International Neuroinformatics
Coordinating Facility

Big Data, Little Knowledge

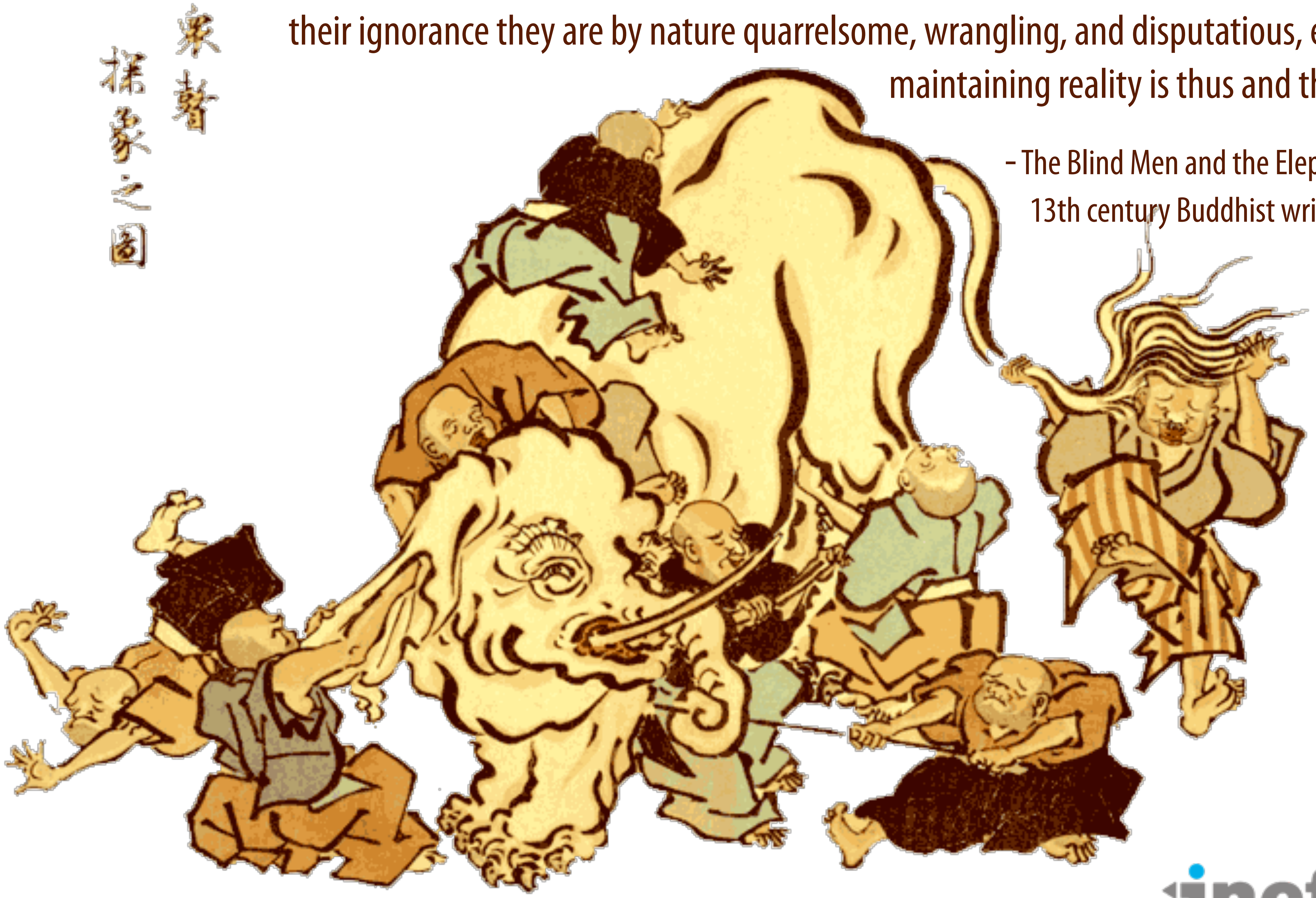
Challenges in Transforming Scientific Data into Knowledge

Prof. Sean Hill
Executive Director
International Neuroinformatics Coordinating Facility
Stockholm, Sweden

Head of Neuroinformatics Division
Blue Brain Project, EPFL
Lausanne, Switzerland

> “Just so are these preachers and scholars holding various views blind and unseeing.... In their ignorance they are by nature quarrelsome, wrangling, and disputatious, each maintaining reality is thus and thus”.

