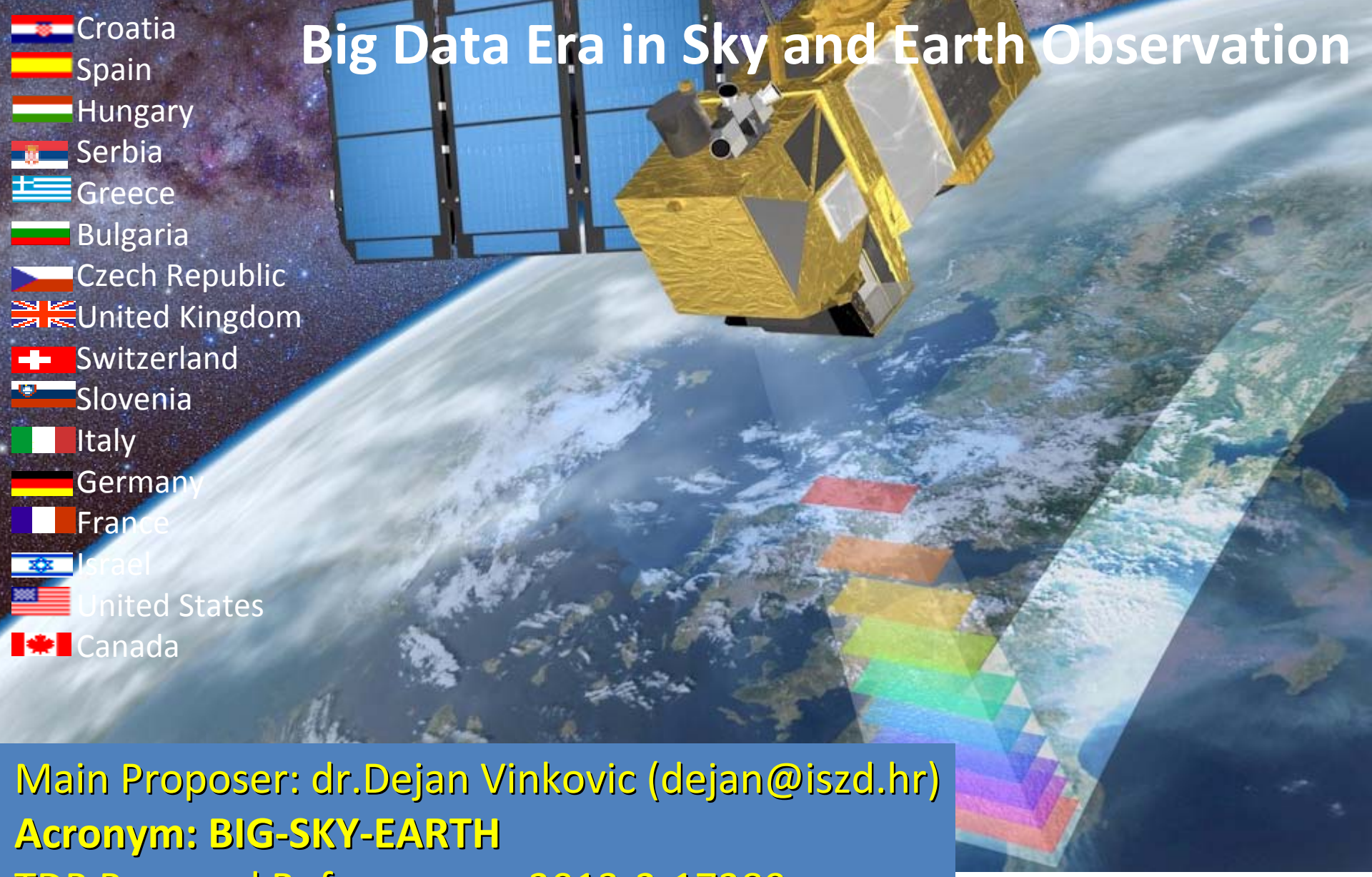


Big Data Era in Sky and Earth Observation

Main Proposer: dr.Dejan Vinkovic (dejan@iszd.hr)

Acronym: BIG-SKY-EARTH

TDP Proposal Reference oc-2013-2-17289

Impact of BIGSKYEARTH

- **We want to set the ground for a long-term networking across the earth-space domain**

