

# Introduction to GSDC

(Role Expansion to National Data Center)



**G**lobal **S**cience experimental  
**D**ata hub **C**enter



October 10, 2017  
Seo-Young Noh

- 1. Data & Infra-driven R&D Era**
- 2. Data Infrastructure: KISTI-GSDC**
- 3. Role Expansion to National Data Center**

---

# Data & Infra-driven R&D Era

# Research Paradigm Shift

## Data & Infrastructure are Key in Scientific Discovery

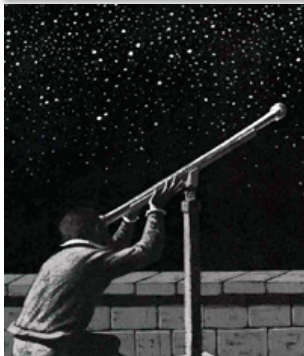
Describing natural phenomena based on **Observation**

**Modeling** and **Theory**

**Computing** and **Simulation**

**Data Analysis** of tremendous data produced from large experimental facilities

Research Paradigm Shift to Data Intensive Scientific Discovery



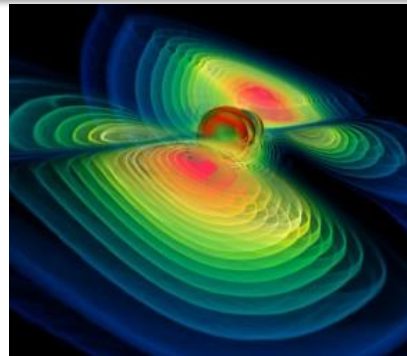
1<sup>st</sup> Generation:  
**Observation**

Galileo's telescope



2<sup>nd</sup> Generation:  
**Theory**

Higgs Theory



3<sup>rd</sup> Generation:  
**Simulation**

Black Hole Simulation



4<sup>th</sup> Generation:  
**Data**

CERN's CMS and ATLAS experiments  
→ Higgs discovery

More chance to do research with advanced equipment, higher chance to get Nobel prize

87% of Nobel prizes have been given to researchers who produced outstanding scientific discoveries using advanced experimental equipment since 1914.

Source:  
The Fourth Paradigm



# "Open Science"...hot keyword among Policy Makers

- OECD produced [the first Open Science report](#), mainly focusing on Open Access, Open Collaboration and Open Data (2015)
- Several expert groups in GSF have been formed to build advisory policy for Open Science: [Research Infrastructure, Data Infrastructure for Open Science](#)

OECD Principles and Guidelines for Access to Research Data from Public Funding



Acknowledgment on the importance of openness

OECD Publishing

Please cite this paper as:

OECD (2015), "Making Open Science a Reality", OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris. <http://dx.doi.org/10.1787/5f529632s1-en>

OECD Science, Technology and Industry Policy Papers No. 25

Making Open Science a Reality

OECD

Open science is *more than open access to publications or data*; it includes many aspects and stages of research processes. [...]

- ... is *a broader concept* that includes
- [interoperability of scientific infrastructure](#)
  - [open and shared research methodologies](#)

- ➔ Provides [cost-effective access to digital research data from public funding](#)
- ➔ Enhances [utilizations of research data to scientific communities as well as societies](#) including corporate sectors

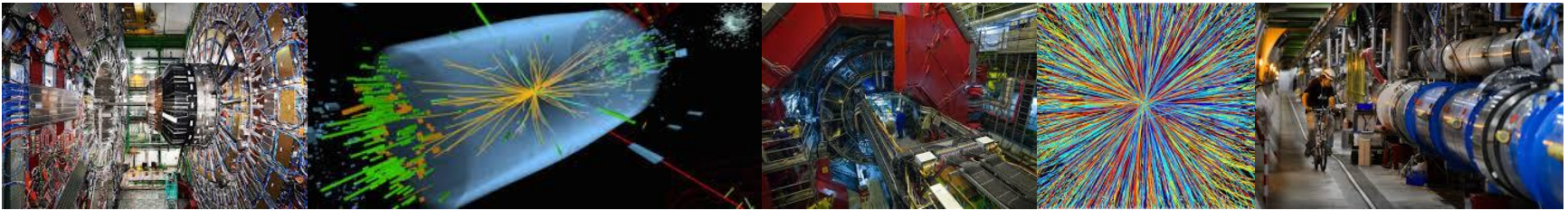


Open Data, Open Access and Open Collaboration through Information and Communication Technology

# Data Infrastructure...that is what we need

Science relies on data, requiring infrastructure for data.

Data is getting more important and growing fast.



Data Infrastructure is the one of key factors for successful science and tackling big problems of humankind.



