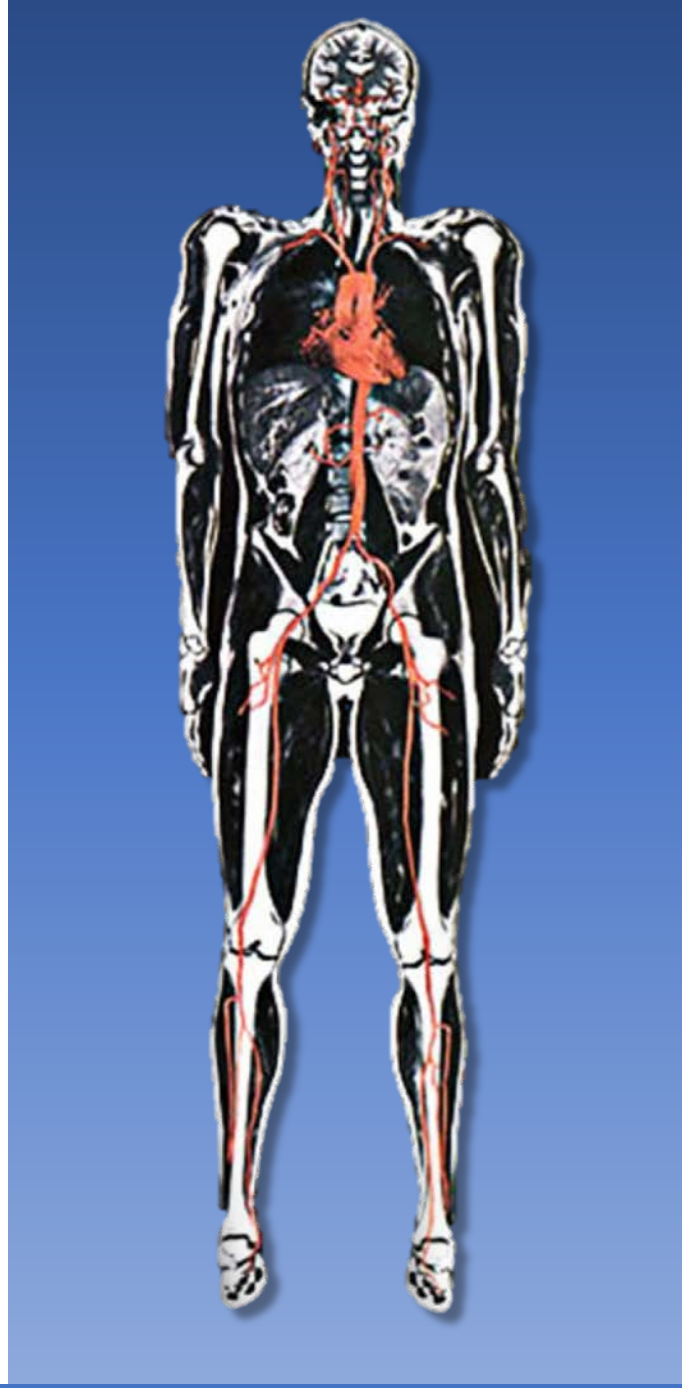
Medical Data

Summary of discussion

Philippe Lambin
Eric Grancher
Hans Hoffman
Francois Grey
Rolf Lewenshon
Renaud Jolivet
Sparsh Navin
Raj Jena



