Review of Software Projects at IN2P3

Luisa Arrabito, LUPM
with the contribution of
A. Beck, P. Delsart, J. Frankland, H. Grasland, B. Khélifi,
S. Incerti, L. Maigne, A. Matta, O. Stézowski, R. Terrier,
A. Tsaregorodtsev, M. Verderi

IN2P3 Scientific Council, 23rd June 2022

Overview

- Review of software projects with a strong participation of IN2P3
- We present 11 projects from different domains : HEP, astro, nuclear physics, medical physics grouped in 3 categories
  - Simulation
  - Data processing and analysis
  - Workload, data and metadata management
- Not a detailed review
- We will not cover those projects already presented in the thematic sessions

Simulation

- Long-term involvement of IN2P3 in the development of Monte-Carlo simulation software
- Today, 3 main simulation projects around Geant4
  - **Geant4**
    - Toolkit for MC simulation transport of particle in matter
  - **Geant4-DNA**
    - Specialized in the interactions of ionizing radiation with the biological medium
    - Fully integrated in Geant4
  - **GATE**
    - Developed on the top of Geant4/Geant4-DNA
    - Specialized in medical imaging and radiotherapy applications
- Very good synergy among the 3 projects
- Responsibility role of IN2P3 staff with the 3 spokespersons

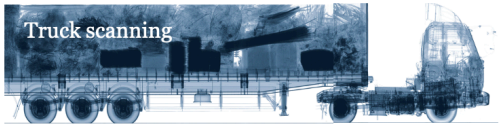
Simulation

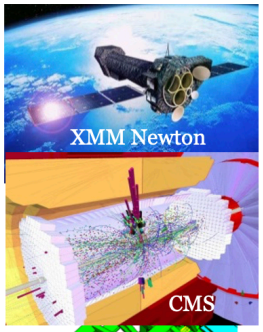
- **nptool**
  - Specialized in the simulation and data analysis of low energy (0-1 GeV) nuclear physics experiments
  - Also based on Geant4
- **Smilei**
  - Simulation of plasmas
  - Using the Particle-In-Cell (PIC) method
  - Massively parallel code

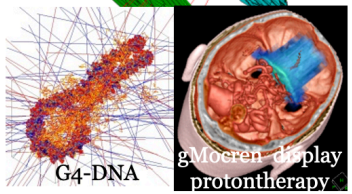
The Geant4 project

Geant4 in a nutshell

- Software toolkit for **Monte-Carlo simulation transport of particles in matter**
  - OO design, written in C++, free, open software
  - https://geant4.web.cern.ch/
- Wide physics coverage
  - Electromagnetic physics:
    - EM « standard » [1 keV – O(PeV)]
    - EM « low energy » [O(100 eV) – O(GeV)]
    - Geant4-DNA down to O(eV) + chemistry phase (free radical transport in water)
    - X-ray & optical
    - Phonons (mK)
  - Hadronic physics:
    - From rest to O(TeV) through
      - string models (high E)
      - cascade ones (intermediate E)
      - Evaporation, capture, etc. (low E)
    - Radioactive decay







- **Used in HEP, Nuclear Science, Medical, Space, Material Science, Homeland Security, etc.**
  - **Mission critical** in many of these areas
- Started in 1994, as the RD44 CERN project
  - Assess interest of OO technologies for particle transport simulation for needs of (future ;)) LHC
  - Alpha version in Apr. 1997, beta one in Jul. 1998
  - **Geant4 version 1.0 Dec. 98**
  - **1st Jan. 1999 : RD44 → Geant4 collaboration**
    - Today: ~130 members, ~30 FTE, 16 Work. Groups
- **Parallel processing** from 2013 with **Geant4 10.0**, with an "event parallelism" scheme
  - Multi-threading processing
    - Ensure all cores have enough memory to run: done by sharing geom. & phys. tables across threads
  - Add flexibility in 2021 with Gean4 11.0 with "tasking"
    - Easier access to hybrid computing as by-product
- **Collaborative & worldwide distributed development model** (GitLab based)
  - With repository and testing infrastructure at CERN
  - Step toward « open development model » this year

