For Better Innovation: From e-Science to Data Driven Discovery

Dr. Mingqi CHEN
The office of The leader Group of Cyberspace Affairs, CAS
24th Oct. 2017
The Big trend of today's scientific research
The Coming Era – Data Intensive Discovery

Innovations within Data Intensive Programs

More and more innovations are happening in data-intensive programs.

Scientific Data is now acting significant role as Engine of Innovations
1. CyberInfrastructure

- Network
- Supercomputing
- Scientific Data
China Science and Technology Network (CSTNet) is one of the earliest internet networks in China.
Supercomputing

- Three-tier grid
  - 1 head center and operation center
    - Eea, 2.36 Petaflops
  - 12 regional centers
  - 18 institution centers, 11 GPU centers
- Applications
  - Computational Chemistry, Physics, Material science, Life science, CFD, Industrial computing… etc
- CAS Computational Science Application Research Center

CNIC: Public HPC service provider in CAS
CNGrid: national HPC environment

Northern Main Node

Sunway TaihuLight (World #1)

Southern Main Node

Tianhe II (World #2)
Scientific Data

- A Long-term mission started in 1986 which funded by CAS
  - many institutes involved
  - long-term, large-scale collaboration
  - data from research, for research

- Collecting multi-discipline research data and promoting data sharing
  - More than 350 research databases and 500 datasets by 61 institutes
  - Over 200TB data available to open access and download
Data analysis pipeline for gene annotation, visualization and functional analysis

- Massive
- Distributed
- Heterogeneous

The Institute of Microbiology

2、Typical Cases: Microbial Big Data Platform

- One-stop data search engine
- Automatic resources allocation

Data Computing

Pfam

PDB

Typical Cases:
Microbial Big Data Platform

The Institute of Microbiology
Microbial Big Data Platform

- Propose the first ISO standards in the field of microbial resources information

ISO / PWI 21710 biotechnology-Data Management and publication in microbial biological resource centers.

Global Catalogue of Microorganisms

- 98 bio-resources centers from 42 countries sharing data through the platform
- largest international cooperation project in the field of microbiology

Microbial Big Data Portal for Chinese National data project (863 project)
Super Monte Carlo: Simulation Program for Nuclear and Radiation Process

Institute of Nuclear Energy Safety Technology (INEST)

support design optimization and safety assessment of nuclear systems, **600+ institutions in 58 nations**
10M-Core Scalable Fully-Implicit Solver for Nonhydrostatic Atmospheric Dynamics

- The World’s first 10m-core scalable fully implicit solver

- **Related Achievements:** Won the ACM Gordon Bell Prize in 2016

Greatly enhanced simulation capability of atmospheric dynamics at extreme-scale

Won the ACM Gordon Bell Prize for China for the first time in the 29-year history of this prestigious award
3. Planning

11th five years planning (2006-2015)

Focus on Cyberinfrastructure and e-Science
12th five years planning (2011-2015)

Focus on Cloud Cyberinfrastructure

- e-Science Application Promotion Project
- Management and Strategic Decision Support Project
- Continuation Education, Science and Education Integration Project
- Scientific Data Resource Integration and Sharing Project
- Science Communication Service based on Internet
- Network Security Guarantee and Service Project
The 13th Plan (2016-2020) for Network Security and Informatization

Five main Projects

<table>
<thead>
<tr>
<th>Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Project of Chinese science and technology Cloud</td>
</tr>
<tr>
<td>The Project of Intelligent CAS</td>
</tr>
<tr>
<td>The Project of Scientific Big Data</td>
</tr>
<tr>
<td>The Project of e-Science Applications</td>
</tr>
<tr>
<td>The Project of Network Security</td>
</tr>
</tbody>
</table>

Focus on Cloud and big data driven innovation
4. Culture

e-Science Technology & Application

created in 2008, published bimonthly in Chinese with abstracts in English (now)

✓ Publication number:
  CN 11-5943/TP ISSN 1674-9480

✓ Authority: CAS

✓ Organizer: CNIC, CAS
  Science Press Co., Ltd.

✓ Publisher: Science Press Co., Ltd.
China Scientific Data created in 2016. The journal is dedicated to promoting the sharing and citation of scientific data, and to making them findable, accessible, intelligible and reusable. Data papers describing (but not limited to): Datasets or data products generated from major scientific activities; Derived datasets or data products refined from raw data; Datasets linked to existing publications.

Publication number: CN11-6035/N, ISSN 2096-2223

✓ Authority: CAS, CODATA China National Committee
✓ Organizer: CNIC, CAS Science Press Co., Ltd.
✓ Publisher: Science Press Co., Ltd.
Training

- CIO Training Program: Started in 2010
  - Objective: Let trainees understand and value eScience and promote eScience applications in their Institutes
  - Attendance: Directors of the institutes of CAS(CIOs), 300+ person-time totally.
China e-Science Forum

1st Forum, 3 December 2009
• Co-organized by CAS, MOST, NSFC
2nd Forum, 15-16 Dec. 2011, Beijing
3rd Forum, 21-22 Oct. 2013, Beijing
4th Forum, 8-9 Dec. 2015, Beijing

8 government departments: CAS/CAC (Cyberspace Administration of China) /MOE/MOST/MIIT/CASS/CAE/NSFC

5th Forum, 4 Dec. 2017, Wuzhen Zhejiang
held by CAS and China Association for Science and Technology

200~ Attendance

450~ Attendance
China’s e-Science Blue Book Series

- First Blue Book in China
- CAS
- Ministry of Education
- NSF of China

Part 1 Trend and Strategy
- International and domestic situation of e-Science and development trend and strategy of eScience

Part 2 Technology
- eScience 2.0, high performance computing, cloud computing

Part 3 CyberInfrastructure
- Network, supercomputing and data facilities for eScience

Part 4 Applications
- Typical eScience applications in various disciplines (10 application in China)
Thank you!

mqchen@cashq.ac.cn

Q&A