第五屆 電子檔案管理技術國際研討會

The 5th International Conference on Electronic Records Management & Technology

December 13-14, 2016

SPEAKER BIOGRAPHY

Name	Gerhard J. A. Schneider

CURRENT POSITIONS

CIO of the University and Director of the Computer Center,
The University of Freiburg
Germany

EDUCATIONAL AND CAREER EXPERIENCES

- 1978 Diploma in Maths (Erlangen)
- 1979 MSc in Maths (Oxford)
- 1981 PhD in Maths (Essen)
- 1988 Habilitation in Maths (Essen)
- 1992 Professor for Decentralized Systems (Karlsruhe)
- 1997 Director of GWDG, Göttingen
- 2002 Director of the Freiburg University Computer Centre
- 2003-2008 ViceRector of Freiburg University
- 2008-now CIO of Freiburg University
- Member of various boards of DFG and DFN

AWARDS AND HONORS

- 2015: 2nd prize for Innovation in Academic Computer Centres (sponsor: Bechtle AG)
- 2012: Instructional Development Award for e-exams (sponsor: University of Freiburg)
- 2011: best lecturer award (sponsor: Faculty of Freiburg University)
- 2008: Campus Online (sponsor: Stifterverband der dt. Wissenschaft)

PRESENTATION

Preserving run time environments on parallel computers

- an essential aspect for research data management

Abstract:

The successful preservation of data generated or collected during research is a key issue in modern science and for the Open Data concept. Reusability of such data not only requires a reliable bitstream preservation as well as reference systems but also knowledge about how data was presented and how it was interpreted. Old word processing programs highlight this challenge, as old files – though available – can no longer be displayed, let alone understood.

Emulation is vital to re-activate old run time systems; advances on this field were presented in a previous talk.

The complexity of modern “virtual research environments” by far exceeds traditional approaches in research. Nowadays we find cross-wise dependencies of operating systems (including updates) with monitoring software for research devices (like microscopes), research support programs (like electronic lab-books) and data processing software (home grown or standard systems like MATLAB) or data base systems.

A successful management of research generated data therefore requires not only knowledge of the well-known meta data but more or less a complete preservation of the run time system, so that the full environment can be replayed at a later stage, using emulation.

In a research project we try to analyze a process to predict the later re-usability of data produced during a PhD.

The problem becomes more challenging if a parallel computer is used to manipulate research data. Rather than trying to re-run the compute jobs later we now try to run our new parallel computer in a novel way so that we can preserve the complete runtime job right from the outset. Again emulation is the solution; on each node of the computer virtual machines are used rather than the traditional bare metal approach. It turns out that the inevitable loss of performance is negligible. Apart from being able to address the issues of long term preservation, this has the additional effect that tedious recompilations of programs are no longer required, thus simplifying the move between super computers. The successful interpretation of data from CERN done by our physicists proves that our approach is working.