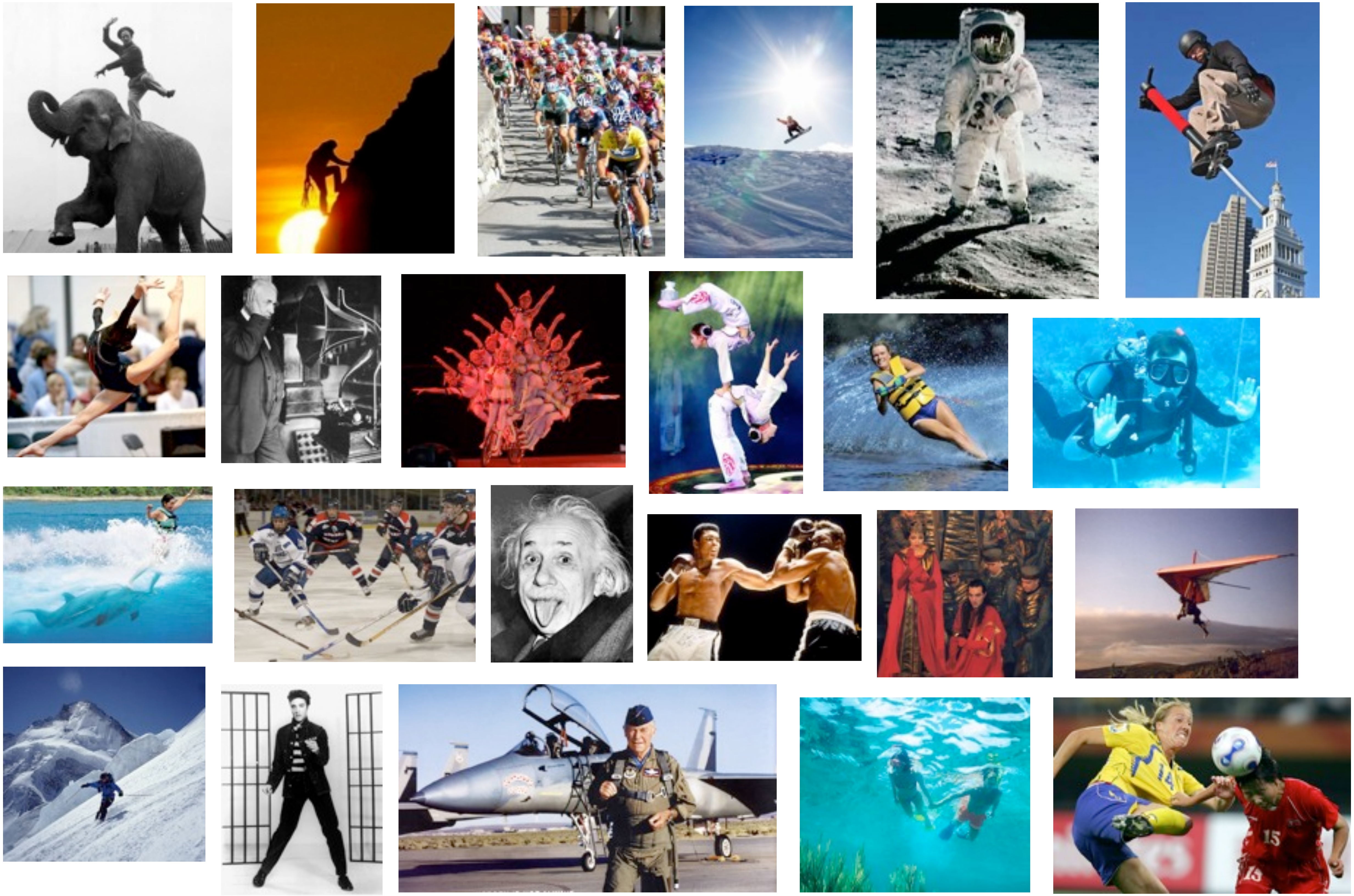
- The Blind Men and the Elephant
13th century Buddhist writings



