

# Some personal thoughts on Training/Education of e-Infrastructure Users

e-IRG Workshop Athens  
10 June 2014

Colin Wright

National Integrated Cyber-Infrastructure System

DST & CSIR Meraka Institute

South Africa

[cjwright@csir.co.za](mailto:cjwright@csir.co.za)

[Colin.wright@wits.ac.za](mailto:Colin.wright@wits.ac.za)



science  
& technology

Department:  
Science and Technology  
REPUBLIC OF SOUTH AFRICA

e-IRG 20140610



# Role of e-Infrastructures

A1

**E-Infrastructures  
support many areas  
of research**

HEP  
Research Infrastructures  
Palaeo.. Societal Challenges  
Climate Change eResearch  
Simulation Human Brain Project  
Comp Psychology Data intensive Research  
Earth Observation Astronomy  
Archives Comp Economics  
.....

**Tier 1**

**E-Infrastructures**

**Tier 2**

**Tier 3**

**E-Infrastructures  
integrate resources and  
services .....**

- Networking
- Computing
- Data
- Software
- User Interfaces

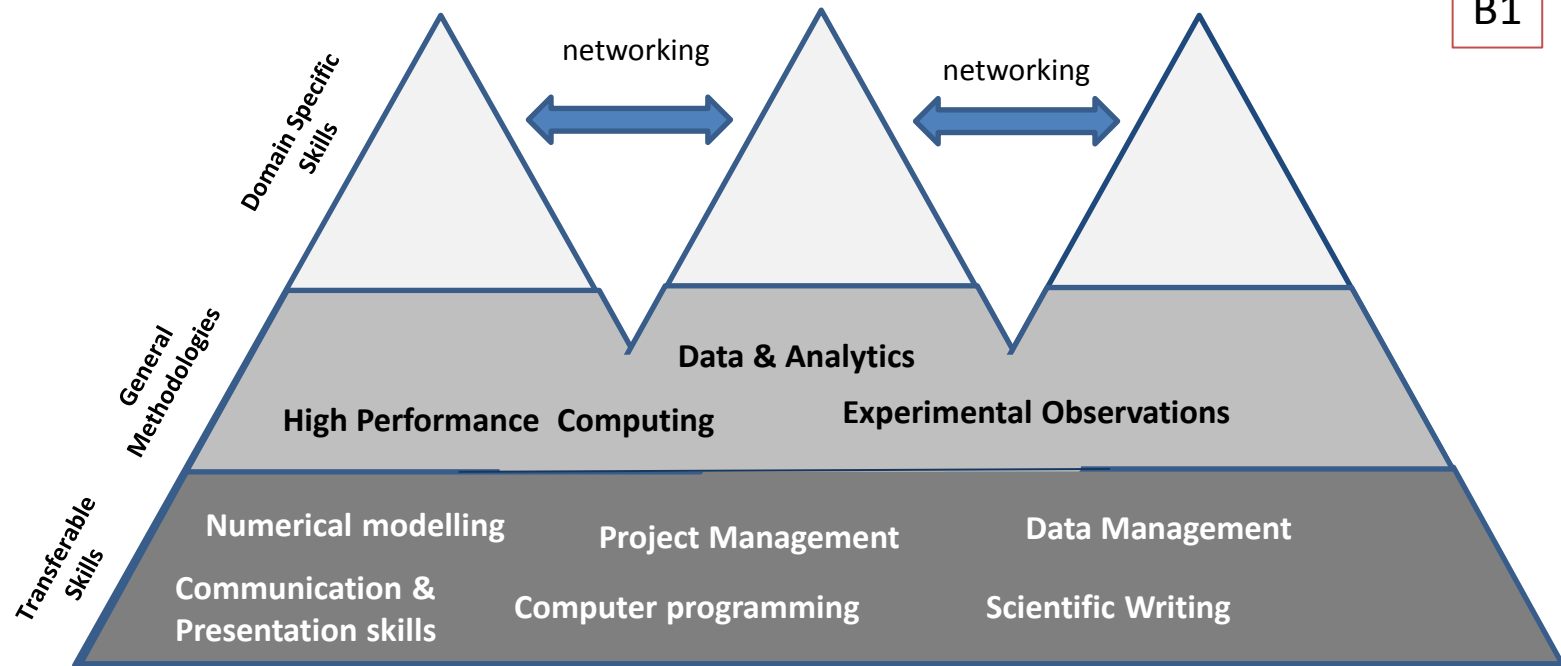


Cyber-Infrastructure

- Network Infrastructure
- Middleware & organisation
- Resources (supercomputers, data repositories, sensors, ...)
- Research Data Infrastructures

