

3rd High-Level Industry-Science-Government Dialogue on 'Atlantic Interactions': Implementing the Air Centre Cabo Verde 7 - 8 May, 2018

Augusto Gadelha

Diretor

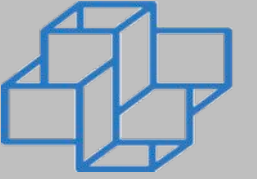
Laboratório Nacional de Computação Científica

MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA,
INOVAÇÕES E COMUNICAÇÕES

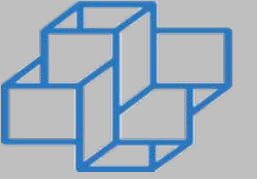


Laboratório
Nacional de
Computação
Científica

National Laboratory of Scientific Computing (LNCC)

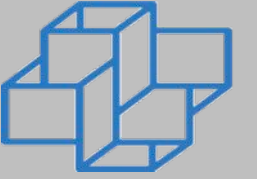


Petropolis, Rio de Janeiro



LNCC – MCTIC

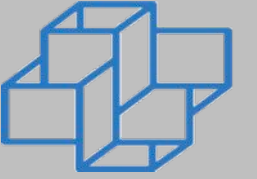
- R & D: Main Research Lines
 - Computational Modelling
 - Numerical Methods
 - Computer Science
 - Computational Biology
- Supercomputer Center
- Graduate Course (M.Sc. & Ph.D.) on Computational Modelling
- Coordinator of the SINAPAD – National System for HPC
- Coordinator of INCT-CID (National S&T Institute – Data Science)



Data Science in LNCC

Goals:

1. To provide infra-structure for Data Science applications
2. To develop methods and techniques for large scale data science
3. Using data to develop in-silico science (science by computer simulation)

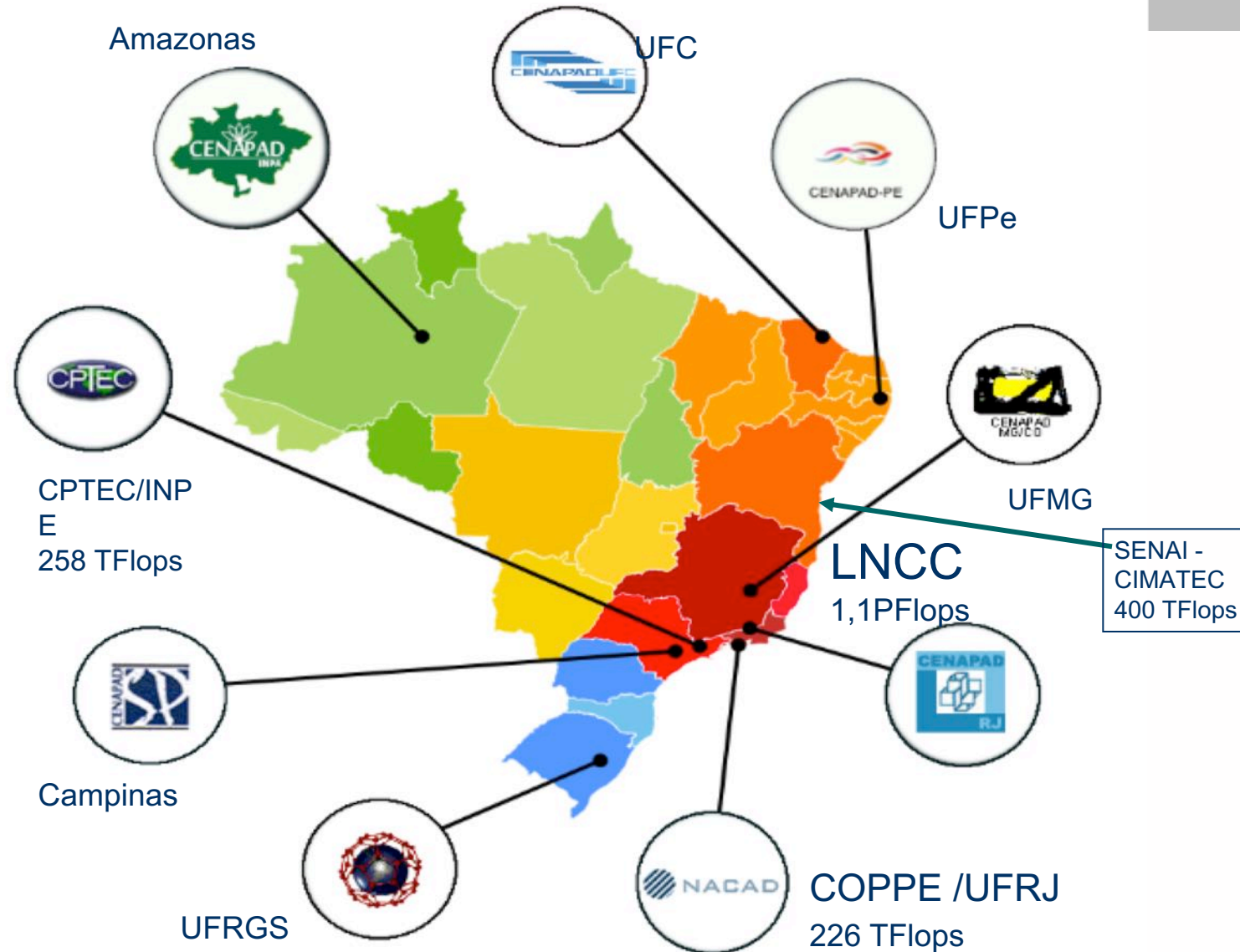
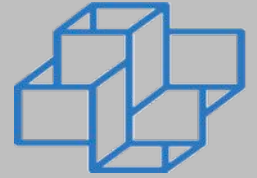