incf

Upon this gifted age, in its dark hour,
Rains from the sky a meteoric shower
Of facts . . . they lie unquestioned, uncombined.
Wisdom enough to leech us of our ill
Is daily spun; but there exists no loom
To weave it into fabric;

Edna St. Vincent Millay, 1939



Developmental Disorders

- Autism spectrum disorders
- ADHD
- Learning disorders, conduct disorders
- Strong genetic disorders (Fragile X, Down’s etc)

Adolescent Disorders

- Depression, Suicide
- Eating disorders
- Bipolar disorder
- Conduct disorders and violence
- Borderline syndrome
- Adjustment disorders
- Anxiety, phobias, suicide
- Tourette’s syndrome
- Epilepsy

Adult Disorders

- Schizophrenia
- Epilepsy
- Mood disorders, hysterias, anxieties and phobias
- Obsessive compulsive disorders
- Eating disorders, sexual disorders
- Sleep disorders, stress disorders
- Impulse control disorders
- Substance abuse disorders
- PTSD/TBI

Aging Disorders

- Depression
- Dementia
- Neurodegenerative disorders
  - Alzheimer’s
  - Parkinson’s
  - Huntington’s
- Memory disorders

Glutamate

Nutrition

Dopamine

Genes

Sugar

GABA

Myelin

Serotonin

Metals

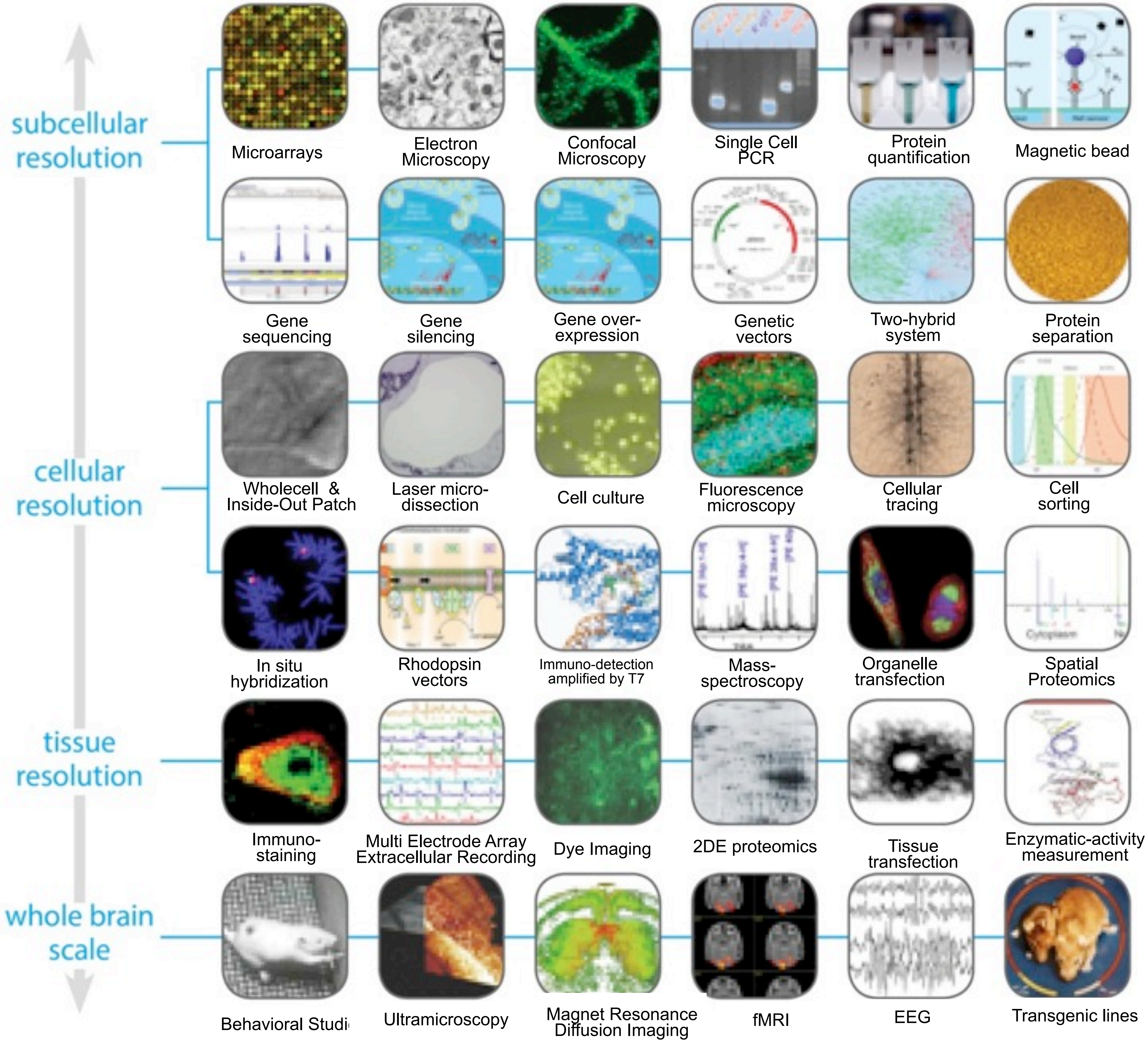
Dopamine

Toxins

Acetylcholine

Protein misfolding

"Big" Neuroscience Data



subcellular resolution

Microarrays
Electron Microscopy
Confocal Microscopy
Single Cell PCR
Protein quantification
Magnetic bead

Gene sequencing
Gene silencing
Gene over-expression
Genetic vectors
Two-hybrid system
Protein separation

cellular resolution

Wholecell & Inside-Out Patch
Laser micro-dissection
Cell culture
Fluorescence microscopy
Cellular tracing
Cell sorting

In situ hybridization
Rhodopsin vectors
Immuno-detection amplified by T7
Mass-spectroscopy
Organelle transfection
Spatial Proteomics

tissue resolution

Immuno-staining
Multi Electrode Array Extracellular Recording
Dye Imaging
2DE proteomics
Tissue transfection
Enzymatic-activity measurement

whole brain scale

Behavioral Studi
Ultramicroscopy
Magnet Resonance Diffusion Imaging
fMRI
EEG
Transgenic lines