KISTI has been in preparation for big data research era. Our mission is gradually expanding to national role for data intensive research.

---

# Data Infrastructure: KISTI-GSDC

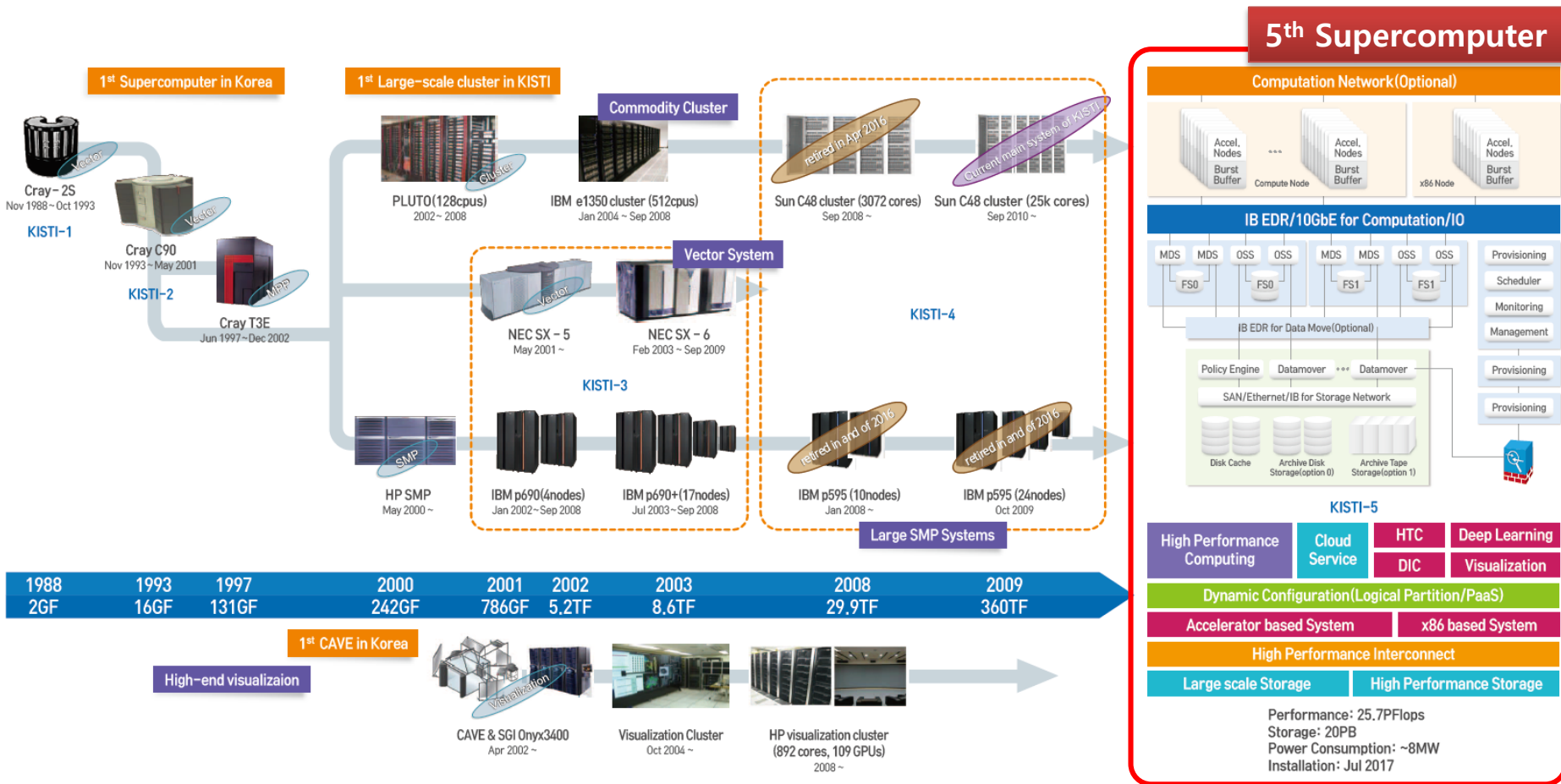
# KISTI...providing powerful ICT infra. service





