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?

Introduction

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.

The workings of e-Science

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.

Striving to advance scientific knowledge

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.

E-Science or e-Research usually denotes data-intensive, IT-intensive and collaborative research, but it also can just refer to research that uses explicitly defined IT support research workflows.

Emerging e-Science can be seen as result of (a) the need for processing huge amounts of diverse data, (b) increased collaboration between labs (and therefore various emerging forms of social computing and workflow modeling for researchers), and (c) what could be called the application of engineering principles to work organization.

One way of looking e-Science is the idea to make scientific workflows easier to create and to share. Scientific research is increasingly digital. Some activities, like data analysis, search and simulation, can be accelerated by enabling scientist to write workflows and scripts that automate routine activities.

Scientific workflows

A Scientific Workflow Systems is a specialized form of a workflow management system designed specifically to compose and execute a series of computational or data manipulation steps, or a workflow, in a scientific application.

What is meant by e-Science? In the future, e-Science will refer to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualization back to the individual user scientists.

e-Science projects usually refers to setups that include the following ingredients

- an environment (e. g. desktop software or a web application) that allows researchers to define a workflow,
- an environment (e. g. desktop software or a web application) that allows researchers to ex cute a workflow (can be the same as above). Execution includes tracking data flow and steps,
- a repository for sharing and reusing workflows made by others.

Scientific Workflows have become an increasingly popular paradigm for scientists to formalize and structure complex scientific processes to enable and accelerate many scientific discoveries. A scientific workflow is a formal specification of a scientific process, which represents, streamlines and automates the analytical and computational steps that a scientist needs to go through from dataset selection and integration, computation and analysis, to final data product presentation and visualization. A workflow system is a holistic unit that defines, manages, and executes workflow processes aided by software. There exist many variants of workflow systems. E. G. based on workflow design (workflow composition systems), workflow scheduling (architecture, decision making, planning scheme, scheduling strategies), fault tolerance and data movement.

Internally, a workflow system uses a Workflow Language ore Meta-Languages for process specification, to define the workflow process logic, to be executed by workflow execution engine or workflow controller.