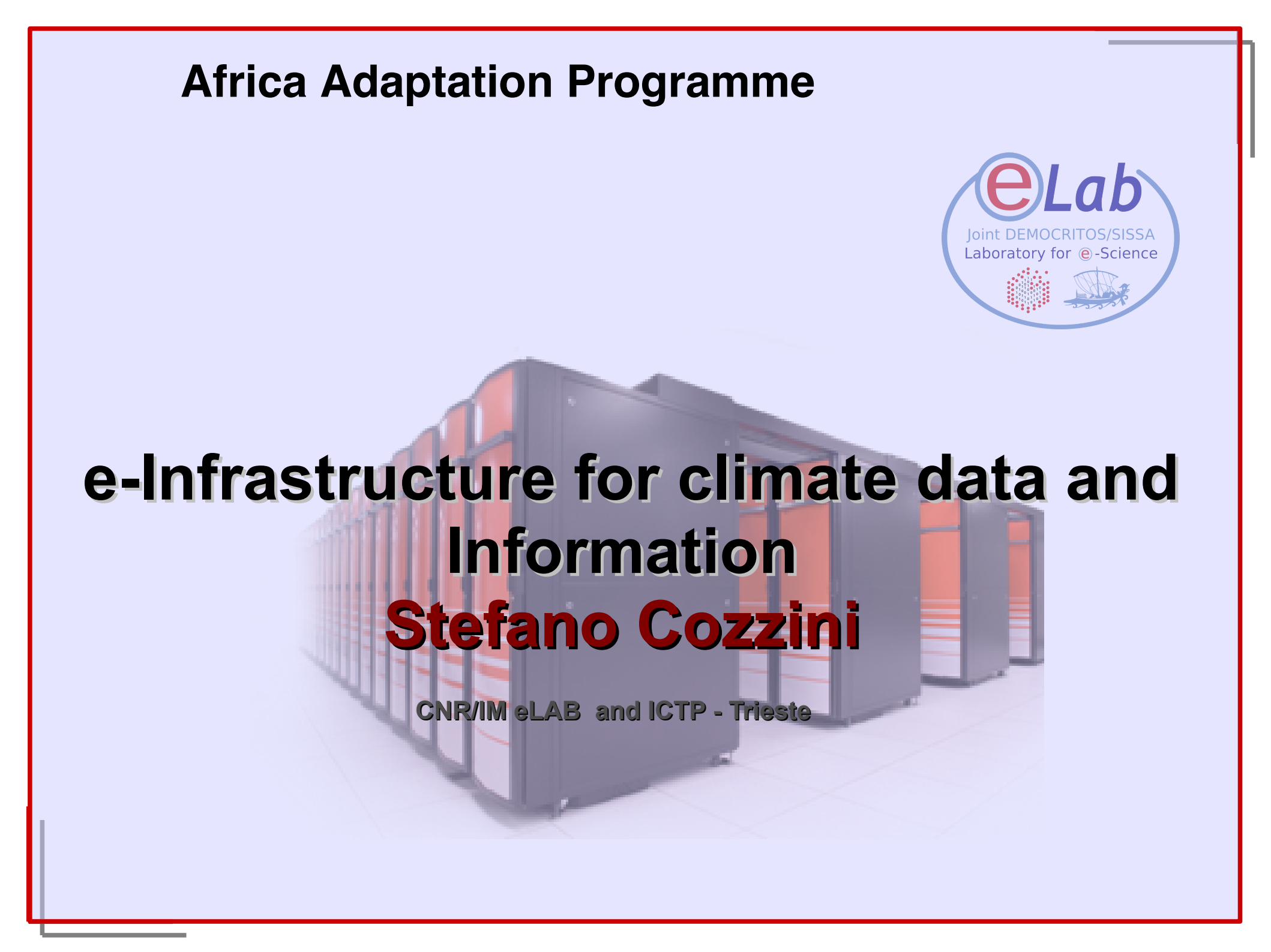
Africa Adaptation Programme



# e-Infrastructure for climate data and Information


**Stefano Cozzini**

CNR/IM eLAB and ICTP - Trieste

# Scientific Institutions in Trieste (2)

- SISSA:
  - International School of Advanced Study
  - phD courses  in different computational science areas
- Democritos National Simulation Center for CNR
  - Founded in 2002: computational material science
- ICTP:
  - International Center for Theoretical Physics
  - Science (and computational science) in developing countries

They all  need computational  e-infrastructures

# The joint laboratory for e-science: eLab

- Goal:

  provide a *computational environment* to satisfy the different requirements posed by in our institutions and beyond

