates for our actions in 2021?
How can the professors emeriti accelerate their involvement in interdisciplinary eCollaboration?
How can we gain the synergy of the eScience use in Central Europe?

E-Science is the basis around which science is working to achieve new discoveries and advancement to improve areas ranging from medicine to dentistry. As the technology scientists use to conduct their research becomes more technological and powerful, so must the ability to store the corresponding data. E-Science is the tool that offers scientists a scope to store interpret, analyse and network their data to other work groups. E-Science will play a major role in every facet of scientific research, starting out with the initial theory-based research, testing through simulation, systematically controlled testing, collection of data in an organised manner and the corresponding interpretation of said data.

E-Science is computationally intensive science that is carried out in highly distributed network environments, or science that uses immense data sets that require grid computing; the term sometimes includes technologies that enable distributed collaboration, such as the Access Grid.

E-Science has been more broadly interpreted since then, as the application of computer technology to the undertaking of modern scientific investigation, including the preparation, experimentation, data collection, result dissemination, and long-term storage and accessibility of all materials generated through the scientific process. E-Science revolutionizes both fundamental legs of the scientific method: empirical research through digital big data and the scientific theory, especially through computer simulation model building.

E-Science is not something that mere individuals can work to accomplish. As computer software is so complex and the data needs of scientific research is so great, you will find that is always large teams, consisting of multi-disciplined personnel, who work on e-Science projects.

These are typically groups made up of well-founded universities, government bodies and research laboratories. E-Science is essential to any advances that are to be made in science. It is also a fact that scientific requires professionals from varying cross-disciplines. It is the case that such research typically requires the input knowledge and scientific skills of scientists across the globe. This makes the ability of e-Science to related large amounts of data and information across the globe of immense importance.

Innovation and development

If science is to develop and new discoveries are to be made, then e-Science will be the one to stimulated such innovation. There is still much that we don't know. For example, there are still diseases that have no known cure, mysterious anomalies that we are unable to explain and frontiers that have not yet been crossed.

If such questions and irregularities are to be answered, then the innovative techniques of e-Science will be essential. E-Science provides scientists and researchers with the architecture through which they can work to discover the unknown. Without the storage capacity and ease with which data can be communicated across the globe, science would end up standing still.

Turing award winner Jim Gray imagined "data intensive science" or e-Sciences as a "fourth paradigm of science /empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing be case of the impact of information technology" and the data deluge.

To support e-Science applications, Open Science Grid combines interfaces to more than 100 nationwide clusters, 50 interfaces to geographically distributed storage caches and 8 campus grids. Areas of science benefiting from Open Science Grid includes:

- astrophysics, gravitational physic, high-energy physics, neutrino physics, nuclear physics,
- molecular dynamics, material science, material engineering, computer science, computer engineering, nanotechnology,
- structural biology, computational biology, genomics, proteomics, medicine.

The term e-Science describes computationally intensive science or science requiring new computational tools to manage massive amounts of heterogeneous, distributed data that must be efficiently stored, processed, analysed and visualized. These computational methods and tools allow and accelerate the production of new knowledge in many areas of science.

Examples of e-Science research:

- Atomistic calculations in chemistry and physics that produce information about the physics-chemical properties of substances.

- Analysis of the enormous data volumes produced in particle physics experiments, as for example in the search for the Higgs particle at CERN.
- Bioinformatics.
- Pattern recognition and machine learning.
- Advanced automatic text translation.
- etc.

Many knowledge-intensive industries with R&D are utilizing some of these methods and tools. For example, in the development of pharmaceuticals, biotechnological products and services, new materials, design of cars and airplanes, among many others.

Today we are surrounded by a myriad of wireless sensor networks and mobile devices that are constantly sensing, capturing and transmitting all sorts of data about the physical worlds, its objects and inhabitants. Many sectors of business, industry and public service are utilizing similar methods and tools to manage the massive ubiquitous production of data in our modern societies.
Examples of non-scientific applications of e-Science methods and tools:

