

## National e-Science Data Service in China

Hui ZHANG

School of Computer Science and Engineering, Beijing University of Aeronautics &  
Astronautics, Beijing

Email: [zhangh@escience.gov.cn](mailto:zhangh@escience.gov.cn)

In this talk, I would like to share our experience and successful stories in developing the national scientific technology infrastructure (NSTI: [www.escience.gov.cn](http://www.escience.gov.cn)) in China to support long-term data sharing and enable co-creation through for many scientific domains. The mission of the data cyberinfrastructure is to promote the scientific data and knowledge sharing across the country by preserving and curating multi-disciplinary science datasets. The major data sources for our system come either from dozens of domain-oriented platforms such as public health portal and microbe data portal, or from the public web sites.

Our system adopts a distributed and cloud-based architecture. We established a data center with 200 nodes and 1000TB storage to host the uploaded datasets and run large-scale data analysis for information retrieval. We introduced a novel data pyramid model for metadata management, which can capture all the attributes of science data. Based on the semantic data model, we designed a powerful search engine to provide rich query services. User can easily locate the scientific datasets that are available from the domain portals and subsequently contact the data providers for retrieve the real datasets.

In the next step, we plan to extend our cloud-based data center to preserve massive scale raw datasets. A data submission system has been developed to facilitate researches to upload any forms of scientific datasets ranging from experimental reports to data samples collected from instruments. This next generation of data center is able to support much deeper data analysis and data mining tasks. More importantly, we would open these services to scientists, educators and students following the cloud "SaaS" paradigm. In this way, a vibrant e-Science ecosystem will be emerging from our data platform and accelerate the growth of scientific community on the data collaboration and curation.