- Elements of computational environment and activities
  - Platforms (hardware + system software)

    e-Infrastructure design/installation/management

  - Applications (simulation/data management software)

    Software developing/advanced services

  - People

    Training activities

Joint DEMOCRITOS/SISSA Laboratory for e-science

# ICTP activities for Africa Adaptation progamme

- ICTS section
  - Activities in setting up e-infrastructure in Africa
  - Contributions to scientific software development for climate models and data
- Earth System Physics (ESP) section:
  - It studies a wide spectrum of the Earth system, from its fluid components (oceans and the atmosphere) to the planet's interior.
  - The ESP section maintains a range of models and datasets
  - coordinates the Regional Climate research NETwork (RegCNET), encompassing over 1000 participants worldwide.
  - Maintain a Regional Climate Model (REGCM4) widely used and adopted in many african countries

# ICTP and eLab joint training activities

- Topics: scientific computing and HPC GRID Infrastructure
- More than 10 events from 2002 in Trieste for overall countries
  - More than 400 people trained
  - About 100 from African countries
- A few dedicated events to Africa:
  - Ghana 2005 ; 2 weeks training school on Linux cluster in Kumasi
  - Within the Special program for Africa:
    - Three weeks Trieste June 2009
    - 1 week Regional school in August 2010 in Addis Ababa
    - 1 week regional school in Cameroon ( 22-27 November 2010)

# ICTP eLab Past e-infrastructure initiatives in Africa : some examples

- Ghana 2005
  - Software installation of the existing HPC facility for climate modeling, Kofi Annan Centre of Excellence for ICT (KACE), Accra, Ghana

Kenya 2007

  - ICTP(OEA) donation of 4 node (16 CPU) HPC cluster including installation of software for climate prediction to ICPAC, Nairobi, Kenya.

Congo 2008

  - ICTP(OEA) donation of 1 node (8 CPU) HPC server including installation of software for material science research at the Dept. of Physics, University of Kinshasa.

## Computational Physics (HPC) Centres

2 actions:

Infrastructure Development

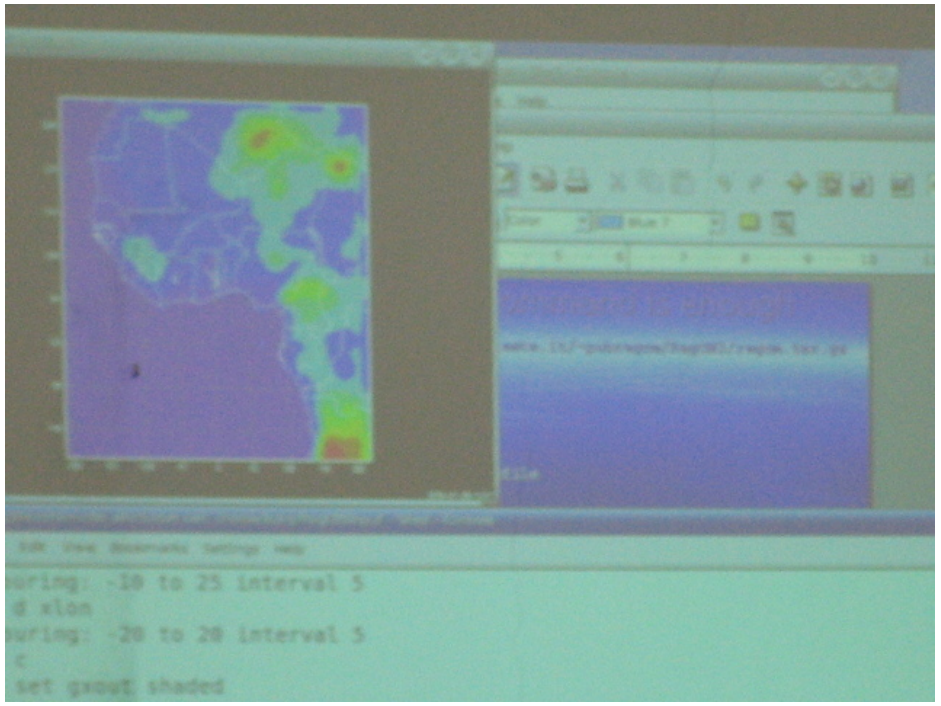
-Implementation of low cost Linux based commodity clusters for computanioally intensive scientific research

Capacity building (sustainability and self-reliance in)

_Scientific user community driving as main force to setup the infrastructure

_Application/user support for e-infrastructure

# African partners: Ivory Coast



- University of Cocody, Abidjan, Ivory Coast
  - Climate modeling HPC centre (West Africa)
  - Climate modeling HPC application support.
  - Infrastructure development – 20K *planned for 2011*

# African partners : Ethiopia

- Addis Ababa University, Addis Ababa, Ethiopia
  - Climate modeling HPC centre (Eastern Africa)
  - Climate modeling HPC application support.
  - Infrastructure development: 25K - *done October 2009*
  - 1 week meeting of African installation team in collaboration with a 1 week meeting on Climate modeling.