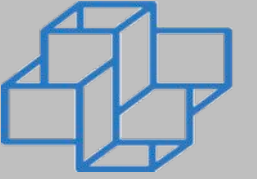
Data Science in LNCC

Infrastructure for Data Science applications

SINAPAD

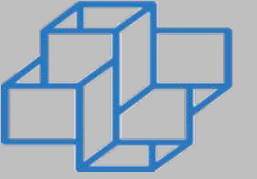
9 HPC National Centers





Santos Dumont Supercomputer - LNCC





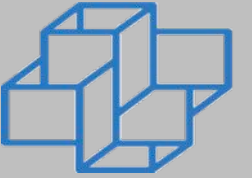
LNCC – SDUMONT Configuration

18.144 CPU cores, 756 hybrid computing nodes:

- **756 computing nodes:**
 - » 504 model B710, 64GB,
 - » 198 model B715 (2 Tesla K40 each - NVidia)
 - » 54 model B715 (2 Intel XeonPhi each)

- **1 computing Node model FAT NODE :**
 - 16 CPU INTEL IVY BRIDGE (15cores, 2.3GHZ), 6 TERABYTES RAM

LNCC – Big Data Cluster



Number Nodes	Configuration
1	768GB RAM, 8 HDs 1.2 TB, 2 Intel Xeon
1	128 GB RAM, 8 HDs 1.2 TB, 2 Intel Xeon
12	96 GB RAM, , 8 HDs 2 TB, 2 Intel Xeon
2	16 GB RAM, 1 HD 1TB, , 2 Intel Xeon

