Journal of Computational Science 14 (2016) 85–89

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

I have a DRIHM: A case study in lifting computational science services up to the scientific mainstream

Michael Schiffers [a,*], Nils gentschen Felde [a], Dieter Kranzlmüller [a,b]

[a] Ludwig-Maximilians-Universität München, Oettingenstr. 67, 80538 München, Germany
[b] Leibniz Supercomputing Centre, Boltzmannstr. 1, 85741 Garching, Germany

ARTICLE INFO

Article history:
Received 1 July 2015
Received in revised form 25 November 2015
Accepted 12 December 2015
Available online 22 January 2016

Keywords:
Research infrastructure
Extreme scale computing
Hydro-meteorology
Chasm model

ABSTRACT

While we are witnessing a transition from petascale to exascale computing, we experience, when teaching students and scientists to adopt distributed computing infrastructures for computational science, what Geoffrey A. Moore once coined "the chasm" between the visionaries in computational science and the early majority of scientific pragmatists. Using the EU-funded DRIHM project (Distributed Research Infrastructure for Hydro-Meteorology) as a case study, we observe that innovative research infrastructures have difficulties to be accepted by the scientific pragmatists because the infrastructure and scientific services are not "mainstream" enough. Excellence in workforce, however, can only be achieved if the tools are not only available but also used. In this paper we report on how "this chasm" is exhibited in the DRIHM case, how it was (partially) crossed, and what can be learned from this experience for more general cases.

1. Background

Although supercomputing performance gains are slowing down, rather than speeding up (see the recent lists at http://www.top500.org/), exascale performance is predicted to happen around 2020 [1]. While the European Technology Platform for High Performance Computing (ETP4HPC) alerts to the growing gap between science and technology education at degree level on the one hand and the needs of the Science & Technology labor market on the other hand [2], the Exascale Mathematics Working Group proposes to "increase the pool of computational scientists and mathematicians trained in both applied mathematics and high-performance computing" [3]. The 2010 Fall report of the American Advanced Scientific Computing Advisory Committee (ASCAC) concurs with this perspective by recommending to renew efforts to recruit and train the next generation of scientists needed to exploit exascale capabilities for *e-science* applications over *e-infrastructures* [4].

However, inherently associated with this endeavor are two challenges: First, any successful exploitation of exascale capabilities requires exploitable items provided by somebody; second, a critical mass of scientists is required to adopt these items for solving their research questions. Obviously, any successful treatment of these challenges not only depends on the various items' readiness levels [5], but also on the "diffusion rate" [6] in the sense of community adoption speed. While we do not address the technical issues of the former challenge, we emphasize the latter. It relates to the velocity of lifting up computational science services from fad to the scientific "mainstream", or – in other words – from a knowledge of the few to a knowledge of the (scientific) crowd.

In this paper we propose to look at computational science services not only from a pure technology-driven or science-driven perspective. Instead, we suggest to look at them also from a *marketing*-driven angle. Our argumentation is based on the experiences gained while running several EU-funded projects over the last five years. In particular we refer to the DRIHM project (Distributed Research Infrastructure for Hydro-Meteorology[1]) as a typical multi-disciplinary representative of the Hydro-Meteorological Research (HMR) community.

The paper is inspired by Everett Roger's bell-shaped technology adoption life cycle model [6] and Moore's revised version (Fig. 1) for marketing disruptive high tech products [7], the latter also known as the "chasm model". Moore identified a "deep and dividing chasm" between "early adopters" and the "early majority" because the former are looking for some kind of "change agent" while the latter group – those that dominate the "mainstream" – aims at getting a "productivity improvement".

* Corresponding author.
E-mail address: michael.schiffers@ifi.lmu.de (M. Schiffers).

[1] http://www.drihm.eu.

http://dx.doi.org/10.1016/j.jocs.2015.12.008
1877-7503/

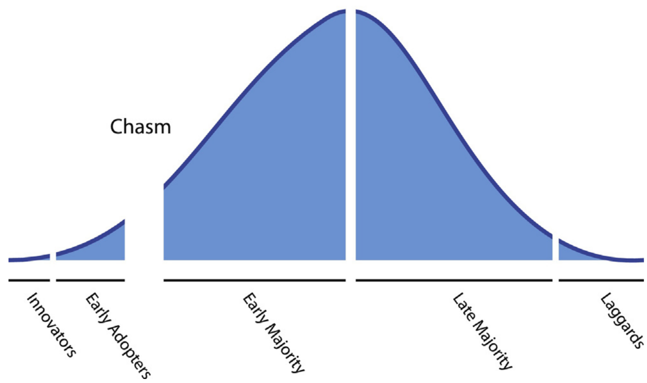

Fig. 1. Chasm model of the technology adoption life cycle [7].