Second  part:

Joint activity to setup a multi-tier data infrastructure for AAP

## Guidelines

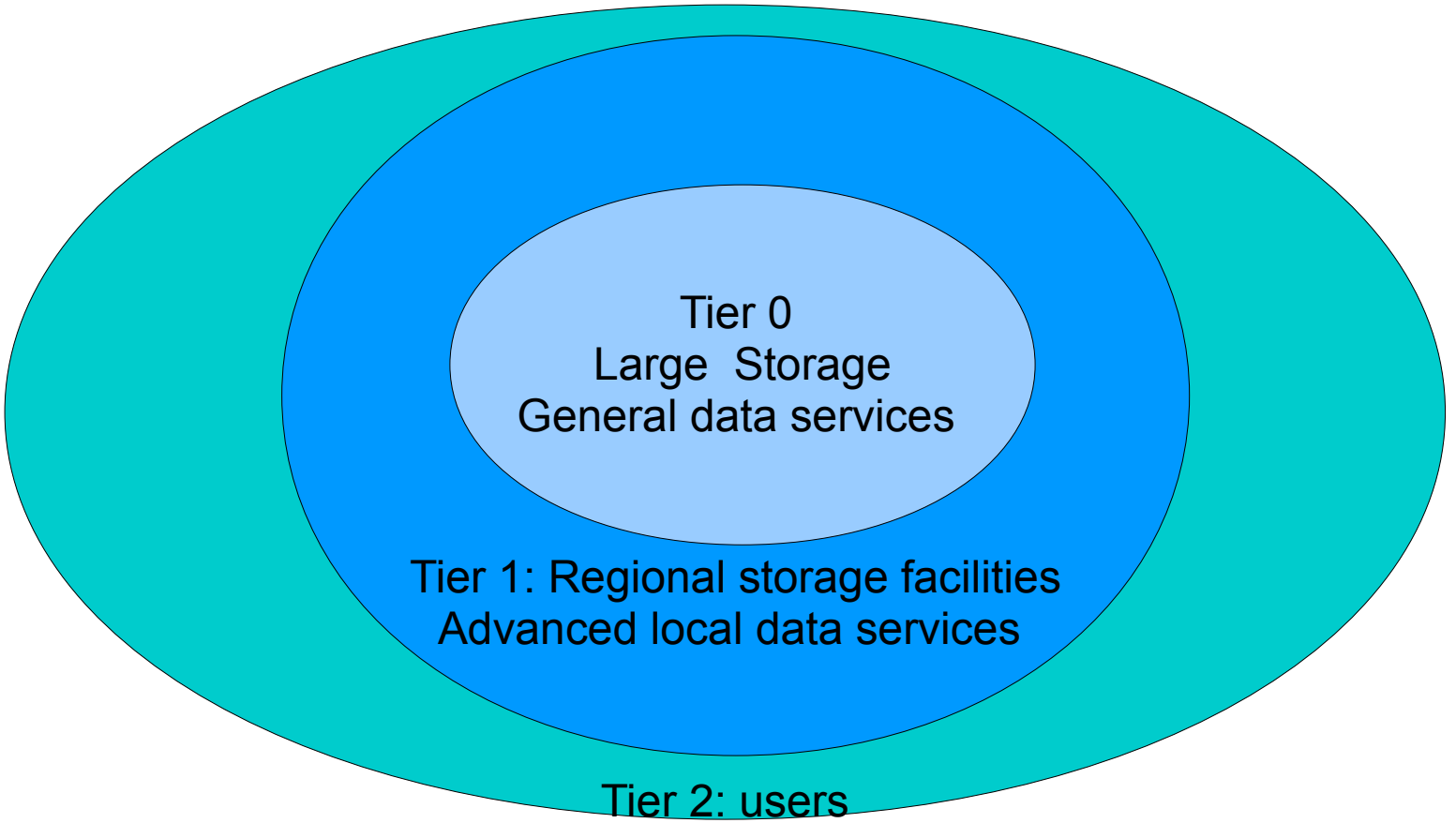
- Identify the best technical solution to solve user requirements
- Leverage on previous success stories:
  - Exploit Technical solution already in place
  - Exploit the same deployment strategy
- Team up with organization/institutions that share the similar common requirement to share effort and ideas