The Geant4 project

Challenges

Computing & Physics Performances

- Issue specially acute for HL-LHC
  - **Need factor O(10) speed-up** to fit simulation production in planned HL-LHC computing budget !
  - Need **physics quality to be improved accordingly**
    - to avoid larger (relative) contribution to systematics
- **Strategy:**
  - **Detailed simulation as much as possible**
    - On CPU, with adiabatic evolution
      - When CPU production sustainable
      - Or when no alternative (ie: too complex physics or geom.)
      - And creativity : eg neutron biasing, Woodcock tracking, …
    - On GPU, in hybrid computing workflow
      - **if** current R&D's outcome convincing (news in ~1 year)
        - HEP MC is **evil case** for GPUs ! (issue : divergence)
      - Ongoing GPU R&D's : AdePT (CERN), Celeritas (US)
    - Maybe on other hardware
  - **Fast simulation as much as needed**
    - Classical fast simulation à la GFlash
    - And/or modern fast simulation, **ML-based**
      - EM one & **maybe hadronic one** too
  - **In close collaboration with experiments !**
- On going efforts, HEP-motivated, beneficial to all domains

Risk of Expertise Disruption !

- Many "senior" members were present since the beginning, 28 years ago
- Most of them will have **retired at the HL-LHC start** !
- **Renewal rate too low !**
  - **Brilliant** young members !
  - But with short term contracts…

A new generation of developers is needed !

- Need **new/young members** on the **long run**
- **Smooth transition** must be planned
  - Some expertise **hard to find, specially in physics.**
  - Overlap must be long O(5 years) to become "senior"
- **Conflict** with short term contract policies !
  - An issue all long term projects complain about…
- Working climate & perspectives should be **secure enough**
  - To minimize the risk of loosing people
  - The more talented people, the higher the risk !
- **We alert our funding agencies on the critical need for a new & long term generation !**
  - **This issue is a risk for all Geant4 users !**

The Geant4 project

Geant4 @ in2p3

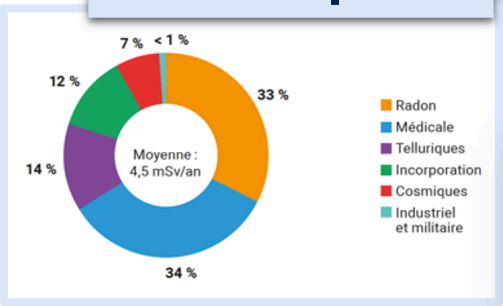
- In2p3 involved in Geant4 since the beginning
  - Started with Michel Maire, LAPP (now retired but still active in G4) an historical developer of EM physics
  - Then joined for fast simulation infrastructure
  - And visualization, analysis
  - …
- In2p3 members have quite **responsibilities** in Geant4:
  - See Steering Board page
  - Responsibilities in Working Groups:
    - Electromagnetic Physics :
      - Largest WG (5 coordinators)
      - **Sébastien Incerti** coordinator for DNA part
    - Generic Processes and Materials :
      - Holds in particular Fast Simulation & Biasing
      - **Marc Verderi** coordinator
    - Novice and Extended Examples :
      - Key entry point for new users and source of "how to"
      - **Ivana Hrivnacova** coordinator
    - For information: Visualization:
      - Laurent Garnier coordinator, was in LAL, moved to Irisa
  - Management:
    - **Marc Verderi**, reelected spokersperson in May this year
      - Second 2-years mandate
    - Was deputy spokesperson for 10 years before
- Geant4 at in2p3 is « Geant4-core » and « Geant4-DNA »
  - DNA part of the Geant4 development cycle and distribution
  - G4-DNA is also a collaboration, driven by Sébastien Incerti
- **Current and to be pursued activities at in2p3:**
  - LP2i Bordeaux:
    - Low energy EM physics and Geant4-DNA
      - Note this includes **GPU usage** for the transports of chemical species in water after the irradiation phase
    - **Sébastien Incerti, Zhuxin Li (thesis), Claire Michelet, Hoang Tran**
  - IJCLab:
    - Analysis package, visualization, basic and extended examples and geometry
    - **Ivana Hrivnacova, Guy Barrand**
  - LAPP:
    - EM physics validation
    - **Sabine Elle**
  - LLR:
    - Bethe-Heitler 5D gamma conversion model (electrons and muons) inherited from developments in HARPO
    - Biasing (variance reduction) techniques
    - Fast simulation infrastructure & GFlash models
    - **Denis Bernard, Igor Semeniouk & Marc Verderi**
- Yearly Geant4 tutorial in Orsay (PHENIICS + ANF every 2 years)
  - Managed by **Ivana**, + Igor and Marc as lecturers

The Geant4-DNA project

Context: exposure to ionizing radiation

1 mSv/an
Public

Chronic exposure



IRSN (2014-2019)

Increasing dose*

Repetitive medical imaging

UCSF & Harvard MS

~10 mSv / scanner abdomen

>100 mSv
Long-term effects

Space missions

ESA

~2,6 times ISS

ESA

6 months = ~ 160 mSv

ESA

Return trip 6 months:
~ 600 mSv

Radiotherapy

NCC

Major questions in radiation protection

- What are the **biological effects** of **low doses** of ionizing radiation?
- What are the **risks** associated with **high doses**?