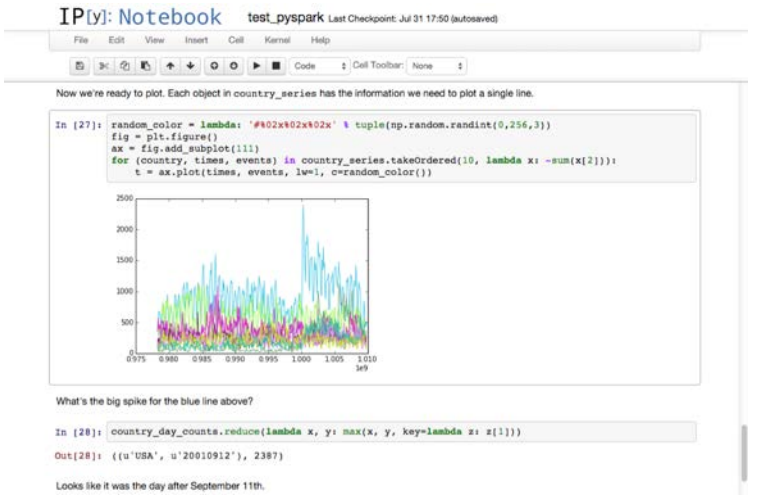
Emerging themes

- Facilitating open data
- Crowdsourcing of annotated data
- Develop distributed learning platforms in private and hybrid clouds
- Using Root as a platform for medical computing
- Simulation - something old and something new
- A challenge : Global integration of patient generated (IoT) and clinically derived data

TL;DR

Facilitating Open Data - 1

- Radiobiology
  - Use interactive notebooks to encode radiobiology experimental data
  - Encapsulate beam data, methods, raw data and analysis
  - Facilitate cross comparison of data for building RBE libraries
  - Include analysis goals (esp for incomplete experiments)
  - Leverage crowdsourcing for analysis - create community challenges



Facilitating Open Data - 2

- Showcase storage of annotated clinical & research data
  - Annotated data & Models from cancerdata.org
  - Try to bind to notebook format
  - Crowdsource – ask community to build better models
  - Set global challenges from this platform (like computer vision / ML community)

Facilitating Open Data - 3

- Store preclinical data (esp genomics)
  - Fulfill data repositary criteria of publishers
  - Link to cancerdata.org, expose DOIs
  - Use Zenodo as the umbrella platform for integrating with other data

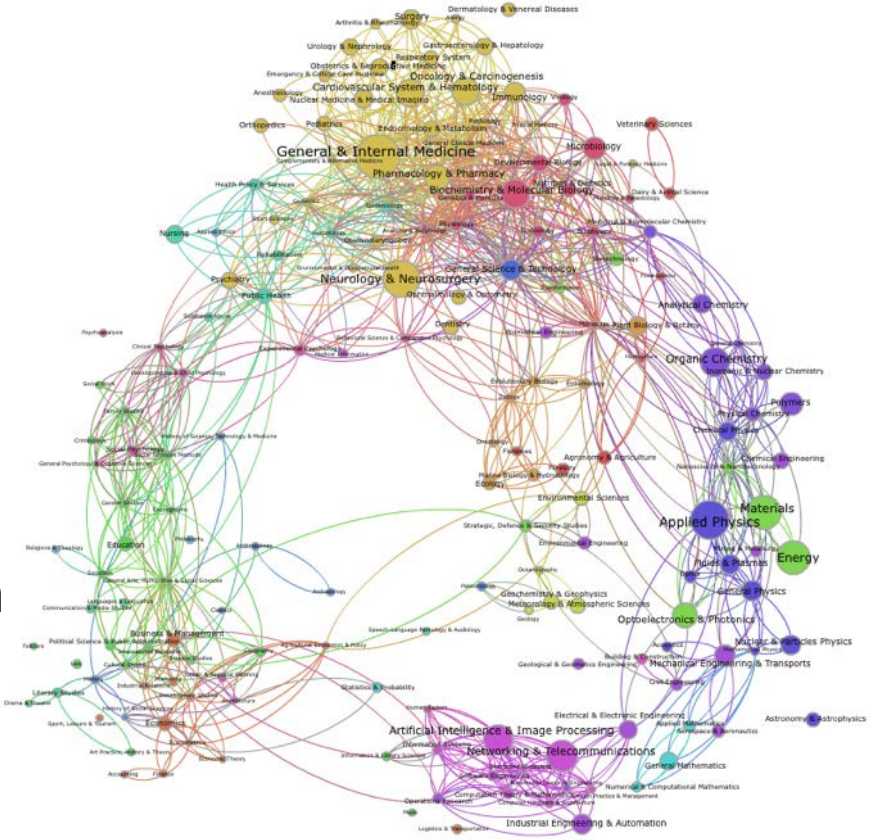
Distributed learning in the cloud

- WLCG/Openstack (LHC computing cloud) – distributed sites in 40 countries
- Hybrid cloud solutions – private cloud server inside DMZ of a hospital IT network
- Re-factor existing tutorials for hospital use
- Opportunity for community goals