# Supercomputer...not for specific, but for open to various R&D

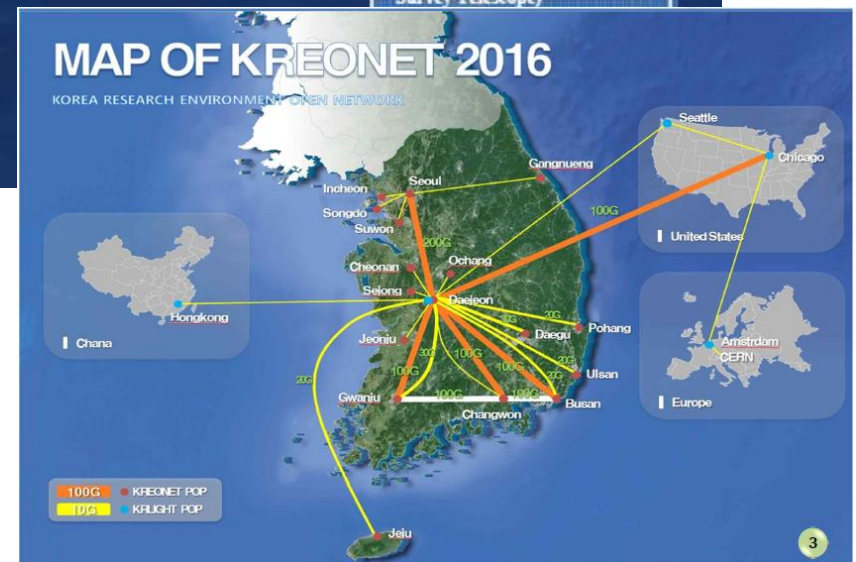
- Current supercomputer, introduced in 2008, provides 360TFlops(Rpeak)
- Supporting various R&D areas including academy and industry
- Preparation for new supercomputer having **25.7PFlops and 20PB**, targeting the 4<sup>th</sup> quarter this year(2017)

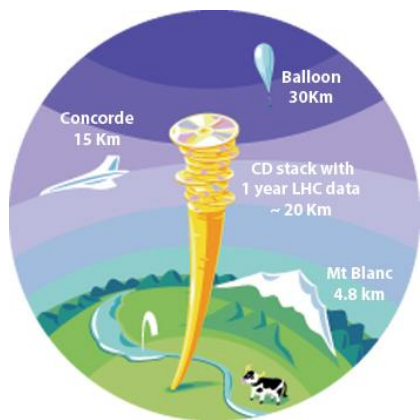


# Advanced KREONET Center...fast & secure data transmission



Providing domestic researchers with a constraint free collaborative research environment through KREONET(locally) and GLORIAD(globally)





**Large-scale Scientific Data:**  
20Km CD stack with data produced per year in CERN

# Global Science

# experimental Data

# hub Center



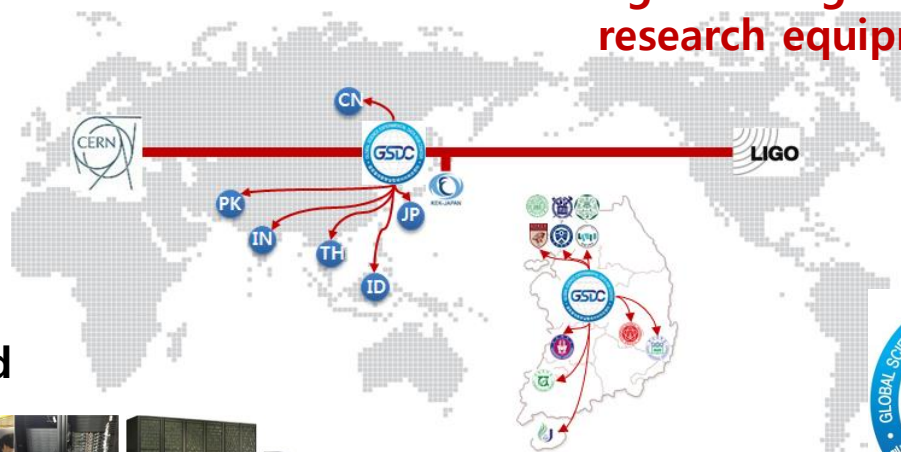
**Collaboration with global laboratories**



**Data from large and high-valued research equipment**

⇒ **(Global)**  
Contribution to global scientific research

⇒ **(Domestic)**  
Scientific data management and analysis platform service



# History



Collaborations

- ① Particle Physics
- ② Detector Construction and Exploitation
- ③ **LHC Computing Grid**
- ④ CERN's training programs and schools

Korea-CERN  
**Agreement**

Korea-CERN (LHC)  
**Protocol**

Enhancement of Grid  
Computing Support for  
large-scale research facility  
(Science & Technology Master Plan 577)

Strategy Study on  
Computing Infrastructure  
for experimental Data  
sharing

2006.10

2007.07

2008.08

2009.12

2010.07

Asia's  
Top 1  
WLCG  
Tier-1

2016

Top Quality of  
Service  
(~11<sup>th</sup> ranked)