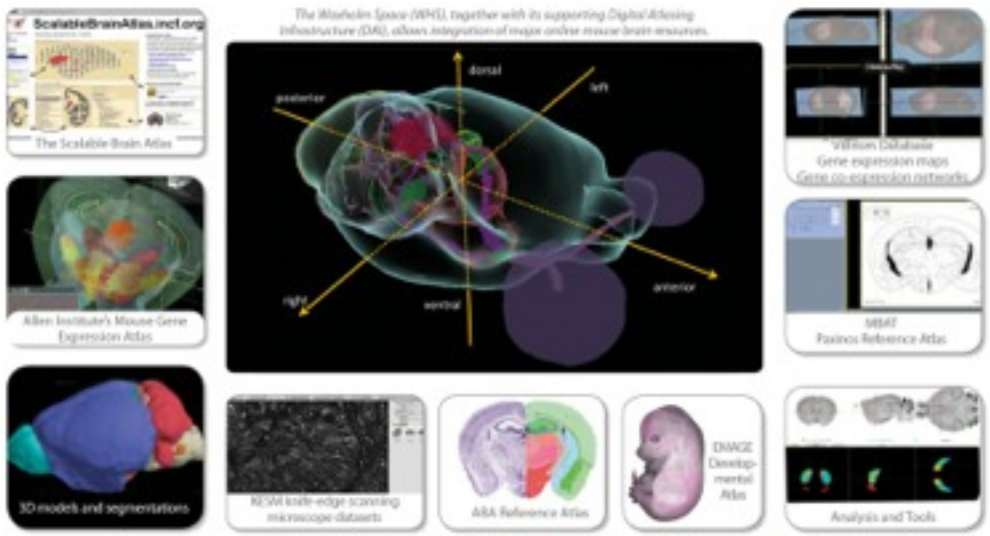
Need a scientific culture that values and promotes deep data integration

- Reward data integration and production of *integratable* data
- Define and annotate in a standardized way:
  - Who?(Ontologies, genetic identity, globally unique patient id)
  - What? (Data model, format, etc)
  - Where? (Space)
  - When? (Time)
  - How? (Protocols, experiment descriptions)
  - Why? (Models, Simulation, Analysis)
- Do good science (not as easy as it sounds!):
  - Understand the sources of variability (intrinsic and extrinsic)
  - Minimize extrinsic (measurement noise)
- Build unifying models that integrate and predict many different types of data
- Share data (ensure attribution, legal structures, etc)
- Share provenance, reproducible workflows, analyses
- Build infrastructures that facilitate these processes and exploit data integration

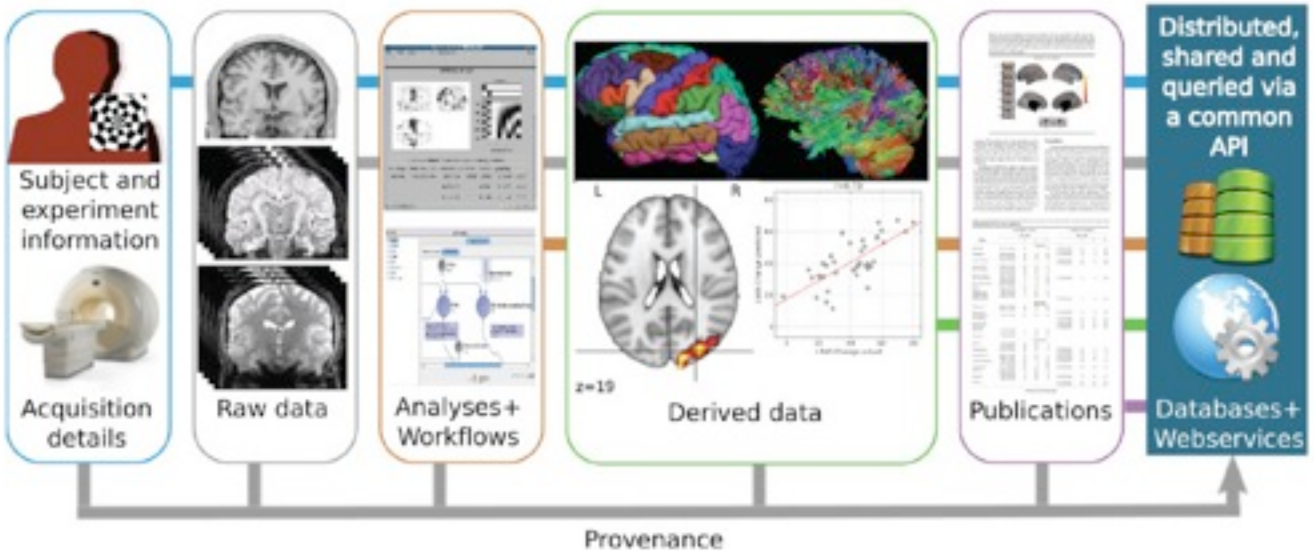
INCF - International organization initiated by OECD to coordinate and facilitate data integration standards and infrastructure for neuroscience