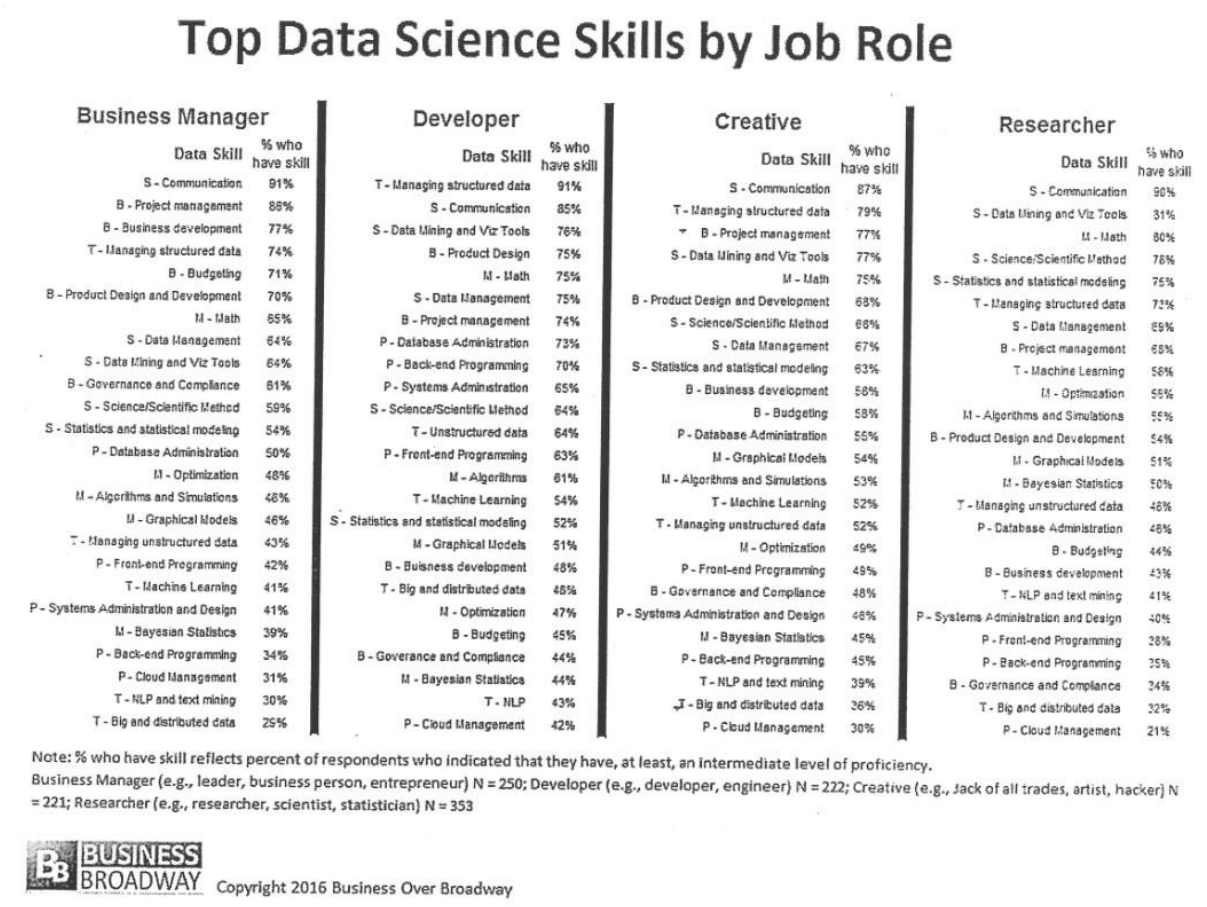
- Weather forecasting.
- Social media e. g. Facebook and Twitter.
- Search engines e. g. Google.
- Mobile communications.
- Computer game development.
- etc.

The list of top data science skills depends on the type of data scientists you are considering.

Developers top skills include Programming skills, Researchers top skills include Math-heavy skills.

Top data science skills for Researchers are primary in the area of Statistics. Additionally top data.

Science skills that are unique to Researchers are highly quantitative in nature, including Machine Learning and Optimization.



Top Data Science Skills by Job Role

Business Manager

Data Skill	% who have skill
S - Communication	91%
B - Project management	86%
B - Business development	77%
T - Managing structured data	74%
B - Budgeting	71%
B - Product Design and Development	70%
M - Math	65%
S - Data Management	64%
S - Data Mining and Viz Tools	64%
B - Governance and Compliance	61%
S - Science/Scientific Method	59%
S - Statistics and statistical modeling	54%
P - Database Administration	50%
M - Optimization	48%
M - Algorithms and Simulations	48%
M - Graphical Models	46%
T - Managing unstructured data	43%
P - Front-end Programming	42%
T - Machine Learning	41%
P - Systems Administration and Design	41%
M - Bayesian Statistics	39%
P - Back-end Programming	34%
P - Cloud Management	31%
T - NLP and text mining	30%
T - Big and distributed data	29%