The analogy between Moore's marketing view and observations in the DRIHM project (as a representative for both a provider of computational science services and a user of such services) renders obvious:

- Moore and DRIHM both share a focus on discontinuous innovations with change-sensitive products in [7] and change-sensitive HMR services in the DRIHM project (especially when dealing with extreme scalability).
- Moore and DRIHM both share a perspective on target groups distinguishable by their characteristic response to discontinuous innovations (as the groups in Fig. 1 indicate). Moore calls them "psychographic groups" [7], DRIHM calls them "stakeholder groups".
- Moore and DRIHM both share the observation of dissociation – that is, the difficulty any group has in accepting an innovation if it is presented in the same way as it was to the group "to its immediate left" in Fig. 1.

The rest of the paper is organized as follows: In Sections 2 and 3 we first describe the DRIHM project before reporting on observations made during the course of the project and demonstrating for computational science some kind of "chasm" between groups of scientists. Given the chasm(s) we apply Moore's model of chasm crossing in Section 4. Finally, Section 5 concludes the paper.

2. The DRIHM case of hydro-meteorological research

Predicting the impact of weather and climate phenomena on the environment is one of the main challenges of this century with significant societal and economic implications [8,9]. Central to all solution approaches are both the ability to transparently couple hydro-meteorological data and models, and to facilitate the cross-organizational collaboration between meteorologists, hydrologists, and earth science experts for jointly advancing HMR.

DRIHM aims at supporting HMR by providing a Grid-based high performance computing infrastructure to orchestrate and execute linear chains of multiscale/multiphysics models [10,11]. Inspired by such disastrous flood events like Genoa 2011, Genoa 2014, and Serbia 2014, the project could demonstrate the value of the deployed services (infrastructure ones and scientific ones [12,13]).

1. Using the multi-model ensemble for meteorological models, both the intensity and the timing of the event could be predicted in post-event research mode with a very high degree of accuracy.
2. By running hydrologic models from the ensemble of meteorological outputs, it was possible to precisely predict the extreme discharges.
3. The hydraulic model composition for studying damages – for example those caused by passages of water under a railway station – predicted the hazard to life as well as damages very much in accordance with the local authority's estimates.
4. The achievements were presented at the March 2015 UN World Conference on Disaster Risk Reduction, Sendai, Japan.[2]

Technically speaking, DRIHM provides a web-based portal for HMR scientists (see Fig. 2).

The portal implements all necessary authentication and authorization mechanisms and it supports the orchestration and execution of model chains on Grid resources (e.g., resources granted by the European Grid Infrastructure (EGI)[3]), Cloud resources (e.g., resources available through EGI's Federated Cloud Initiative[4]), HPC resources (e.g., resources provided through the Partnership for Advanced Computing in Europe (PRACE)[5]), or proprietary resources accessible via web services interfaces or any other specific clients.

3. Observations

DRIHM represents a class of inter-disciplinary projects that are aiming at large scale computing while at the same time processing big data sets across organizational boundaries. While technical issues are just one side of the coin, regulatory frameworks to adhere to, adequate education and training programs, and tailored IT-Management processes have to be considered equally [14]. Fig. 3 illustrates this schematically as a matrix spanned by the vertical technology-related e-Infrastructure axis and the horizontal topic-related aspect axis.

While DRIHM used the matrix as a reference framework for guiding the development process, we briefly share sample observations in selected cells A, B, D, and column C of Fig. 3.

Cell A addresses the big data issue for extreme scale scientific software. Because there is currently no common modeling and big data processing methodology available for HMR, big data analysis is more focused on exploiting parallelism in processing than on the data itself. The most common mechanisms to handle big data consists of storing data in some kind of key-value store that is easily distributable, and then use extreme parallelism to analyze the data.

The problem with this approach, however, is that too much information about the data and its structure is lost over the complete workflow of model chains. Multi-level algorithms that need data with different granularities do not match well with this model of storage. Consequently, generic data management was abandoned in favor of casuistic hard-coded, but un-portable, solutions to support HMR. This yields

Observation I: The technological context is primarily set by "familiar" work habits.

Cooperation across organizational boundaries raises questions of data access rights, file pinning permissions, and authorized execution of models on dynamically selected resources. Unfortunately, providers of resources, data, and services belong to multiple legal domains with often incompatible regulatory frameworks (cell B) – typically expressed as intellectual property rights, privacy protection rules, and certificate handling policies. In DRIHM we decided to take the line of the least resistance by completely adhering to European regulations – although they differ as well. A general legal framework is missing.

[2] http://www.wcdrr.org/home.
[3] http://www.egi.eu.
[4] https://wiki.egi.eu/wiki/Fedcloud-tf:Main.
[5] http://www.prace-ri.eu/.