17 Member countries: USA, Europe, Asia, Australia

Digital Brain Atlasing


ScalableBrainAtlas.incf.org
The Scalable Brain Atlas
Allen Institute's Mouse Gene Expression Atlas
3D models and segmentations
KESM knife-edge scanning microscope datasets
ABA Reference Atlas
Analysis and Tools
Paxinos Reference Atlas

Data Sharing



Subject and experiment information

Acquisition details

Raw data

Analyses+ Workflows

Derived data

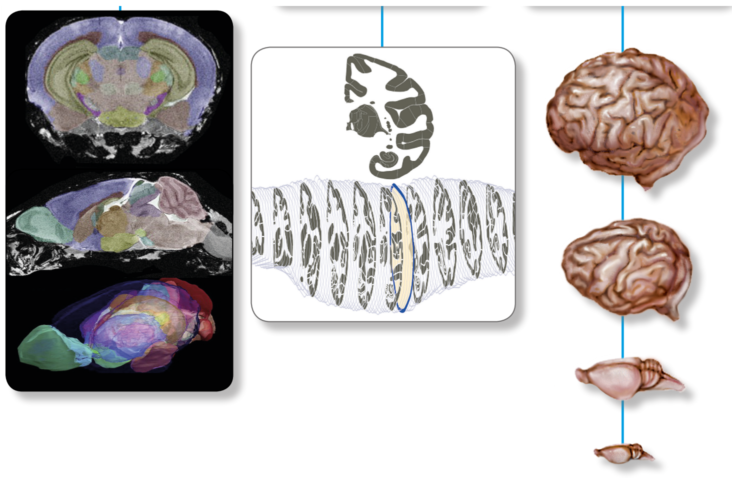
Publications

Distributed, shared and queried via a common API