# Skills Sets and Capabilities needed by Researchers in Academia and Industry

B1



Research Communities

Pekka Manninen: “Skills and Human Resources for e-Infrastructures within Horizon 2020”

# e-Infrastructures skills and human resources: specific aspects

B2

1. e-Infrastructures development
  - Context dependent needs for specific skills
  - Skills and practice to communicate with scientific users
2. e-Infrastructures service provision
  - Skills for understanding service context
  - Low appreciation of the job profile
  - Hardware and Telecoms Skills
3. ***Scientific usage of e-Infrastructures***
  - ***Computation***
  - ***Simulation***
  - ***Data Analytics skills***
  - ***Visualisation***
  - ***Software skills: software engineering***
4. ***Changes in research institutions***
  - ***Need for institutional leadership***
  - ***Need to support professional development***
5. ***New careers***
  - ***Professional recognition***

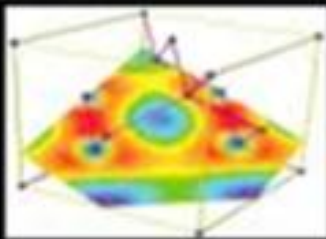
## I. Experiment



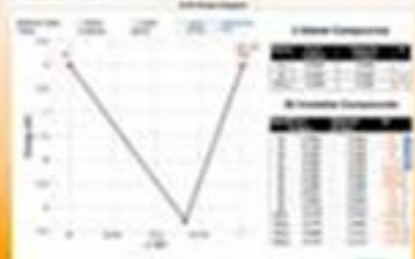
## II. Theory

$$H(t)|\psi(t)\rangle = i\hbar \frac{d}{dt} |\psi(t)\rangle$$

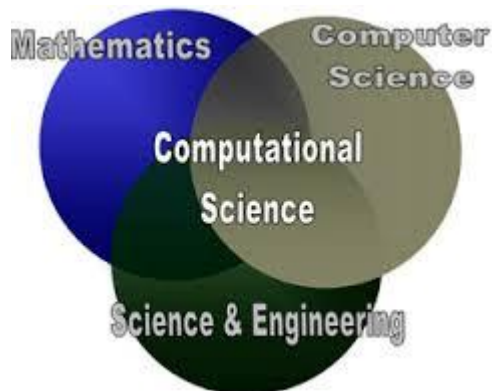
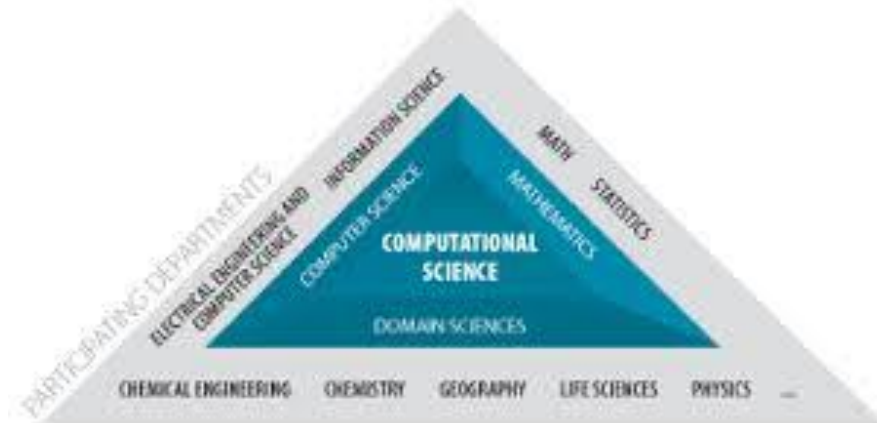
## III. Simulation



## IV. Data-intensive



COMPUTATIONAL SCIENCE: INHERENTLY INTERDISCIPLINARY



- App Maths → Comp & App Math
- Computational Sciences
- Faculty & Institutional lethargy
- Curriculum renewal ...