  ( Do not re-invent the wheel  approach )

## Concepts:

- The data infrastructure should be distributed and easily available to all users
- The data infrastructure  which would house the datasets would also have the capability of providing computationally-intensive services for data analysis and visualization
- The data analysis and visualization component would be provided through a web portal with a collection of Open source  tools, data and methods (IDV, Google Earth, TRENDS, NSFM, OpenGIS, R, RClimdex, netCDF, etc)

# Multi-tier infrastructure for climate data management in AAP countries



# Some details of multi-tier infrastructure

**Tier 0:**

The central node large enough to host considerable amount of data

Hosted by a ICTP:

large user community there with a quite common need

e-Infrastructure expertise in place

**Tier 1:**

African regional centers ( 5/6)

Large enough to guarantee the all together provide the same amount of data of tier 0

Hosted by AAP countries where competences and needs and capability are already in place

**Tier 2:**

All AAP countries

Basic usage based on web browser

Advanced/ or “ad hoc” usage may require some light-way User Interface software

# Why a multi-tier infrastructure ?

- Key Reasons

  Climate modeling is far too expensive for countries to do it on their own

  Mission critical system based on the data repository should rely on high availability e-infrastructure

  Economics of scale:

  - Cheaper and more availably to have a shared system across

- Long term reasons:

  Sustainable approach on long term basis: no single point of failures

  Opportunity for Africa to be on cutting-edge approach to address the problem of data information

## Actions

- education/training
- hardware acquisition and installation
- software deployment
- data collection/integration

## Education and training:

- 2 kinds of training activities:
  - toward sys.adm.
    - to allow installation and maintenance of the storage servers and data management software
  - toward final users
    - to make them use efficiently the infrastructure
- In both training activities we invite Unidata to mount a course on tools and methods
  - Specifically IDV and RAMMADA THREEDS

# First training events: february/march 2011

- Sys. Adm. training: 1 week in Trieste
  - 12/15 people (2/3 people team from tier 1 centers) + lecturers
  - january/february 2011
- Early adopters training:
  - 1 week in Trieste after the previous one
  - 5/10 people + lecturers (shared with previous activities)
- Dates march 2011 (close to the Cordex meeting)
- hosting Institution ICTP

## Second round training events:

- installation events in tier 1 locations:
  - each sys. adm. team will proceed in installing and configuring in its location the data server
  - installation is done in collaboration with an expert
- tutorial events for local people conducted by early adopters
  - materials and tools discussed in event 1 will be proposed to local users
- dates: TO Be DEFINED
- duration one week each:
  - two days for installation
  - three days for tutorials

# Hardware&Software activities

- Tier-zero hardware
  - expansion/allocation  of storage capabilities at ICTP
  - Target dimension : 100 Terabyte
  - Installation and Customization of specific tools for data management
- Tier-one hardware  for 5/6 site
  - acquisition of two redundant data servers
  - Target dimension: 20 Terabyte
  - Installation on tier 1 sites and associated training/user support
- Software development of lightweight user interfaces for tier 2

# Data acquisition and integration

- Sources:
  - Climate data:
    - IPCC data sets / Cordex Experiments/ Amma project/ NCEP (USA) / ECMWF Data/ CRU datasets / UK metoffice / AAP countries DATASET
  - Social Economic data:
    - Population data / livelihoods living standard
  - Hazard data
    - Floods/Drought/Storms etc..
- Data integration procedures should be put in place
  - Formidable task

# What is data integration ?

- For applications where there are a number of data sources (recall previous slide)
  - Geographically distributed
  - Having data on different platforms
  - (may be) on systems with different query capabilities (e.g., different DBMSs, files, spreadsheets)
  - Perhaps even having different data models
  - Having different schema
  - BUT about the common, general theme of climate change
- We plan to construct a general-purpose information system
  - all these data sources can be co-accessed as if they belong to a single data source
  - It can produce "combined information objects" on-demand for ad hoc queries to facilitate climate change analyses performed through other software products (workflows, atlases, statistical packages …)
  - It is integrated with AAP knowledge management system and other components

# Conclusions

- Lesson learned:
  - No infrastructure if there is no need of it
- Sustainability through user need
- Many projects shared common need
  - Successful experiences already in place
  - Room for fruitful collaboration
- Data integration requires a considerable amount of efforts
  - We need to strongly integrate data e-infrastructure with knowledge management system