Developer

Data Skill	% who have skill
T - Managing structured data	91%
S - Communication	85%
S - Data Mining and Viz Tools	76%
B - Product Design	75%
M - Math	75%
S - Data Management	75%
B - Project management	74%
P - Database Administration	73%
P - Back-end Programming	70%
P - Systems Administration	65%
S - Science/Scientific Method	64%
T - Unstructured data	64%
P - Front-end Programming	63%
M - Algorithms	61%
T - Machine Learning	54%
S - Statistics and statistical modeling	52%
M - Graphical Models	51%
B - Buisness development	48%
T - Big and distributed data	48%
M - Optimization	47%
B - Budgeting	45%
B - Goverance and Compliance	44%
M - Bayesian Statistics	44%
T - NLP	43%
P - Cloud Management	42%

Creative

Data Skill	% who have skill
S - Communication	87%
T - Managing structured data	79%
B - Project management	77%
S - Data Mining and Viz Tools	77%
M - Math	75%
B - Product Design and Development	68%
S - Science/Scientific Method	68%
S - Data Management	67%
S - Statistics and statistical modeling	63%
B - Business development	58%
B - Budgeting	58%
P - Database Administration	55%
M - Graphical Models	54%
M - Algorithms and Simulations	53%
T - Machine Learning	52%
T - Managing unstructured data	52%
M - Optimization	49%
P - Front-end Programming	49%
B - Governance and Compliance	48%
P - Systems Administration and Design	48%
M - Bayesian Statistics	45%
P - Back-end Programming	45%
T - NLP and text mining	39%
T - Big and distributed data	36%
P - Cloud Management	30%

Researcher

Data Skill	% who have skill
S - Communication	90%
S - Data Mining and Viz Tools	81%
M - Math	80%
S - Science/Scientific Method	78%
S - Statistics and statistical modeling	75%
T - Managing structured data	72%
S - Data Management	69%
B - Project management	68%
T - Machine Learning	58%
M - Optimization	55%
M - Algorithms and Simulations	55%
B - Product Design and Development	54%
M - Graphical Models	51%
M - Bayesian Statistics	50%
T - Managing unstructured data	48%
P - Database Administration	46%
B - Budgeting	44%
B - Business development	43%
T - NLP and text mining	41%
P - Systems Administration and Design	40%
P - Front-end Programming	28%
P - Back-end Programming	25%
B - Governance and Compliance	24%
T - Big and distributed data	22%
P - Cloud Management	21%

Note: % who have skill reflects percent of respondents who indicated that they have, at least, an intermediate level of proficiency.
Business Manager (e.g., leader, business person, entrepreneur) N = 250; Developer (e.g., developer, engineer) N = 222; Creative (e.g., Jack of all trades, artist, hacker) N = 221; Researcher (e.g., researcher, scientist, statistician) N = 353

BUSINESS BROADWAY

Copyright 2016 Business Over Broadway

25 Skills Assessed in the Data Science Survey

Skill Area	Skill Detail
Business	1. Product design and development
	2. Project management
	3. Business development
	4. Budgeting
	5. Governance & Compliance (e.g., security)
Technology	6. Managing unstructured data (e.g., noSQL)
	7. Managing structured data (e.g., SQL, JSON, XML)
	8. Natural Language Processing (NLP) and text mining
	9. Machine Learning (e.g., decision trees, neural nets, Support Vector Machine, clustering)
	10. Big and Distributed Data (e.g., Hadoop, Map/Reduce, Spark)
Math & Modeling	11. Optimization (e.g., linear, integer, convex, global)
	12. Math (e.g., linear algebra, real analysis, calculus)
	13. Graphical Models (e.g., social networks)
	14. Algorithms (e.g., computational complexity, Computer Science theory) and Simulations (e.g., discrete, agent-based, continuous)
	15. Bayesian Statistics (e.g., Markov Chain Monte Carlo)
Programming	16. Systems Administration (e.g., UNIX) and Design
	17. Database Administration (MySQL, NOSQL)
	18. Cloud Management
	19. Back-End Programming (e.g., JAVA/Rails/Objective C)
	20. Front-End Programming (e.g., JavaScript, HTML, CSS)
Statistics	21. Data Management (e.g., recoding, de-duplicating, integrating disparate data sources, Web scraping)
	22. Data Mining (e.g. R, Python, SPSS, SAS) and Visualization (e.g., graphics, mapping, web-based data visualization) tools
	23. Statistics and statistical modeling (e.g., general linear model, ANOVA, MANOVA, Spatio-temporal, Geographical Information System (GIS))
	24. Science/Scientific Method (e.g., experimental design, research design)
	25. Communication (e.g., sharing results, writing/publishing, presentations, blogging)

Respondents are asked to indicate how proficient they are for each of the 25 skills using the following scale: Don't Know (0), Fundamental Knowledge (20), Novice (40), Intermediate (60), Advanced (80), Expert (100).

AnalyticsWeek

BUSINESS BROADWAY

Copyright 2015 AnalyticsWeek and Business Over Broadway