## 2. XSEDE, USA



- Succeeding TeraGrid
- An integrating e-Infrastructure for digital resources in US
- \$121m, 5 years
- *Education & Outreach Blog*
- *Curriculum and Educator Programs*
- *Campus Bridging*
- *Campus Champions*
- *Training*
- *XSEDE Scholars Program*
- *Student Engagement*
- *Speakers Bureau*
- *Educational Resources*
- *Campus Champions Fellows Program*

6/3/2014

Prace Training Portal: Events



- [Training Courses](#)
- [Tutorials](#)
- [Material](#)
- [News](#)
- [Blog](#)
- [Sign up](#)
- [Login](#)

Training events are advertised on the following pages:

- [PRACE Training Events](#)
- [PRACE Partners' Training Events](#)
- [Past Events](#)

[Access the PRACE Best Practice Guides](#)

[Visit the XSEDE Training Portal](#)



**XSEDE**  
Extreme Science and Engineering  
Discovery Environment



## Services for researchers



CSC provides modeling, computing, and information services for universities, polytechnics, research institutions and industrial companies. The expert services in the field of science are meant for the academic research community as a whole.

Researchers can use the largest collection of scientific software and databases in Finland through Funet network.

### Courses and events

- 12.06.2014 Seminar at HY/Meilahti: How to utilize new CSC computing resources in your research
- 16.06.2014 - Elmer FEM Course
- 17.06.2014 Elmer Coding Day
- 18.06.2014
- 30.06.2014 - CSC Summer School in High-Performance Computing 2014
- 09.07.2014

Course and event calendar

## Projects

- EUDAT
- ELIXIR
- ....

6/3/2014

BoF Groups | rd-alliance.org

## Skills Training ... 2

D2



Research Data Sharing  
without barriers

## Education and skills on Data Science

The objective of the Interest group in research data training and education is the discussion of various levels.

Home About Organisation Working and Interest Groups Plenary Meetings News & Events

6/3/2014

Digital curation training for all | Digital Curation Centre

Contact us



Search

### Digital curation training for all

Our training programme aims to equip researchers and data custodians with the skills they need to share and preserve data effectively. Ultimately, tools and approaches will evolve over time, but if all stakeholders understand the bigger picture they will be in a better position to make critical decisions that best reflect their individual needs.

With this in mind, DCC training courses make use of the curation lifecycle model as a means of contextualizing the range and nature of roles and activities required to maintain access to data over time. The DCC also encourages the transfer of knowledge and best practice among data custodians, producers and users. In this way, you will be able to share your skills and the responsibility for data curation with others on your research team.

We've developed a range of training courses that meet a variety of training needs - from the absolute beginner to the more experienced data curator.

[Request a DCC training session at your institution!](mailto:info@dcc.ac.uk) (<mailto:info@dcc.ac.uk>)

The DCC provides half-day training sessions for both researchers and research support staff. You provide the venue and the participants, we provide the trainers and the materials. [To arrange a session, please get in touch.](mailto:info@dcc.ac.uk) (<mailto:info@dcc.ac.uk>)

6/3/2014

Guides, Training, Support

> [ANDS](#) > Support

## Guides, Training, Support



### ANDS Guides:

To access ANDS Guides and other resources, please go to the [ANDS Guides](#) page.

### Training and Tutorials:

- Learn about RIF-CS from this [page](#).
- Trove and TIM and the ARDC Party Infrastructure
  - Understand what TIM is, what is Trove and how to contribute party records to Trove, and get more information about the ARDC Party Infrastructure project [here](#).
- Understand [Data Citation](#).

### Support Resources:

Useful technical [support resources](#) are listed to help contributors understand ANDS Online Services better.



**Congratulations**  
The Queensland University of Technology has completed its ANDS-funded project.



6/3/2014

Upcoming Training Sessions and Luncheons | Canadian Research Data Centre Network

About the CRDCN People Contact Us


 Search our Publications

Home Data Research Publications Events News KT Corner

Home »

## Upcoming Training Sessions and Luncheons

[Epidemiological Statistics](#)

Date: 1 May, 2014 to 31 August, 2014

Where: Online

### Archives

[2014 \(31\)](#)

[2013 \(46\)](#)

[2012 \(38\)](#)

[2011 \(8\)](#)



## GÉANT Training



In order to get the best out of the services and software developed by GÉANT, training is an essential component. Delivery of up-to-date training on GÉANT Services is a core requirement of the Project and supports knowledge transfer across Europe.