2015.5



KISTI-CERN  
**10Gbps Network  
Established**

2014.04



WLCG Tier-1  
Approved  
(11<sup>th</sup> Nation)



Launched **Global Science**  
experimental **Data hub**  
**Center@KISTI**



# Goal and Roadmap

## National Unified Data Center for Science and National Agenda

Leap

Goal



Born

Phase

2009~2014

**Cornerstone**  
Accelerator centric Data Center (Asia hub)

2015~2018

**Growth**  
Data Center for Data Intensive Research

2019~2024

**National Unified Data Center**

2025~

Functions

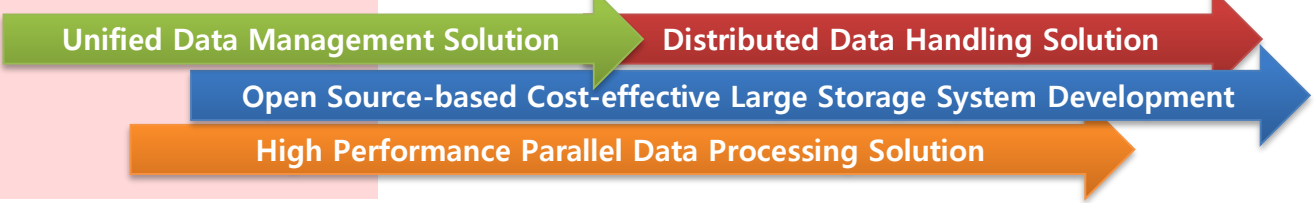
- WLCG Tier-1 Service

- Top 10 WLCG Tier-1
- **Asia representative hub**
- Pipelined service with high-valued facility

- Tailored data analysis platform service
- **Unified scientific data management service**

- National data portal for sciences
- Supporting national agenda

Technologies



# Strategy

Promotion of Data Intensive Research

**GSDC Promoting Science**

R&D Partner for World-class Scientific Achievement

Role of GSDC

National Unified Data Center for Science and National Agenda



Service  
Development

Technology

Data Mgmt.

Efficient Storage

Infra Unification

Tailored Service for Research

Edu

KiAF (ALICE)

Med/Genome

LDG (LIGO)

Climate

CMS T3

TEM

RAON

Brain

NA

Global Data Hub

WLCG Tier-1 Asia Hub

Nat'l Agenda

Collaboration



Infrastructure

Open Science Data Platform



Data

Basic Science

Medical · Bio

High-valued

Safety · Infra



# World-class CERN Tier-1 Center

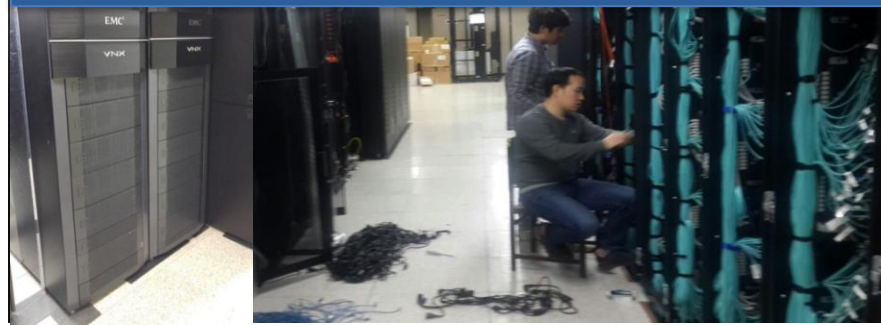
**WLCG Tier-1 officially certified in 2014** (Applied in 2012)  
Worldwide LHC Computing Grid

**Providing reliable and stable service**

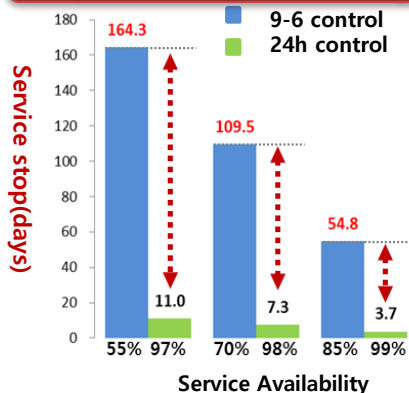
Best equipment procured every year



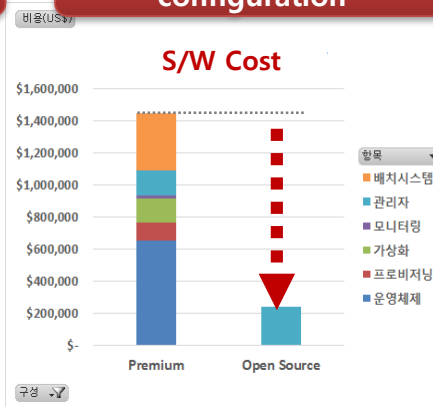
Interlocking with existing systems done by 100% KISTI experts