*« A major challenge lies in providing **a sound mechanistic understanding of low-dose radiation carcinogenesis** »*
L. Mullenders et al. (2009)

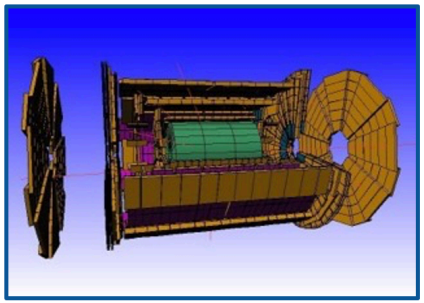
The Monte Carlo approach…

* Equivalent or effective

The Geant4-DNA project

The « Geant4-DNA » project

ATLAS@CERN



geant4.org

Initiated in 2001, based on an idea of the European Space Agency (P. Nieminen)

- To develop the **first open source platform** able to **simulate mechanistically the biological effects of ionizing radiation at cellular scale** during manned space missions

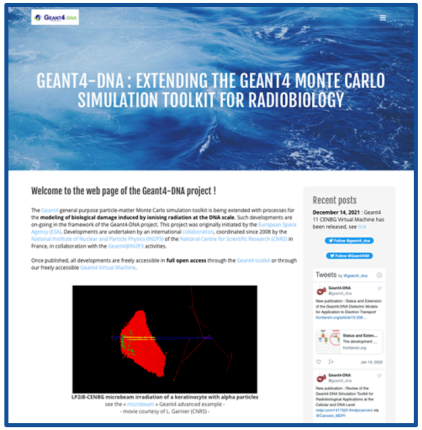
Set of C++ libraries able simulate the different stages of the effects of ionizing irradiation

- Open source, distributable (many OSs), user extensible (OO)
- Radiation **physics** (particle-matter interactions), radiation **chemistry**, radiation **biology**…

Fully integrated in the « Geant4 » Monte Carlo simulation toolkit

- **Ge**ometry **and** **t**racking – version **4**
  - Developed by an international collaboration, initially for the simulation of the large experiments of the LHC at CERN