Effective training helps NRENs and their staff use the leading edge technologies provided by GÉANT to their utmost and maximises the investment made.

GÉANT offers a range of courses to assist the NRENs in their training needs. From self-paced e-learning to multi-day off-site courses, there is a training solution to meet your needs and to help users understand, install, configure, use and troubleshoot GÉANT software and services.

### GÉANT Latest News

04 June | 2014  
eduroam – beyond the campus

28 May | 2014  
The future of healthcare is in big data

22 May | 2014  
TERENA and the GÉANT project announce 2014 Community Award winners

22 May | 2014  
EUMETSAT and GÉANT: ensuring delivery of critical data

[Subscribe to GÉANT News](#)

## Skills Training ... 3

### TERENA Training

#### Overview

As part of the GÉANT project, TERENA is involved in kshops aimed at promoting the use of and providing the updates to some of the tools and services which have been developed under previous projects, and new tools developed within the current project.

The training courses are listed below and their respective webpages are linked.

#### TRANSITS CSIRT Training

[More details >> \[http://www.terena.org/activities/transits/\]](http://www.terena.org/activities/transits/)

#### PerfSONAR

[More details >> \[perfsonar/\]](#)

#### Security toolset



D3



e-IRG's 2020 vision for Europe needs a single “e-Infrastructure Commons” ....




*A necessary consequence of the vision ... is that the e-Infrastructure's user base needs to be expanded to meet the challenges set in Horizon 2020. An analysis of options to expand the user base of e-Infrastructure services is needed, ....*

Horizon 2020: Work Programme 2014—2015: a few that mention training are:

- Mathematics and ICT—Starting Communities
- Distributed, multidisciplinary infrastructure on Big Data and social data mining
- Pan-European High Performance Computing Infrastructure and Services
- Provision of core services across e-infrastructures
- Research and Education Networking—GEANT
- New professions and skills for e-infrastructures
- Centres of excellence for computing applications
- Network of HPC Competence Centres for SMEs



NSF-wide

**CYBERINFRASTRUCTURE TRAINING, EDUCATION, ADVANCEMENT, AND MENTORING FOR OUR 21ST CENTURY WORKFORCE (CI-TEAM) (CI-TEAM)** 

**CONTACTS**

Name  
Joan Peckham

Email  
jpeckham@nsf.gov

Phone  
(703) 292-7344

Room

D4



## SA Drivers

E1



## SKA—the BIG one

Each dish: ~160Gb/s. Design choices re data transport & processing still to be made.

Requirements: Computation in multi petaflops range (but data); Networking terabit/s range at minimum. Storage Exabytes. Infrastructure and skills challenges immense.



COURTESY OF WWW.SKA.SA.ZA

FFF1



Western Cape DEDAT

Big Data HPC Centre:

Pre Feasibility

DRAFT



BIG DATA AFRICA PROGRAMME PROPOSAL  
30 Aug 2013

Executive Summary

## More SA Drivers ...

- **Bioinformatics** a heavy users of SA e-infrastructure. Complex and highly heterogeneous requirements for software by this community.
- **High Energy Physics (CERN)**: formally collaborate on both ATLAS and ALICE experiments. Need data transport, storage, processors and people. Grid Computing.
- **Earth Observation**: large data sets, fast networks and HPC (GEOSS, SAEON).
- **Open Access Scholarly Publishing**: National and international initiatives (SA DIRISA).
- **Palaeo...** (Mrs Ples)
- **Humanities and Social Sciences**: Currently few examples where e-I has been used to support research in HSS. Projects: digital libraries, digitising key heritage datasets (e.g., Bleek archives, Rock Art resources) and HLT.
- **Climate Change** (ACCESS), **Southern Oceans**
- **Simulation**
- **Visualisation**
- **Applied Sciences**
- **Computational Chemistry / Physics / Biology/ Material Science / ...**