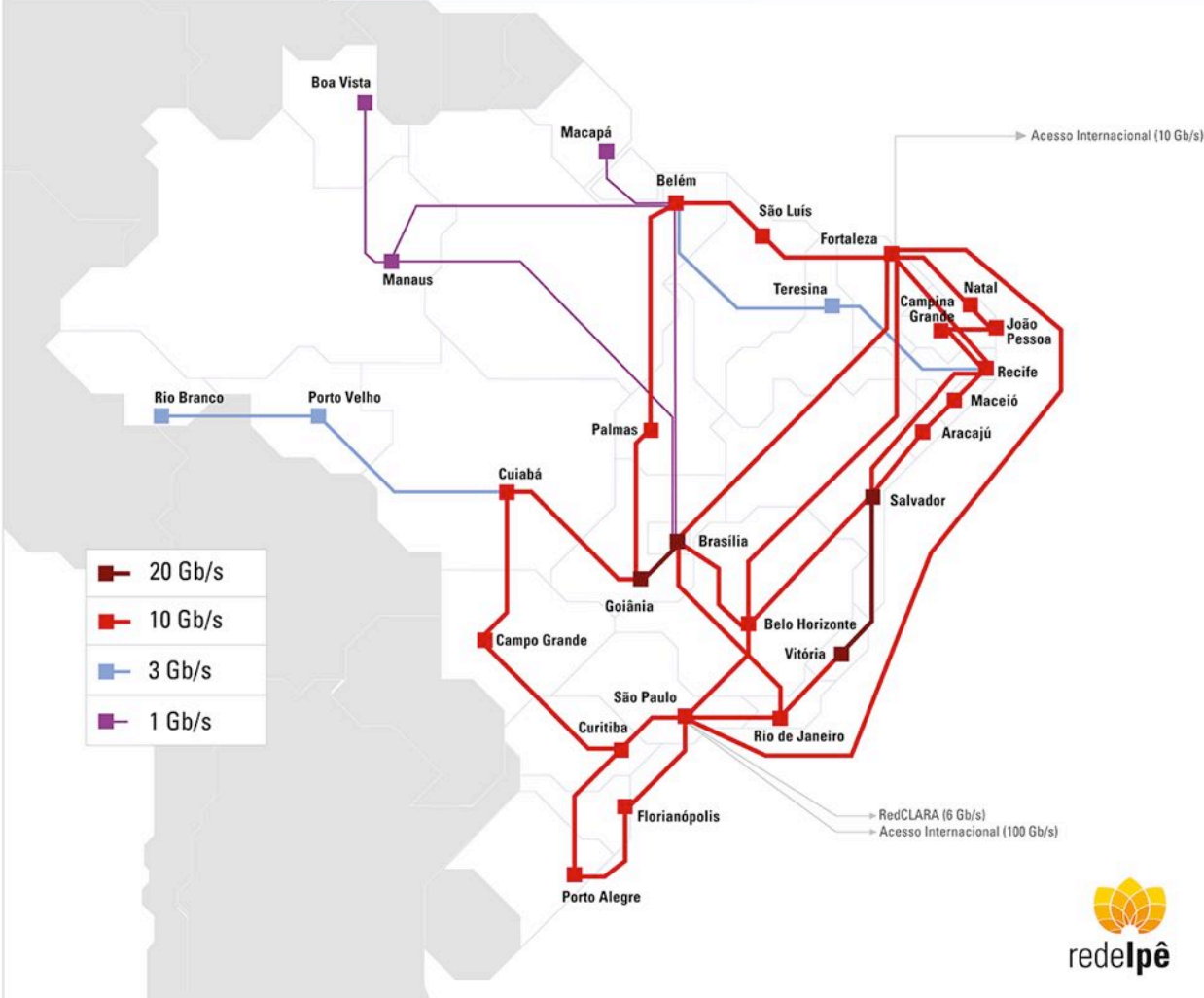


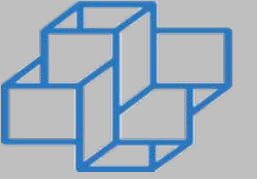
RNP – National Academic & Research Network

Conexão em 2016

capacidade agregada 347 Gb/s

capacidade internacional 116 Gb/s



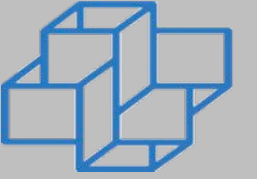


Data Science in LNCC

Methods and techniques for large scale
data science

INCT – CID – Data Science

National S&T Institute – Data Science



- Mission: to foster Data Science in Brazil

- ✓ Building expert human resources
- ✓ R & D on methods and techniques



- Strategy

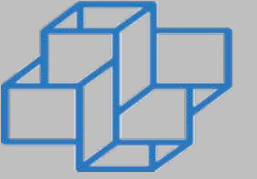
- Gathering with domain groups: health, oil, sports, medicine, astronomy, biodiversity, ecology,.....
- Building on expertise of LNCC and its partners
 - multi-disciplinary research & modeling
 - Strong capacity in Computer Science and Mathematics

- Partners (network)

- UFC, UTFPR, IME-RJ, UFJF, UFPR, CEFET-RJ

INCT – CID

National S&T Institute – Data Science



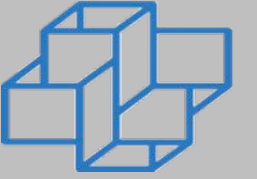
Partners in Domain Areas

- FIOCRUZ-CIDACS and ICICT - Health
- National Observatory – Astronomy
- Botanic Garden – ecology, biodiversity
- Petrobras, Repsol - Oil