- **We want to help European researchers to use the best available tools to deliver cutting edge science in the Big Data era**

sentinel-1A

Launch date: 3. April 2014

up to **2.4 TB/day** of imaging radar data for **7 years** (fully open and free data access policy)

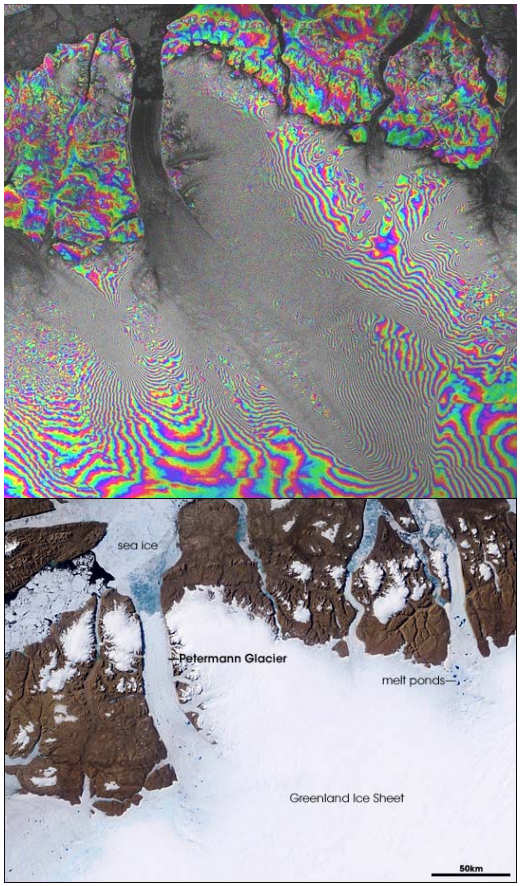
Applications: Oceans and ice, Changing lands, Emergency response

First in a constellation:

1A/B/C, **2**A/B/C, **3**A/B/C, **4**A/B, **5**P/A/B, ...

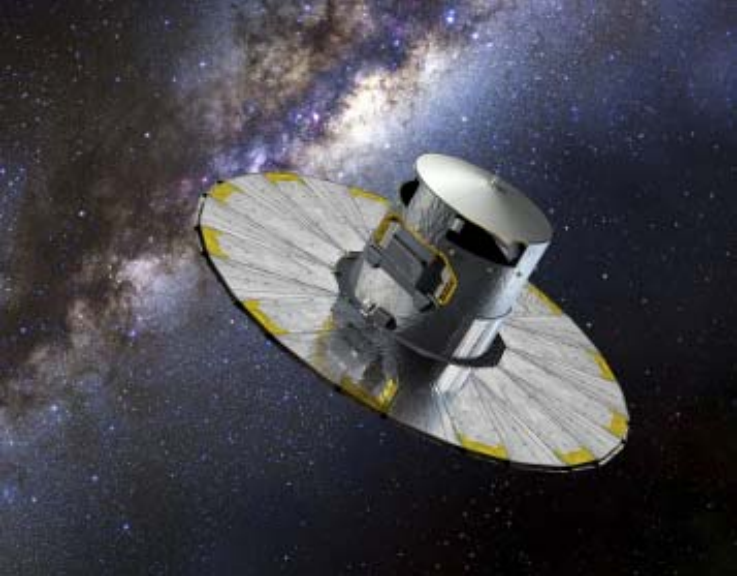
The era of Big Data has arrived!

Example: images + time = surface movements

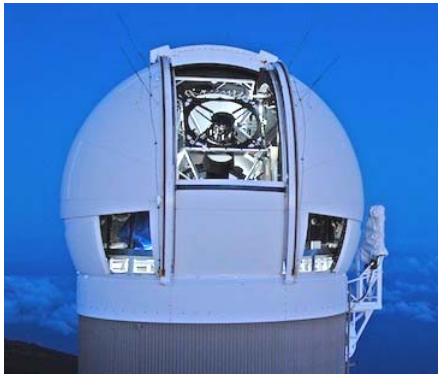


Part of the **European Earth Observation Programme Copernicus:** the most ambitious Earth observation programme to date: 30 satellites: peta-bytes now: zetta-bytes in a decade

copernicus
The European Earth Observation Programme



The era of Big Data has arrived!