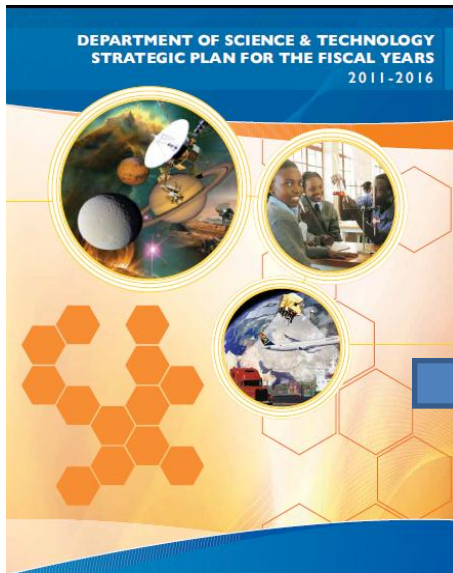
# SA National Cyberinfrastructure System



science  
& technology  
Department:  
Science and Technology  
REPUBLIC OF SOUTH AFRICA

Science Imperatives led DST into the arena

.... NICI System



**CSIR**  
our future through science



**meraka**  
INSTITUTE  
African Advanced Institute for Information  
& Communications Technology

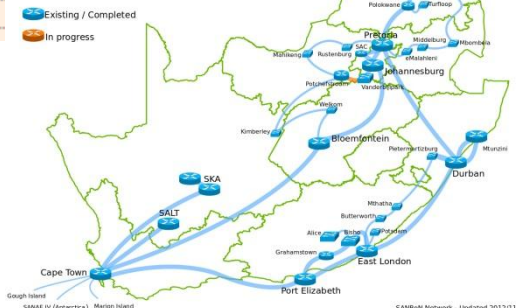
**CHPC**  
CENTRE FOR HIGH  
PERFORMANCE COMPUTING



- PART B: Programme Strategic Objectives and Activities.**
- 8. Programme 1: Administration
  - 9. Programme 2: Research, Development and Innovation
  - 10. Programme 3: International Cooperation and Resources
  - 11. Programme 4: Human Capital and Knowledge Systems
  - 12. Programme 5: Socio-Economic Partnerships
  - 13. Public entities

**SANReN**  
South African National  
Research Network

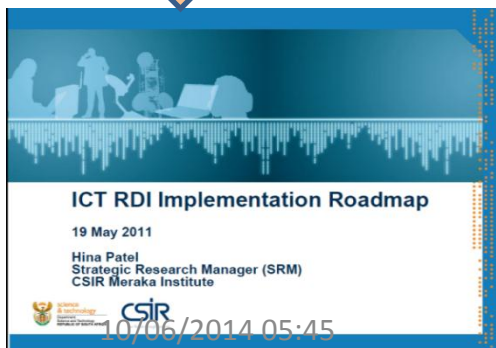
**SANReN**  
South African National  
Research Network



**SOUTH  
AFRICAN  
NATIONAL  
GRID**



**VLDB /  
DIRISA**



e-IRG 20140610

**CSIR**  
our future through science



# SANREN Backbone

F2

SANREN =

**SANREN**  
South African National  
Research Network

+

**TENET**  
Tertiary Education and  
Research Network of South Africa

 Existing / Completed

**SOUTH  
AFRICAN  
NATIONAL  
GRID**



**WACS**

Very Large data Base/Set initiative  
For e-Research Communities  
**SA-VLDB/S**  
Cedric Wright - Manager: Cyberinfrastructure  
2010/11/14  
  