Company Partners

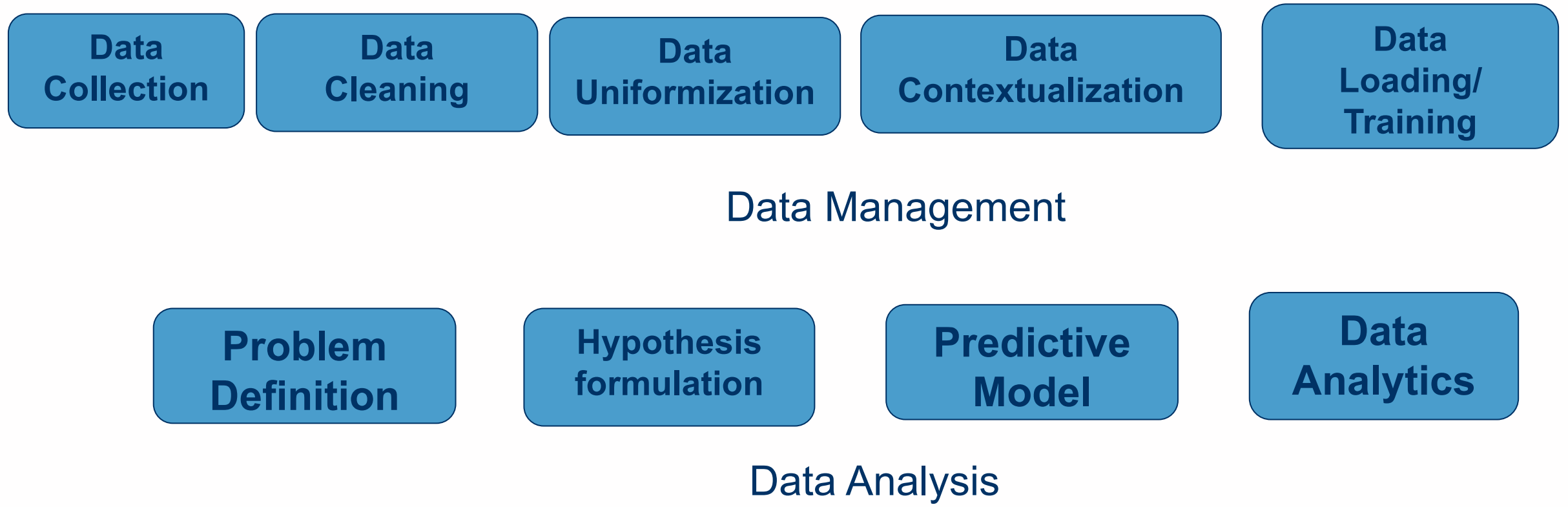
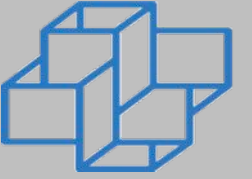
- DELL - EMC – Research Center
- IBM Research
- DBS2 (a start-up)

International: NYU, INRIA, UCSB, Boston University – Data Science Institute

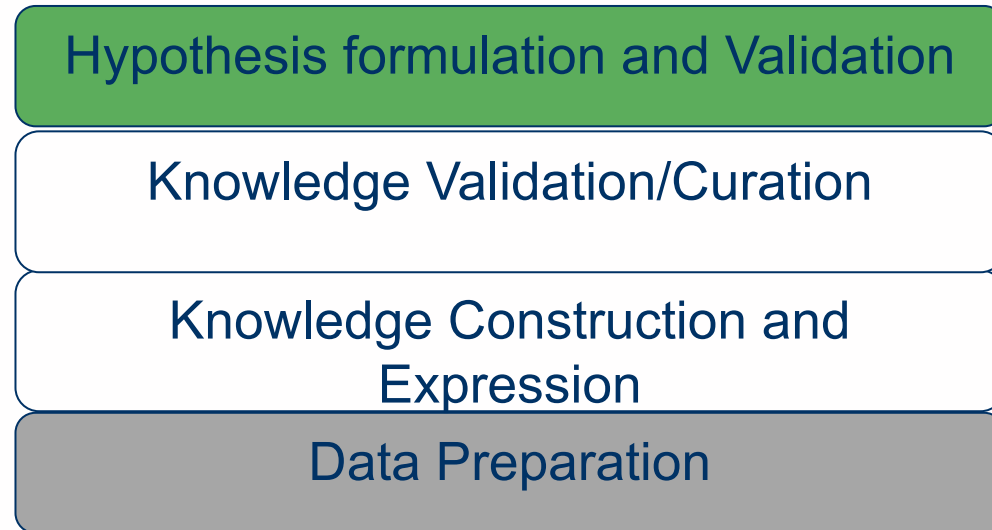
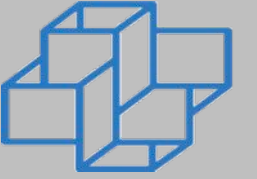