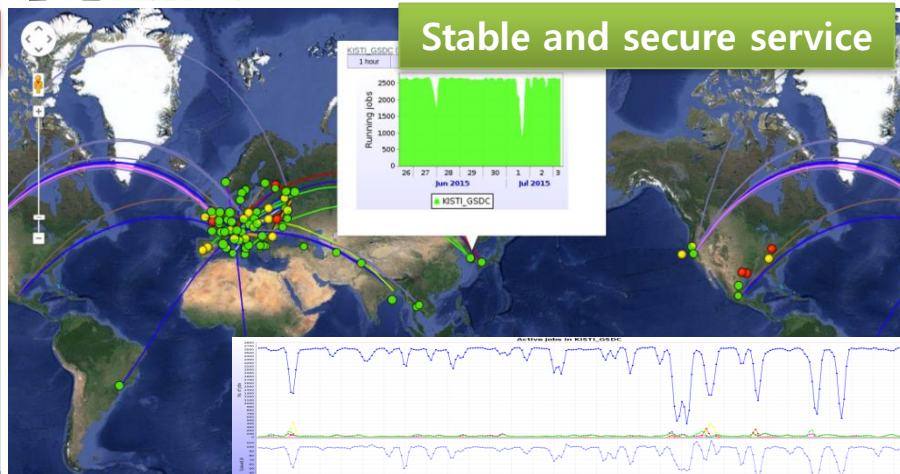
24x365 service quality control (at least 97% availability)



Cost-effective service configuration



Stable and secure service



100% open source used, requiring expertise and advanced skills [NOT FREE]

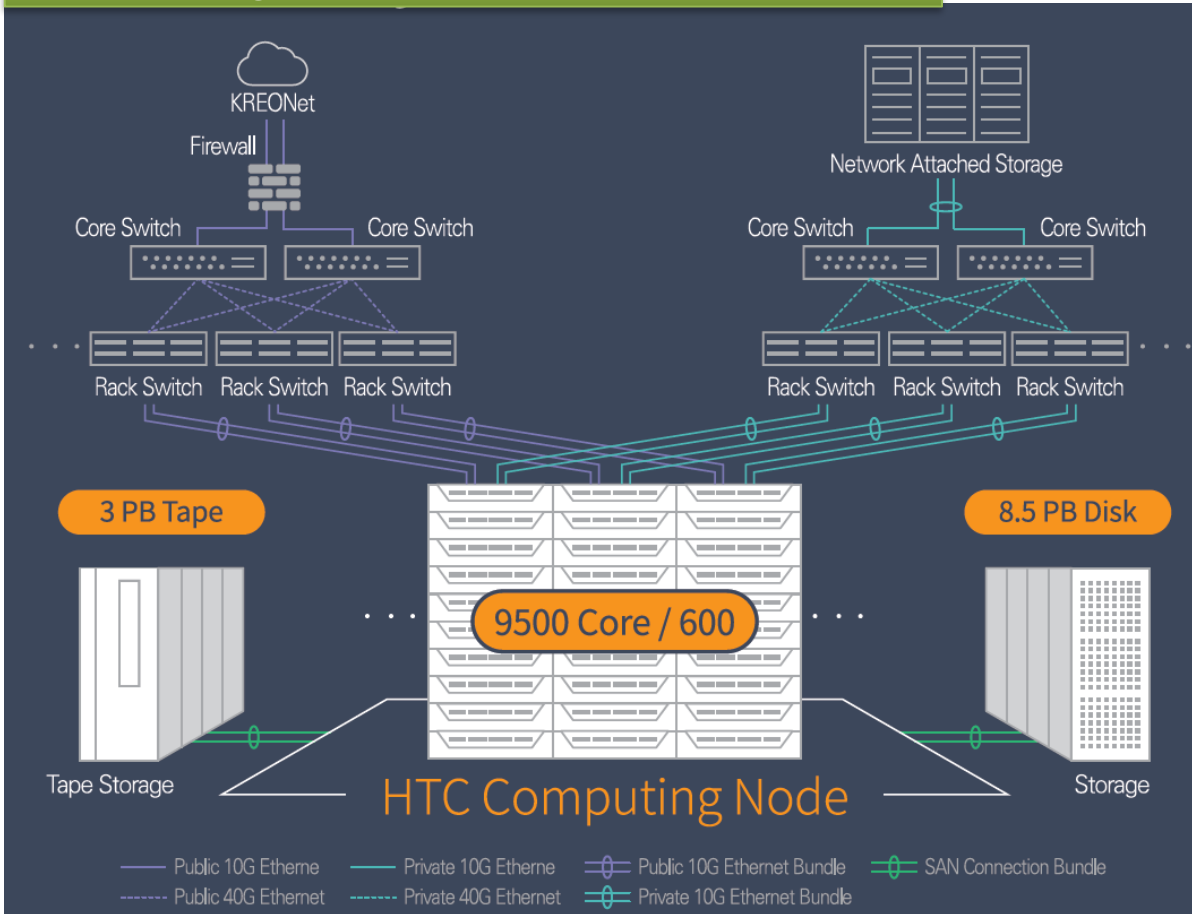
4.5 million data analysis completed per year

