

Software and Computing for Big Science

Michel Jouvin

CNRS/IN2P3/LAL

jouvin@lal.in2p3.fr

HEPiX, Barcelone, 8/10/2018



History Recap

- Idea presented by Y. Kemp (DESY) during HEPiX in Berlin (spring 2016)
 - Originally proposed by Günter Quast (CMS, KIT, Karlsruhe) in 2013 during a HEP conference in Mainz (Germany)
 - Took 18 months for maturing the idea
- Journal launched in Spring 2017
 - Presented at HEPiX Budapest
 - Focused particularly on HEP, Nuclear Physics, Astrophysics, Cosmology and Photon Science
- Big Science (see Wikipedia): science based on big experiments requiring a long planning, collaborations, a lot of money... and a lot of computing!
- Editorial Board: several HEPiX members
 - 3 editors-in-chief: Günter Quast (Germany), Markus Elsing (CERN), Volker Beckmann (France)
 - Associate editors: ~15 from all continents (Europe, N.A., Asia) and horizons including G. Chen, Y. Kemp, H. Meinhard, M. Jouvin

Motivation

- Several scientific communities produce and analyze large/huge amount of data ($O(\text{PB})$) in experiments organized as large (international) collaborations and relying on computing
 - HEP with LHC experiments is currently the archetype but coming astrophysics and cosmology experiments (CTA, SKA, LSST), NP experiments (FAIR) on the same line
 - Similar trends in photon science, biology...
- Sharing many computing technologies and facing common computing challenges
 - We know this at HEPiX! Regular presentations by other communities... like Alf in Berlin
 - Not only on the computing resource side but also on the software application side: in particular the challenge of parallelization and efficient use of new processor/machine architectures
 - No community can address these alone: collaboration is a requirement, publication a prerequisite
- In addition, the challenge of managing/distributing large amount of data: several frameworks/infrastructures already developed
 - But often not really known outside of the experiments that develop them
 - Sharing the information about it would avoid to reinvent the wheel for these complex systems

Publication Challenge

- (Almost) no place to publish about our R&D and solutions
 - Too computing-related to be published where our scientific results are published
 - Not enough research (too much engineering) to be published in computing science reviews
 - CHEP proceedings is the main publication but not frequent enough (18 months) and difficult to share with other communities
 - What is published is too dispersed: difficult to identify and thus failing to contribute to knowledge sharing
- Importance of a validated, peer-reviewed journal
 - Goal: be a reference archive for knowledge sharing in big, data-intensive, sciences about software and computing challenges and solutions
 - Means that contents must be validated and its value recognized: impact factor, even if low
 - Citation indexing required for getting an impact factor : 20-25 articles/year during 2 years before it is considered
 - Help with career recognition of « Research SW Engineers », in particular the young ones

Topics: Everything related to SW&Computing

- Infrastructures for large-scale, high-throughput computing
- Middleware development
- Data processing, hosting and sharing
- Distributed data analysis
- Software development Infrastructures
- Software benchmarking and Performance Assessment
- Frameworks and software integration
- Novel algorithms for efficient data reconstruction and filtering
- Deep learning algorithms
- Event and object classification
- Online/Offline data quality monitoring
- Data visualization



Journal Facts



- Online journal, continuous publication, in partnership with Springer
 - Christian Caron (Executive Publishing Editor @Springer) associated with the project since its early stage (present at HEPiX in Berlin and CHEP in Sofia)
 - Springer provides the manuscript handling system to help with an efficient peer-review
 - Hybrid business model rather than pure Open Access until potential inclusion into SCOAP³
 - Too late for including in phase 2 (2017-2019) but positive preliminary contacts for next phase
 - Also topic-agnostic funding by country/region (e.g. several European countries) leading to open access as soon as there are some authors from the country
 - Springer decided to allow free access during the 2 first years... so in fact Open Access till end 2019
 - 1 paper issue per year, with an editorial from the Editors-in-Chief or invited persons
- Types of articles: topical articles, advanced tutorials, project/technology reviews
 - No proceedings: may cause the journal to be excluded from indexing, other places like CHEP
 - Do not duplicate what is published somewhere else: would damage the journal image



Journal Status

- 2017 : 4 articles published
 - Printed issue available: ask me if interested
 - 1 article with 1K downloads, 145 media share (e.g. Twitter), 11 citations
- 2018 : 6 articles already published, 6 in progress, 4K downloads
 - Several related to machine learning, a hot topic in HEP and astroparticles
 - Many more submissions (10) but several rejected, in particular because of authorship questions (e.g. small group of authors instead of a collaboration)
 - HEP Roadmap for the 2020s (aka CWP) will be published soon (final revision stage)
- A good start but still at a slow pace: objective remains 20 articles/year
 - Takes time to get known: many potential authors contacted didn't know the journal
 - A lot of advanced SW&C activities in Astrophysics (CTA, SKA) and Cosmology (LSST, LISA)
 - Chicken&egg problem with the Impact Factor

This is Your Journal



- No journal without readers... but also without authors!
 - You are among the potential authors
 - Contact us (at E.B.) to discuss your ideas... or submit them directly!
 - <https://link.springer.com/journal/41781>
- The journal is about computing also, not only software
 - So far, no computing-related articles received
 - Some of the work about storage and large-scale computing may be worth an article
- The journal is considering publishing computing school materials as tutorial
 - Must be advanced courses
- Journal: an opportunity to increase collaboration with other communities
 - HEPiX has a role to play as this is something we have done for a long time