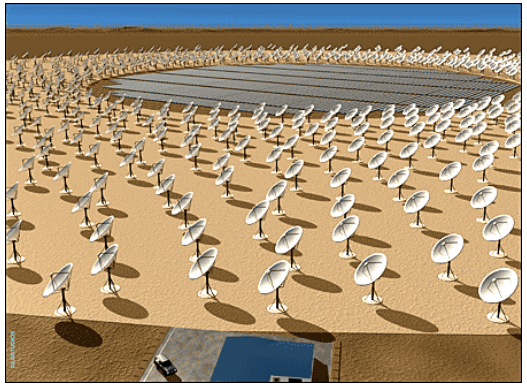
Pan-STARRS (NEO defence)
>100 TB of data

LOFAR (radio telescope):
1 petabyte per year

GAIA space telescope

Launch date: 12. December 2013
1PB of data after **5 years** (40 million observations a day!) (free data catalogue access policy)
Applications: 3D catalogue of ~1 billion astronomical objects

Large Synoptic Survey Telescope:
30TB of imaging data each night



Square Kilometre Array:
1.5 exabytes per year

IVOA

The Virtual Observatory (VO): provides standards describing all astronomical resources worldwide and supports standardized discovery and access to these collections

Shared challenges: data tsunami

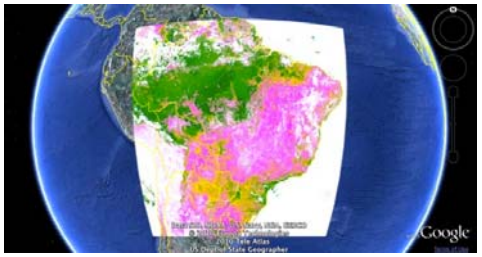
ESA's G-POD

Digital curation and data access

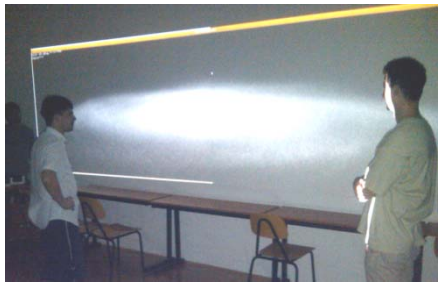
- store, maintain & preserve huge amounts of data
- large **multidimensional & highly interrelated** datasets = paradigm change: **push the computing to the data**

Visualization

- visualizing large quantities of data with: low signal-to-noise ratio, high dynamic range, multidimensional parameter space, multi-layered time-dependent, ...



Google Earth Engine



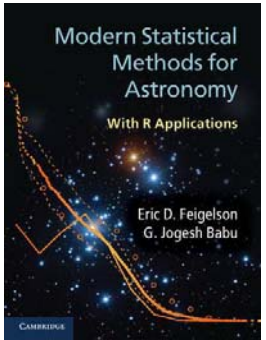
GPUs as numerical co-processors

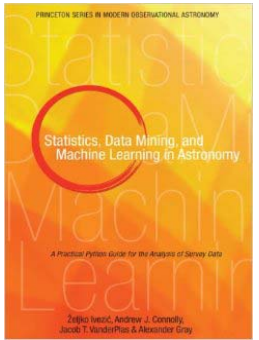
Adaptation to new high performance computing (HPC) technologies

- heterogeneous supercomputing environments
- new programming techniques for GPUs and cloud computing

