Towards Automation of Scientific Methodologies of 2040: Software and Services

Shantenu Jha*, Dong Ahn, Kyle Chard, Dan Laney, Todd Munson & Arvind Ramanathan

The defining property of future scientific methodologies (FSM) in 2040 will be **automation**, and primary requirements will be a concomitant ecosystem of composable software and services. Automation will raise the level of abstraction at which scientific discovery is achieved!

Automation will not just remove the mundane and tedious, but also replace human errors in judgement and observation, as well as biases. The next frontier of automation will transcend simple automation of execution processes, and will substitute human thinking and ingenuity. For example, *ab initio* scientific discovery by coupling computation guided hypotheses generation with wet lab testing. As computational theorem proving is no longer the preserve of human ingenuity, nor the pinnacle of human logical reasoning, hypotheses generation and testing will no longer be associated solely with human creativity and ingenuity.

Automation will impact FSM along multiple dimensions and levels: from **what** to compute (e.g., best property to study), **how** to compute (e.g., which method is optimally accurate for the given amount of time or resource), **where** (e.g., translating an abstract plan into an orchestrated campaign across a federated set of services, which then map into self configuring, automated and heterogeneous software and hardware systems), and even **when** (e.g., optimal plans to achieve multiple objectives).

What are the implications of extreme automation on scientific software and services? How must the design, development and deployment of scientific software evolve to meet these requirements? What must the software ecosystem and sustainability practices be, so as to manage if not reverse the hyper-inflationary cost (price) of reliable software ? We do not have specific answers, or a roadmap, but we believe it is imperative to fundamentally rethink the design, development processes and reproducibility of community software.

We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run. This is also likely to be true for Machine Learning. Whereas it will be difficult to formulate its qualitative and quantitative impact on FSM-2040, we can confidently say it will induce unprecedented if not unimaginable degrees of automation in FSM. However, we risk overestimating the impact of automation in the long run, if we underestimate the challenges of designing and developing the software infrastructure in the short term.

To address these challenges, we are pursuing a building block approach to the design and development of workflow systems (ExaWorks). Workflows are the current incarnation and unsophisticated precursors of automated FSM-2040. Lessons from ExaWorks will provide: (i) an engineering basis and experience of how to develop modular components; (ii) how to make them extensible and composable, and (iii) important lessons and insight in the design and deployment of future self configuring automated services.