\* Service quality measurement based on annual WLCG report

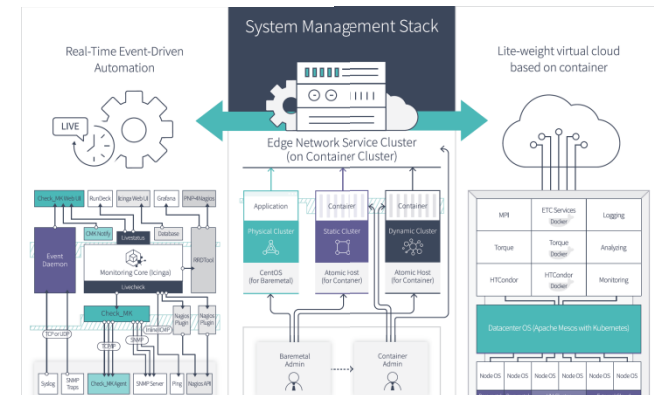
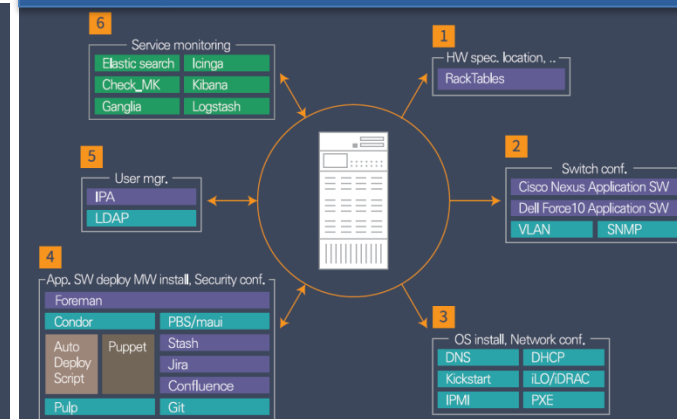
# Infrastructure @ KISTI-GSDC ... keep growing

Major vendors' competition place due to every year procurement, requiring big efforts. **It is impossible without expertise.**