Using data to develop in-silico science
(science by computer simulation)

Data Science Process : 2 Stages

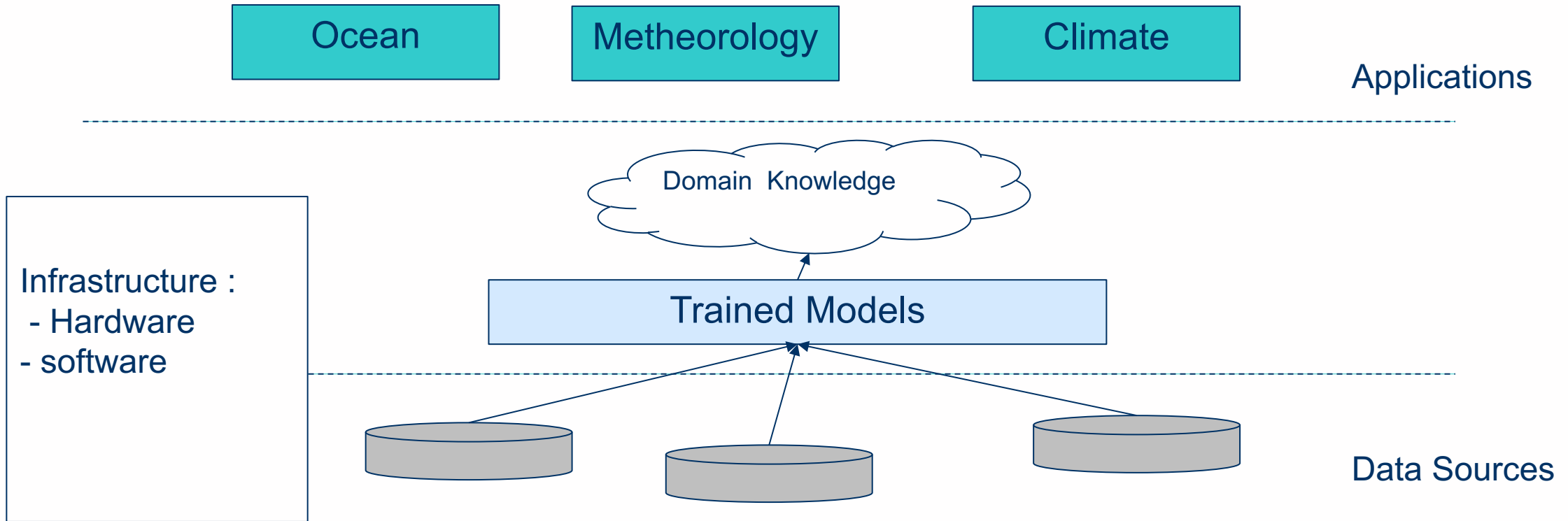
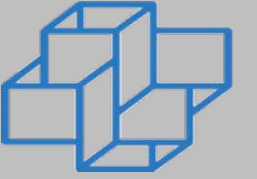


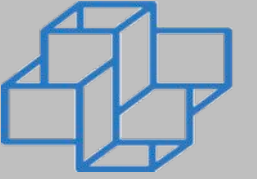
Constructing Knowledge Bases



Going up the ladder

Data Science Infrastructure and Systems





IMPLEMENTING THE AIR CENTRE

Challenges

- Extracting data (through sensors or simulation) and building up large scale data basis
- Expertise on data oriented approach to knowledge construction
- An execution environment to process petabytes of data
- Programming language for parallel computing
- Computing warranties: recovery under faults, monitoring tools, etc.
- Preparing for future HPC challenges → the exaflop scenario

**3rd High-Level Industry-Science-Government
Dialogue on 'Atlantic Interactions':
Implementing the Air Centre
Cabo Verde
7-8 May 2018**

THANK YOU

OBRIGADO