Training of a new generation of scientists

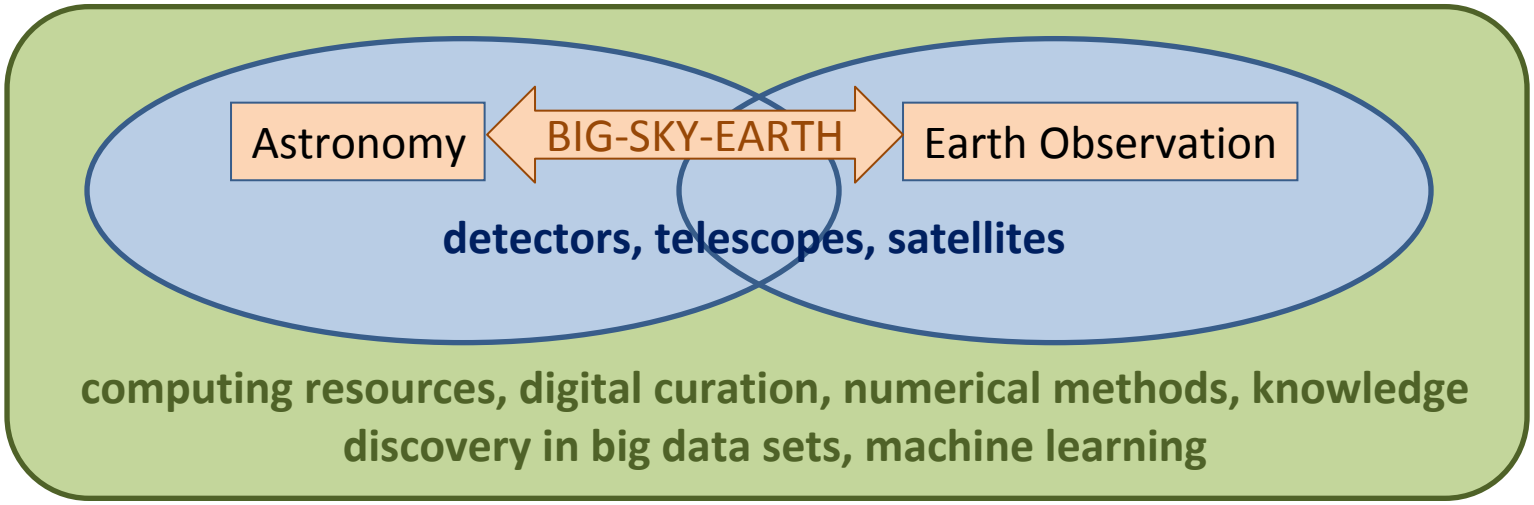
- astroinformatics, geoinformatics, bioinformatics
- natural sciences + IT/CS = exploration with statistics





New books and online courses

Why COST Action?