## 25 storage racks with 5 different vendors



## Six steps to enable basic functions



## Automatic System Management Stack for Service

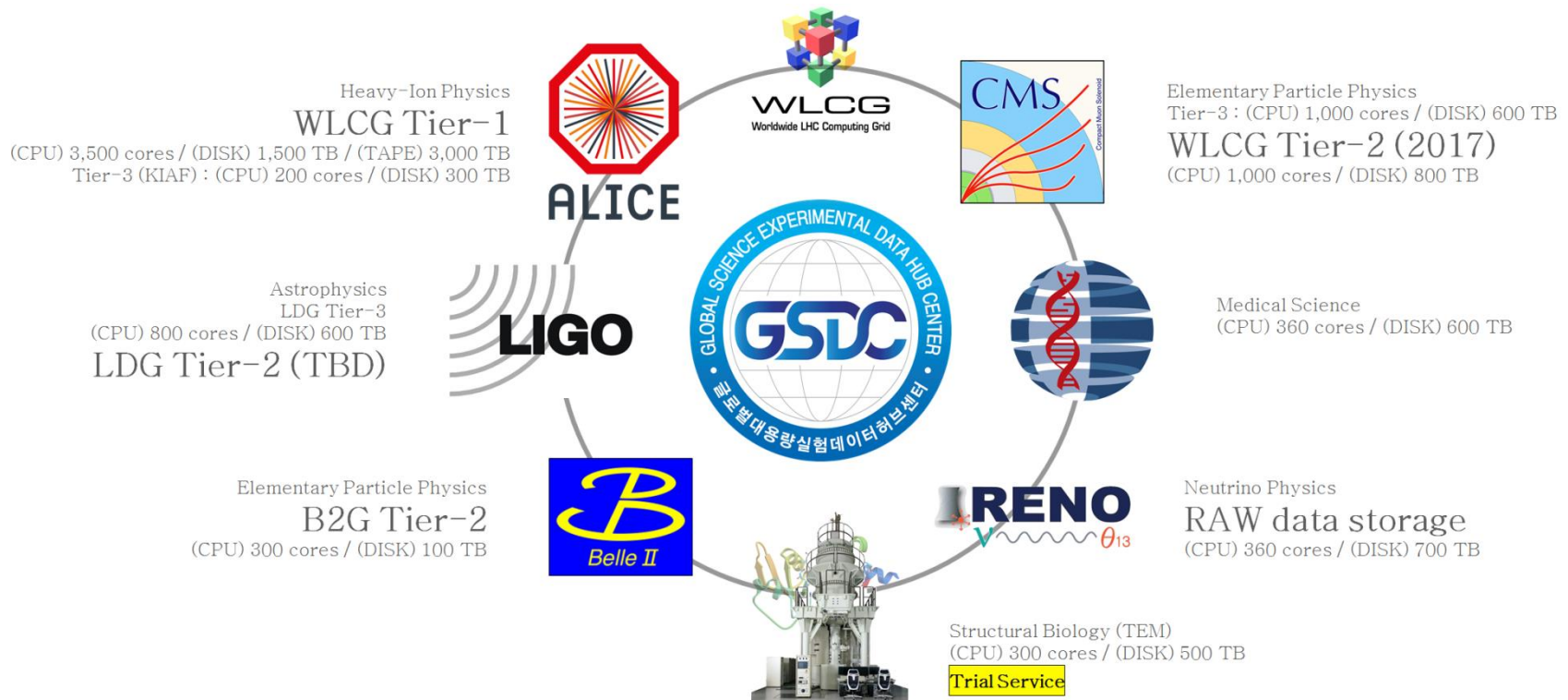


---

# **Role Expansion to National Data Center**

# Expanding to other Scientific Domains

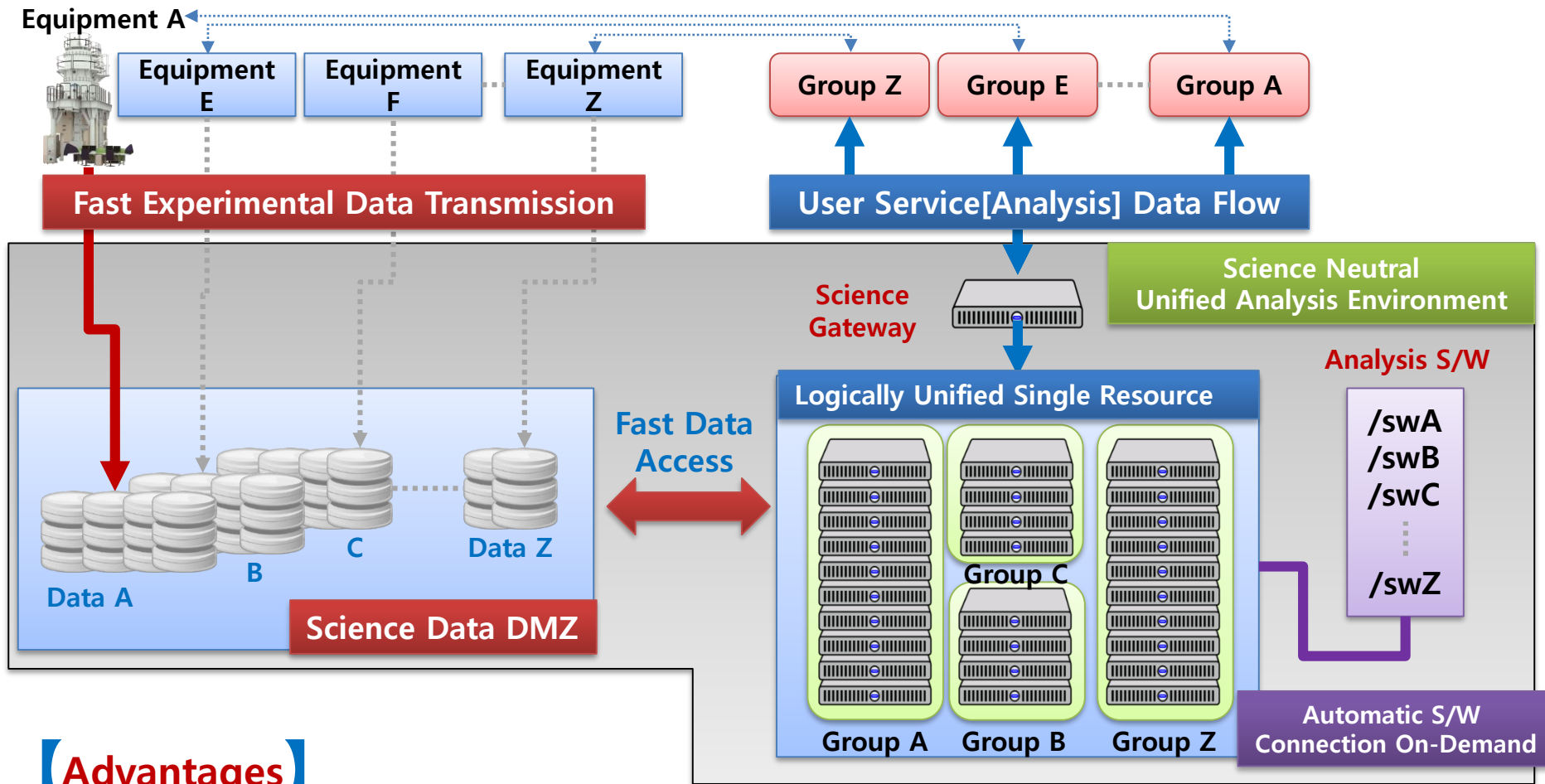
Experience on WLCG Tier-1 operation and service has given **many benefits** to expand its service availability to **other scientific domains in Korea**