**CHPC & VLDB**



**Cape Town**

Gough Island

SANAE IV (Antarctica) Marion Island

**SKA**

**SALT**

Kimberley



Alice

Bisho

Potsdam

Grahamstown

Port Elizabeth



**East London**

Butterworth

Mthatha

Pietermaritzburg

Durban

Mtunzini

Johannesburg

Vanderbijlpark

eMalahleni

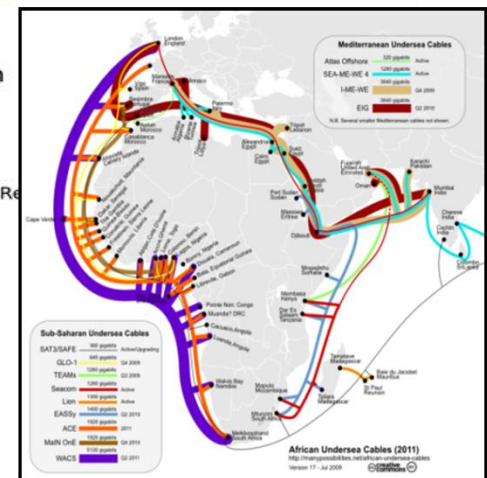
Middelburg

Mbombela

Turloop

Makhado

**Seacom**



- ~165 sites
- FETs ...
- DBE

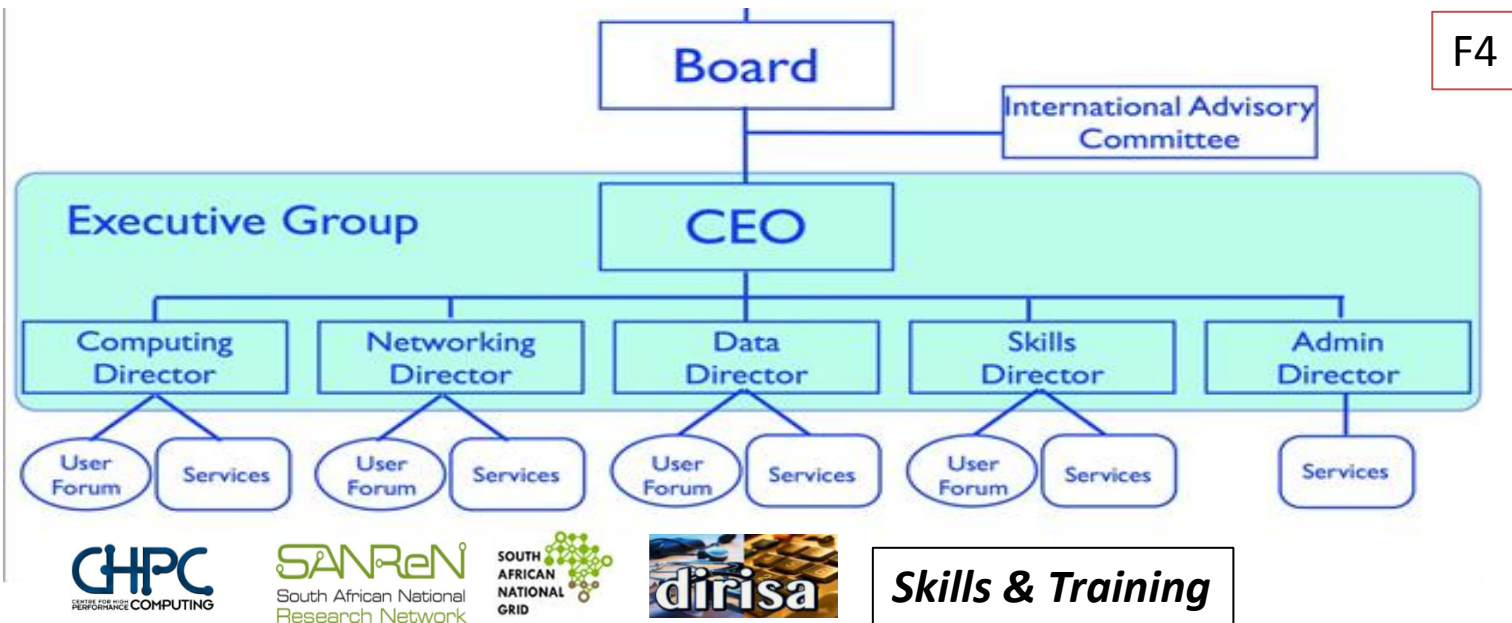
- 1Gbps /10Gbps
- SEACOM 10Gbps
- WACS 40Gbps -> 440 Gbps
- Optical fibre

## ***Internal Structure, Governance and Management***

- *Amalgamate separate organisations into a single integrated CI (e-I) organisation (NICIS) encompassing networking, compute, data.*
- ***Create a new Skills and Training entity within this new integrated organisation, with collaboration between the national organisation and the universities, particularly in the area of human skills and training.***
- *High-level of accountability of the sectors to the stakeholders (DST, universities, etc.).*

Not a “bricks and mortar” institute but rather an enabling, facilitating organisation.