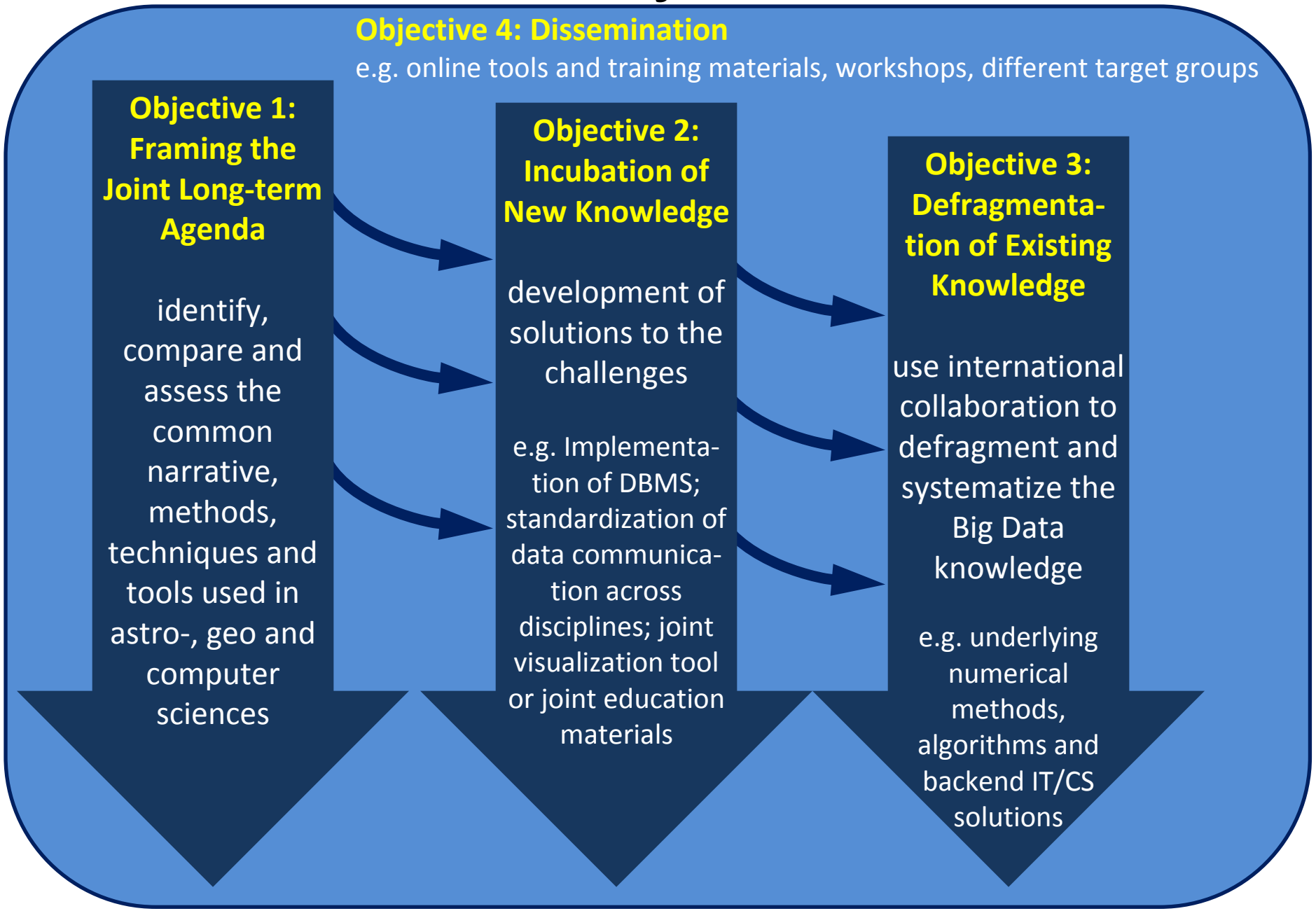


The challenges are inherently global and transdisciplinary!

- boosting the communication within and between disciplines
- identify and cluster relevant common solutions
- go beyond individual or national projects
- diversify the pool of experts
- addressing and documenting various issues in Big Data science
- spread good practices between astro- and geo- communities
- build a joint agenda and training/education resources

Set the ground for a long-term networking!

Action objectives



Management structure

WG1: Optimisation of database tools in astro- and geophysics contexts

(focused more on the back-end tools)

WG2: Data mining and machine learning in the petabyte era as frontiers in astronomy and Earth observation

(focused more on the front-end tools)

WG3: Education of a new generation of experts in knowledge extraction from massive datasets

WG4: Visualization of high dimensional data

(scientific & outreach)

Management Committee (MC)

(management and supervision)

National delegates

MC Chair & Vice-Chair

Technical Manager

Training, Dissemination and Liaison Manager

WG1,2,3,4 Scientific Chair and co-Chair

Inclusion manager

Industrial networking manager

Steering Committee

Implementation

We stand on the shoulders of:

- HPC infrastructure (adjunct to big astronomy or remote sensing facilities)
- FP6/7 (& H2020) astro-, geo- and HPC projects
- astro- & geo-informatics community
- IT/CS industry & SMEs

... and go a step beyond!

Example: industry (e.g. IBM, Google) is involved in collaboration with astronomy facilities to research extremely fast, but low-power exascale computer systems and data mining in PB databases

Implementation

We stand on the shoulders of:

- HPC infrastructure (adjunct to big astronomy or remote sensing facilities)
- FP6/7 (& H2020) astro-, geo- and HPC projects
- astro- & geo-informatics community
- IT/CS industry & SMEs

... and go a step beyond!