Distributed learning in the cloud -2

- Precedent : Human brain project (Anastasia Ailamaki - EPFL)
- https://www.humanbrainproject.eu/fr/-/cern-open-lab-in-conjunction-with-hbp-publishes-a-whitepaper-on-future-it-challenges-in-scientific-research

- Spin-outs
  - Work with ESTRO/ASTRO on an Open Oncology Ontology



ROOT for medical computing

- ROOT is the compute platform used throughout CERN
- Open source
- Nice toybox – tools for SVM and ML
- Build workflows for computational tasks

Multi-scale Simulation

- Upgrade FLAIR / Fluka to allow integration into decision support workflows
- Use Geant4DNA to model DNA damage from cytotoxics and radiation together
- Or build physiological models to complement FLUKA in ROOT

http://www.fluka.org/flair/

Need to import medical expertise into CERN to build these tools

See one, simulate one, teach one

- Lot of tidy medical data and know-how encoded in these data & models
- Strong opportunities for teaching resources
  - Interactive
  - Distance learning

IoT challenge – Fetch this!

- Can CERN help to integrate a global medical dataset?
- Bring together
  - Clinically derived medical data
  - Patient generated data from IoT
  - Patient data access authorisation via TPM
  - In situ processing of patient data via TCM
  - Lessons (self-) learned from Microsoft Fetch?


BIG
DATA
INFORMATION
MANAGEMENT
TECHNOLOGIES
PETABYTES
ANALYSIS
RESEARCH
STORAGE
LARGE
SETS
MILLION
SEARCH
INTERNET
LEARNING
SOFTWARE
VOLUME
PROCESS
GOVERNMENT
SAN
NAS
DEVELOPMENT
TECHNOLOGY
SCIENCE
FUTURE
NEW
BILLION
PARALLEL
INITIATIVE
ECONOMIC
SIZE
EXABYTES
DATABASES
PROCESSING
ANALYTICS
SIMULATIONS
BUSINESS
FRAMEWORK
BETWEEN
STATISTICS
VISUALIZATION
APPLICATIONS
INFRASTRUCTURE
AMOUNT
REAL
STRUCTURE
WORK
HIGH
USERS
DEPARTMENT
SET
QUERIES
DETERMINE
WELL
ALGORITHMS
COMPANIES
HIGHER
RATE
TERMS
YEAR
PRIVATE
CAPABILITIES
SUCCESSFUL
NEEDED
USES
SHARED
COMPARED
LESS
HUGE
EVERY
WORLD
VARIOUS
BASED
TIMES
ARCHITECTURE
PARADIGM
FUNDING
INSIGHT
TOOLS
CENTER
RECORDS
COST
PROJECT
DISTRIBUTED
RELEVANT
USED
ANNUAL
NOW
TERABYTES
MARKET
VARIETY
RELATED
TIME
SENSOR
WOULD
CURRENTLY
PEOPLE
TRAFFIC
STATE
CRITIQUE
USE
MAPREDUCE
NEARLY
MOST
END
CRITIQUES
OUT
APPROACH
MANY
FLOW
NATIONAL
INTELLIGENCE
SECOND
PRIVACY
INTERNATIONAL
IMPLEMENTATION
BECOMES
APPROACHES
PROCESSES
PER
ONE
CAPACITY
HAND
HANDLE
SOURCES
CHALLENGES
LIMITS
COMPLEX
PAST
ANNOUNCED
INCLUDE
NETWORKS
HYPOTHESIS
REPLICATION
DRIVE
SOCIAL
RESULTS
YEARS
SYSTEMS
FACTOR
MANAGE
ONLINE
DAY
WORLDWIDE
MAKING
INTEGRATION
MASSIVE
DEFINITION
USING
EFFECTIVE

Thank you!