# Proposal: NICIS (National Integrated Cyber-Infrastructure System)



- A national “Cyber-Infrastructure” platform
- Recognise unique features and characteristics of each sector.
- Provide bandwidth, compute cycles and data storage; through to value-added services e.g. advanced user support, data curation, cloud and grid services, FID, eduroam, training activities
- Skills and Training Service Area coordinate cross-cutting HCD theme interactively with HEIs.

A first step towards National One stop shop

# Proposal re: NICIS Skills and Training Services Area

G1

- *The shortage of e-I skills is global phenomenon.*
- *Should offer effective coordination of e-I Skills and Training services within a sustainable framework.*
- *Collaborations between the national organisation and the universities, particularly in the area of human skills and training.*
- *Role is to work with and through HEIs and RCs to grow e-I savvy cohort.*
- *Build cohort of data professionals to support research infrastructure development.*

## Examples

- eResearch
- Data Science
- Computational ....
- HPC
- Data analytics
- .....

The main targets of these services are

- Cyber-infrastructure professionals (support personnel), for developing their operational knowledge of CI.
- Researchers who use, or could benefit from using, CI services to enhance their research or collaboration capacities.
- Link to “next generation” of students in computer science who could become future CI experts and students in other disciplines who could become future users.

## Collaboration ....

- Change culture of institutions.
- Co-develop and coordinate courses with stakeholders; support community in developing educational activities, programs;
- Across disciplines and faculties. Engage with key disciplines (e.g. computer science, numerical analysis, statistics and electrical engineering departments, software eng, ...);
- Relate to the entire value-chain of knowledge creation;
- Remote education (MOOCs) in CI, standardise training material & make openly available;
- Campus Champions as liaisons to the NICIS services and local sources of knowledge (EXSEDE model);
- Organise events; road shows and ensure presence at conferences.



***Training:***

- Develop skills at multiple levels: e.g incorporate e-I training into established doctoral training programmes.
- Training module scope to include:
  - computational science, numerical algorithms, grid-computing, parallel programming, cloud computing, data-centric computing, e-science, computer animation / graphics.
  - data management needs particular attention in view of data deluge; including maintaining essential research data infrastructure; data science including analysis and visualisation, curation and long term preservation; and auditing.
- Oversight of curriculum development in area of data management in higher education institutions.
- Enable users to use e-I provided advanced services.
- Commerce & industry training modules on “commercial basis”.
- Domain specific workshops in partnership with research community.
- Ad hoc and programmatic interventions.

### Student Engagement:

- prepare them to be future researchers and educators;
- recruit nationally, with focus on under-represented groups and institutions;
- provide students with real-world research and development experience to encourage them to pursue a future career or advanced degree in digital science.

# Closure: Issues for Consideration

H1

- Training offered by amongst others: PRACE, EXSEDE, CSC, GEANT, TERENA, DCC, ANDS, RDA, CHAIN-REDS, e-Science, ... Projects. Strategic leadership: e-IRG, NSF, H2020 ..... and yet there is still much to do!
  - Culture change
  - Is existing training coherent across providers?
  - Enable optimal use of e-Infrastructure Commons
  - Dearth of Data Scientists
  - Research in increasingly more disciplines is becoming e-I dependent
  - Re-skill mid career researchers (ad hoc)
  - Appropriately skill early career researchers (programmatic)
  - Silo mentality ... collaborative model.
- Jobs are changing hence need solid disciplinary underpinning.
- Cross international boundaries—some parts of developing world have a dearth of CI / e-I skills.
- Appropriate educational skills are not all in one place.

## Issues for Consideration ... 2

H2

- e-Infrastructures → Institutions → Users.
- Enable users of T0, T1, T2 and T3 e-Infrastructures—who should train?
- Train innovative thinkers.
- Upskill users to utilise technological advances (S/W & H/W) .
- Vertical programme integration; Horizontal integration across disciplinary boundaries.
- Integrate computational training and thinking into curricula at all levels.
- Doctoral and Postdoctoral training of future e-Infrastructure leaders and users.
- Intensify engagement with users.
- Computational and Data skills.
- Monitor Performance (KPIs).

Thank You