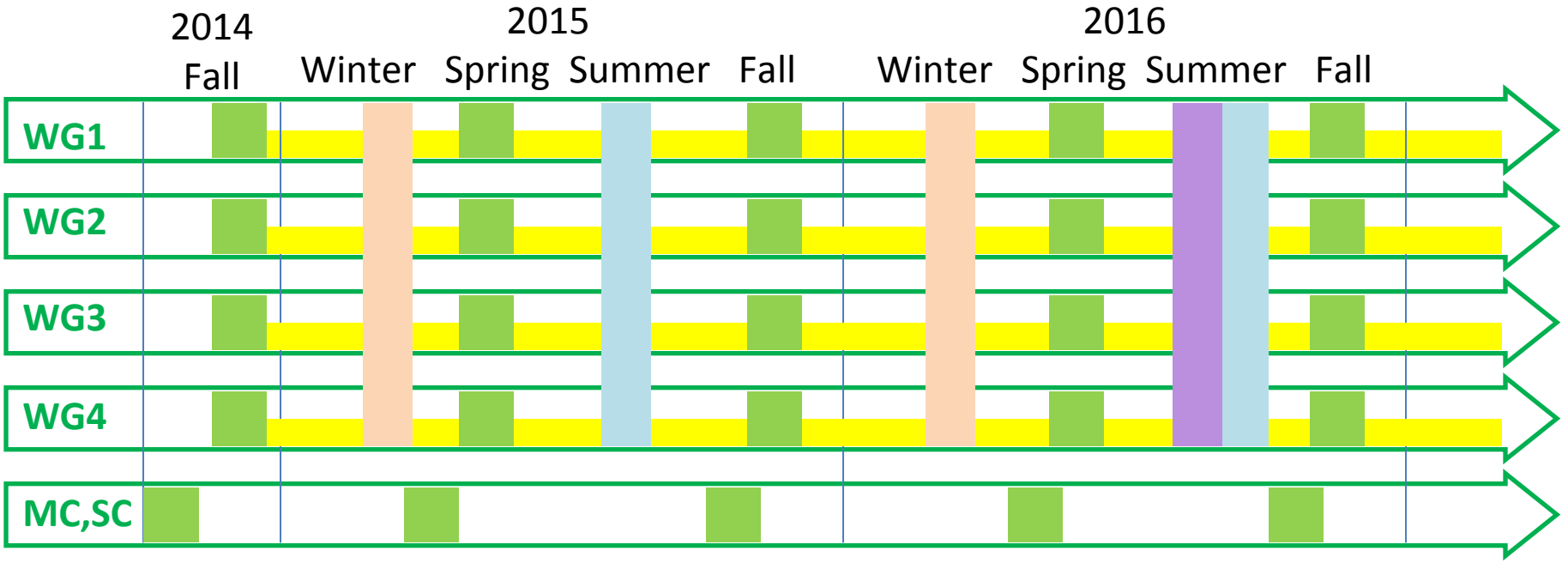
WG & MC/SC meetings

Short-Term Scientific Missions, Joint Student Supervision, Website, etc.

Workshops

Training schools

Confrence: “Big Data in the Earth and Space Sciences”



Impact

Big Data is not just bigger, it is different!

Success in research will depend on the ability to mine knowledge from that data.
And some of the most interesting science probably hasn't even been imagined!

Across the earth-space domains of Big Data:

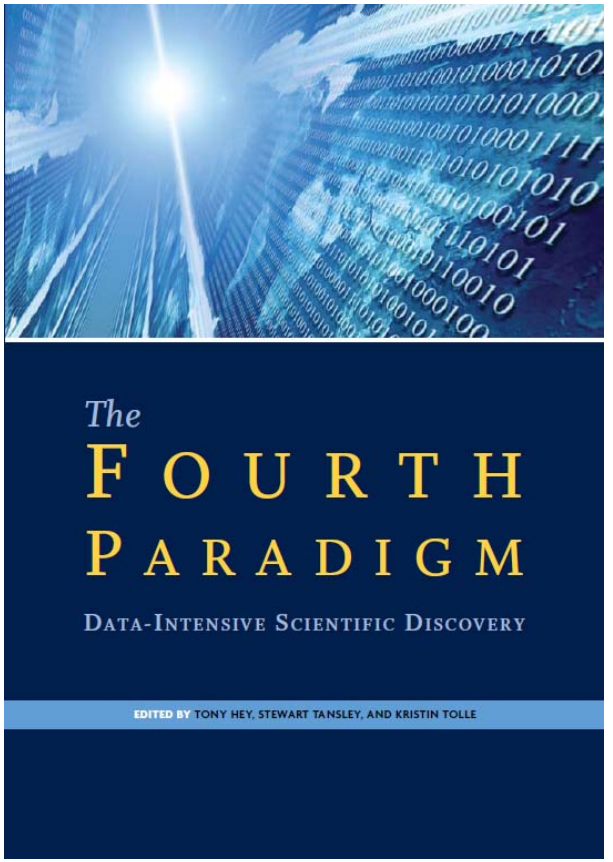
• recognizing and documenting well defined common problems

• cultivating a community of practitioners who work together on sharing a common informatics techniques

• leveraging the best practices, methodologies, conceptual approaches and tools

• improved knowledge extraction from Big data (e.g. finding outliers, finding trends in multidimensional sets, ...)

• New up-to-date educational tools and materials



Big Data is the 4th paradigm in addition to theory, measurements, and modeling