and it is still expanding to many other research areas.

**Service for additional domestic experiments is under preparation.**

# Unified Data Analysis Platform @ KISTI-GSDC



## Advantages

1. **Pluggable Science** → Supports in unified way for various groups and equipment
2. **Data Infra. Sharing** → Reuse and full utilization of infra. saving tax-payer's money
3. **Simple R&D Process** → Fast results from data acquisition to data analysis

# Role Expansion

## Transmission electron microscopy Data Sharing·Analysis Farm

World Best

Universal Level

Total cost for large facilities (accumulated, 2014)

\$11 billion

\$7 billion



Data Gen.

KRENET

Science Gateway

Equipments

From Data Acquisition to Analysis

50% of Time Reduction



Raw Data



Computer Clusters



Org. Data 2D 3D Identification

Improving R&D efficiency by linking data center and large research facility

미래창조과학부



※ Survey and analysis on the status of national large research facilities

## Officially joined KEK Belle II Computing Grid



Inter-University Research Institute Corporation  
HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION  
KEK  
E-PS  
E-20  
E-50  
E-60  
E-80  
E-90  
E-100  
E-150  
E-200  
E-250  
E-300  
E-350  
E-400  
E-500  
E-600  
E-700  
E-800  
E-900  
E-1000

To:  
Korea Institute of Science and Technology Information  
305 Suwon-si, Gyeonggi-do, 471-040, Korea

Dear Mr. or Madam,

This is to inform you about the finalization of the agreement for the Belle II Computing Grid with the KEK High Energy Accelerator Research Organization (KEK). The Belle II Computing Grid is a large-scale computing grid for Belle II, which is a new super collider project. The data taking is scheduled to start in 2018. The Belle II Computing Grid is a large-scale computing grid for Belle II, which is a new super collider project. The data taking is scheduled to start in 2018. The Belle II Computing Grid is a large-scale computing grid for Belle II, which is a new super collider project. The data taking is scheduled to start in 2018.



Inter-University Research Institute Corporation  
HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION  
KEK  
E-PS  
E-20  
E-50  
E-60  
E-80  
E-90  
E-100  
E-150  
E-200  
E-250  
E-300  
E-350  
E-400  
E-500  
E-600  
E-700  
E-800  
E-900  
E-1000

work among Belle II Grid sites.  
Sincerely,

金子敏明

Tobias Kazuo, Ph.D.  
Computing Research Center, Head  
High Energy Accelerator Research Organization KEK

MOU  
KISTI & Belle II



2016.11.16

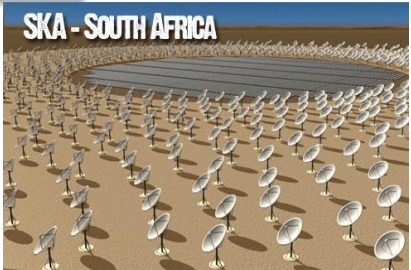
## RAON

New Accelerator in Korea



Utilization of Tier-1 know-how for data management

## SKA



Regional data center (under discussion)

## KAGRA

Gravitational Wave Detector in Japan



Officially participation in data management

## TEIN-GLORIAD-KR

Network Connection Improvement



Enhancing collaboration in Asian community

**Thank you.**