CrossMark

EDITORIAL

CSBS Editorial

Volker Beckmann[1] · Markus Elsing[2] · Günter Quast[3]

Dear Reader,

Thank you for giving attention to this first set of articles of the new journal "Computing and Software for Big Science", which aims to present innovative concepts for large-scale, collaborative computing and software developments in the fields of particle, astro-particle and nuclear physics, but also for observational astronomy and cosmology , and high-brilliance light sources. In the absence of a generic term, we chose to encompass these areas of research, characterised by major international collaborations, large data volumes, and huge demands on computing resources, under the title "Big Science".

Advancements in science have always been linked to technological developments that open up novel opportunities to seek answers to the most complex scientific questions. In the past few decades, the exponential growth in computing power, together with the tremendous progress in software development and the availability of global networks have revolutionised the way we do scientific research. In scientific projects today, software development is becoming an ever more important part of the overall effort. Indeed, in some areas, science is driving the development of cutting-edge software and computing technology.

As an example, the first accelerator-based particle physics experiments were invented in the second half of the twentieth century, at the same time as computers. Ever since then, progress in particle physics has been intimately linked, not only to the development of accelerators and detectors, but also to advancements in computing technologies [1]. Due to the statistical nature of the underlying laws of physics, every new generation of experiments has been required to process ever increasing data volumes. Large international collaborations have come into existence, pushing back technological boundaries and facilitating many fundamental discoveries. Within these big science collaborations, physicists have developed paradigms for collaborative and geographically distributed software development. Novel algorithmic approaches have been invented to extract fundamental physics results from data. Over the past few decades, the needs of large physics experiments and scientific collaborations has also resulted in several ground-breaking developments in the field of software and computing. One of these, the invention of the Web [2] at CERN, not only changed the way information is exchanged between collaborators in the Large Electron Positron Collider experiments of the 1990s, but radically changed the way society has used information technologies ever since. The need to store and process the unprecedented data volumes of today's Large Hadron Collider experiments led to the development of the worldwide LHC Computing Grid [3] in the late 1990s, a federation of more than 150 computing centres, located in universities and national laboratories all around the globe. The Grid provided a novel answer to the sheer scale of the LHC computing problem, and this long before the term Big Data had become common place, or indeed a line of business for today's big internet companies.

Other scientific communities have followed these developments, and new scientific opportunities in the fields mentioned above involve data intensive computing challenges [4]. Pioneering projects are already under way or under

✉ Markus Elsing
   Markus.Elsing@cern.ch

1 Institut National de Physique Nucléaire et de Physique des Particules, CNRS, Paris, France

2 CERN, Geneva, Switzerland

3 Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany

preparation. Sky survey projects, like the Large Synoptic Survey Telescope, are employing sophisticated data processing strategies to achieve high throughput processing for the extraction of physics information. In the gamma-ray domain, the Cherenkov Telescope Array is considering a computing grid analysis approach, and upcoming projects like the Square Kilometre Array will reach data volumes at the exabyte scale. "Triggerless" readout schemes, currently pioneered by LHCb and ALICE, are based on powerful event selection algorithms deployed in online farms and will boost the physics potential of such particle physics experiments. When it becomes operational in the year 2025, the High-Luminosity LHC will pose tremendous challenges to reconstruction and trigger selection algorithms because of the unprecedented complexity of an average of 200 overlapping proton–proton interactions in every recorded event. Accelerator-based neutrino experiments require novel concepts for three-dimensional event reconstruction. All these present and future experiments require innovative techniques for event and object classification, processing, hosting, distribution, quality monitoring, and visualisation of data. Precise modelling of all the relevant physical processes and of the detector response requires sophisticated and CPU-hungry simulation codes.

On the other hand, big science projects depend on developments in computer science and information technology. A very active field in data science considers algorithmic approaches based on machine learning. These offer a huge potential to solve complex scientific problems, from classification to feature extraction and data modelling. Machine learning is more than just an add-on to traditional algorithms—it really has the potential to revolutionise scientific data analysis once again. The rapid changes in processor technologies pose an equally important challenge for the future of scientific computing. Today's processors deploy many cores and are able to vectorise calculations. This requires carefully crafted memory access patterns. Traditional, algorithm-driven data processing strategies need to be rethought to match these new processing paradigms. Event processing frameworks need to support the development of CPU-efficient software, and benchmarking and performance assessment tools are needed to optimise applications for the computing hardware. At the same time computing grids, emerging scientific clouds, current high-performance computing centres, and even volunteer resources offer solutions to different computing problems, from data-in-tensive processing to CPU-intensive calculations. Such resources are leveraged by sophisticated middleware services for distributed computing.

The scientific community is committed to successfully tackling the software and computing challenges of present and next generation experiments. In contrast to what happens in industry, software development for scientific projects is organised around geographically distributed teams from many institutes and laboratories. Developments are often driven by science students working in small teams with backgrounds in software engineering. For many decades, the education of students in computing and software for big science has been at the forefront of information technology, but today big science projects are competing with big data companies for the best data scientists [5]. And it is important for a career in science to be able to publish high quality articles in order to gain recognition for vital contributions to the field, an opportunity which this new journal offers to the computing and software community of today's big science projects.

As a peer-reviewed journal "Computing and Software for Big Science" is therefore specifically dedicated to the publication of high-quality material originating from the collective effort by this scientific community, where experimental research is increasingly organised in large and global collaborations around large-scale instruments with a huge output of data. Faced with the above-mentioned challenges, our scientific community requires fundamental and novel concepts for large-scale and collaborative computing and software development, as well as novel algorithms and techniques for data processing. The journal aims to capture novel trends and innovations in infrastructures for large-scale, high-throughput computing, as well as in related middleware developments. New concepts for data processing, hosting, and sharing will be a focus, alongside novel ideas for distributed data analysis. Important subjects will be data processing frameworks, software integration, benchmarking, and performance assessment. Novel algorithms for efficient data reconstruction and filtering, as well as event and object classification methods, will be presented. An important subject will be the application of Deep Learning techniques to scientific data processing, replacing more classical approaches and opening up new perspectives. And last, but not least, monitoring and data visualisation techniques will be prominently covered in the journal.

The papers solicited include primarily research articles presenting new and original results, review papers, the scientific aspects of white papers describing future collaborations and/or facilities, advanced, self-contained tutorials, as well as documentation papers with the explicit aim of collecting and combining knowledge spread over many unpublished, internal documents to foster proper technology transfer. Opinion papers may appear from time to time, upon invitation. Submissions of software manuals or algorithm descriptions without significant scientific context will not be considered.

The journal encourages as far as possible the publication of relevant data and software—this can happen either in the form of electronic material supplementary to articles accepted in this journal or via dedicated repositories that

offer persistent identifiers (such as the DOI) for proper citation and findability. More details are given in the journal's instructions for authors in the section Research Data Policy [6].

We look forward to working with you to support the ongoing development of computing and software for current and future big science projects!

The Editors-in-Chief,

Volker Beckmann, Markus Elsing, and Günter Quast

References

1. Carminati F (2012) Preface of the book "From the Web to the Grid and Beyond". Springer, Heidelberg
2. Berners-Lee T (1989) Information management: a proposal, CERN/DD, Geneva
3. Forster I, Kesselmann C (1998) The grid: blueprint for a new computing infrastructure. Morgan Kaufmann Publishers Inc., San Francisco
4. Berghöfer T et al (2015) Towards a model for computing in european astroparticle physics. AP-PEC white paper, 2015
5. Heiß A, invited opinion paper
6. http://www.springer.com/journal/41781. Accessed 11 Sept 2017