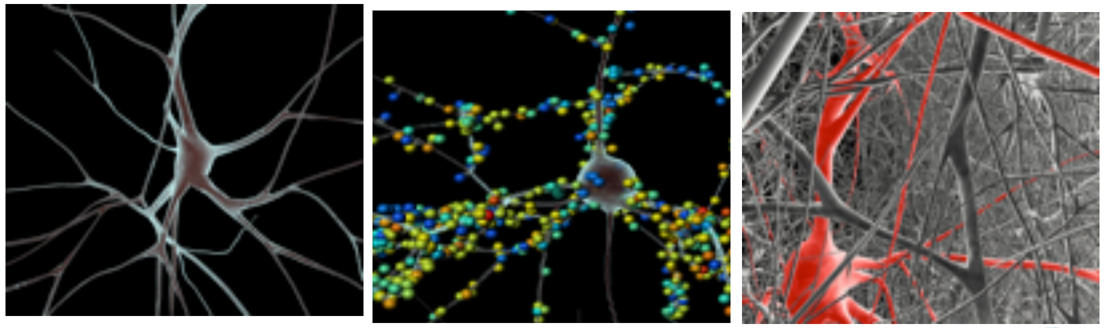
Databases+ Webservices

Provenance

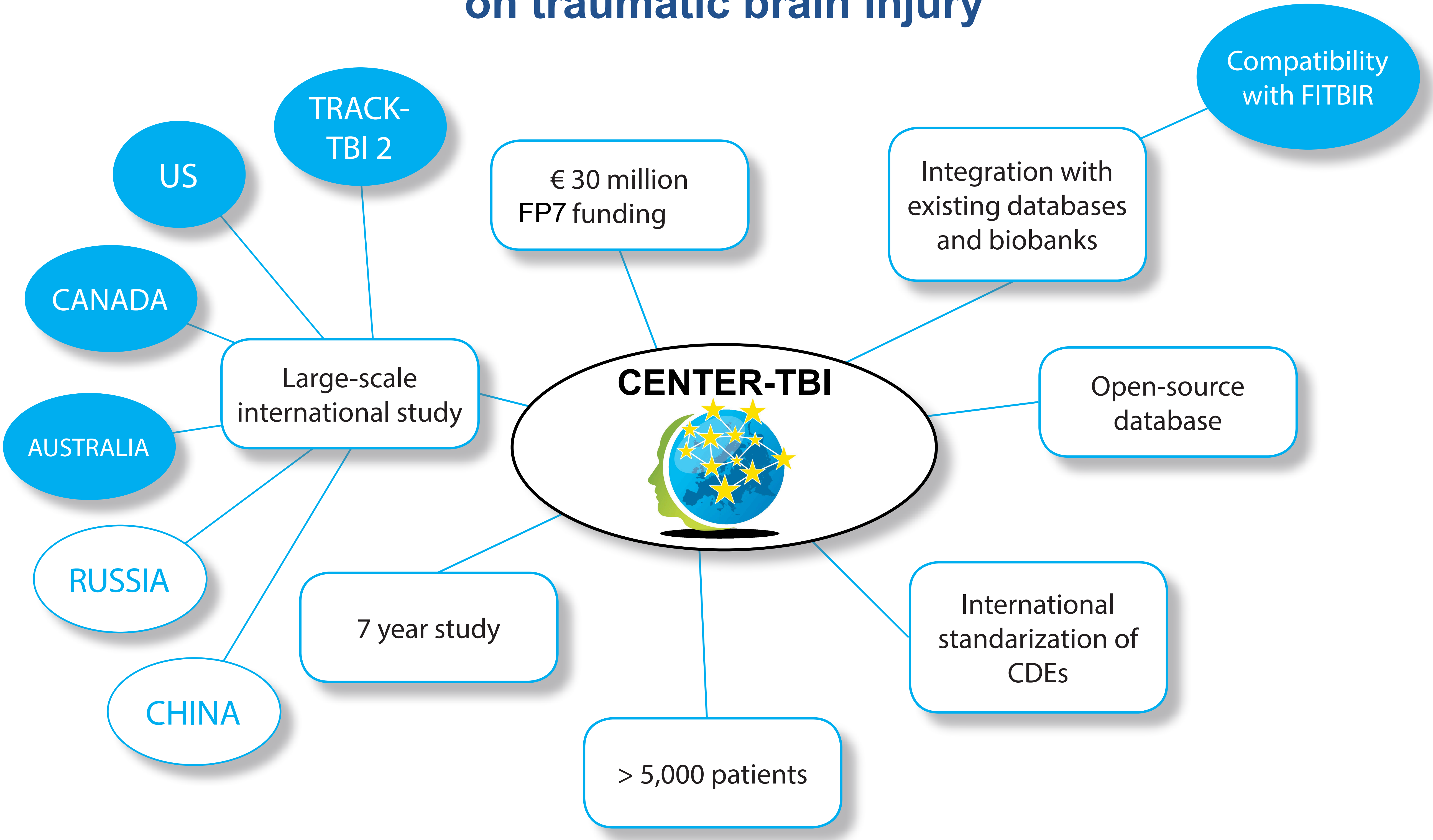
Ontologies of Neural Structures



Multiscale Modeling



International collaboration to produce “Big Data” on traumatic brain injury



CENTER-TBI

Large-scale international study

- US
- TRACK-TBI 2
- CANADA
- AUSTRALIA
- RUSSIA
- CHINA

€ 30 million FP7 funding

Integration with existing databases and biobanks

Compatibility with FITBIR

Open-source database

International standarization of CDEs

> 5,000 patients

7 year study