geant4-dna.org



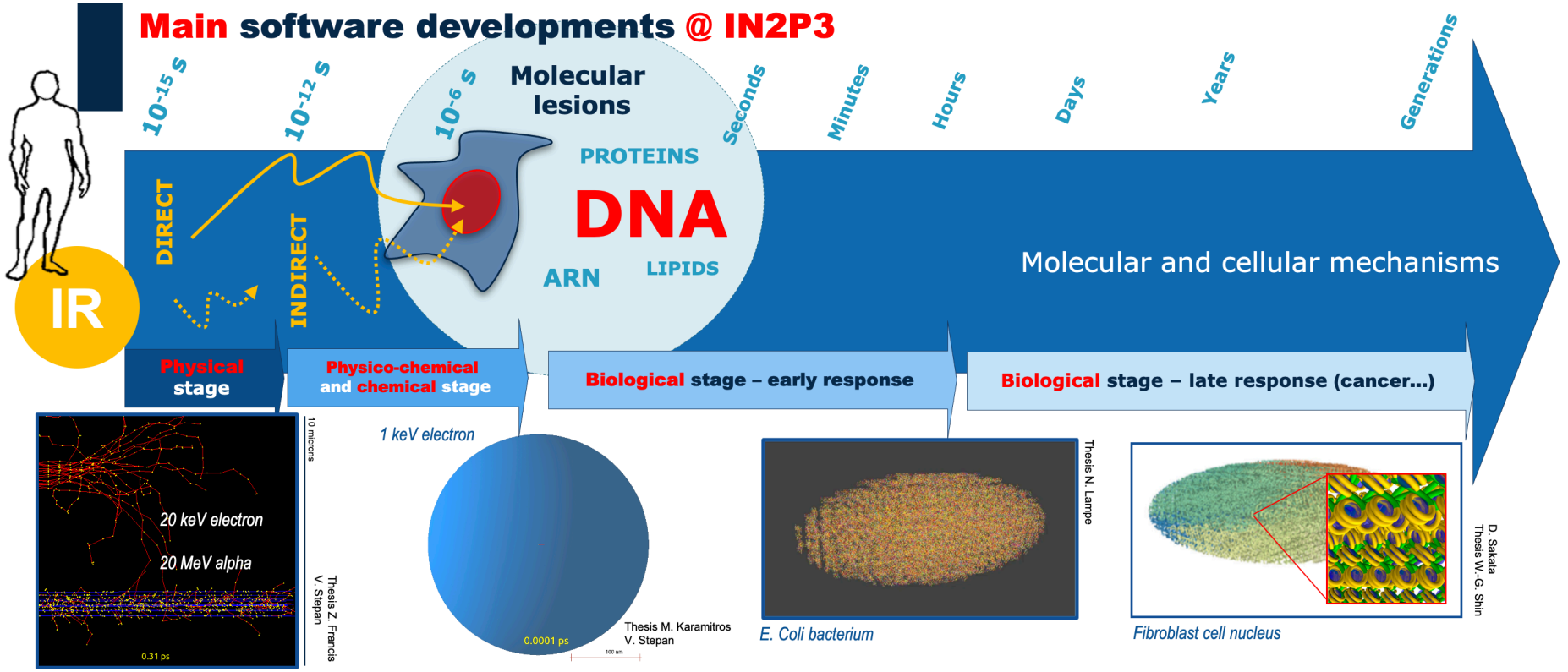
First release of « Geant4-DNA » physics models in 2007

- Step-by-step simulation of particle transport down to low energy (~10 eV) in the biological medium
- **IN2P3** contribution: LP Clermont and LP2i Bordeaux (ex-CENBG)

International collaboration funded in 2008

- ESA proposed to **IN2P3** to lead this collaborative project
- 60 collaborators in 2022 – **13 @ IN2P3** (IPHC, LP2i, LPC, SUBATECH)

The Geant4-DNA project



Simulate « **physical** » **interactions** ($<10^{-15}$ s) with biological matter: **direct** effects and **dissociation** of water molecules.

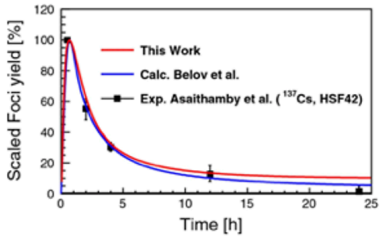
Models describing the **« step-by-step » physical** interactions of **electrons, protons, alpha particles and ions** with biological matter (water, DNA constituents) down to low energy (~ 10 eV)

Simulate « **physico-chemical** » ($<10^{-12}$ s) and « **chemical** » ($<10^{-6}$ s) processes with water molecules in the biological environment: creation of numerous molecular species (e.g. °OH) that cause **indirect effects** on DNA

Alternative approaches to model radiolysis: Steb-by-step, Independent Reaction Time

Simulate **early DNA damage** and its **repair** following physical, physico-chemical and chemical steps (min - h) in biological targets.

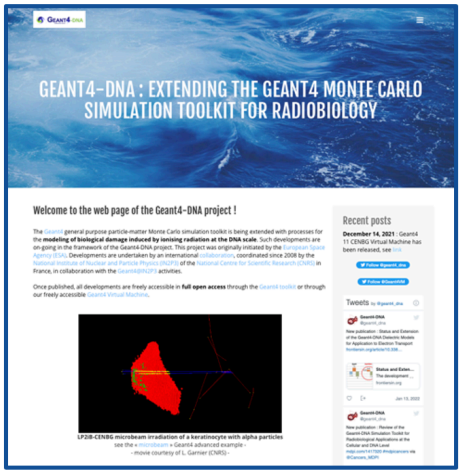
- **Multi-scale geometries** of biological targets
- **Complete simulation chain** for early damage quantification in bacteria and cells
- **Repair** models



DSB foci repair as a function of time

The Geant4-DNA project

Scientific production and animation



All details at:

geant4-dna.org

> 120 publications & open source code (geant4.org – 2 Geant4 releases per year)

Key contributions by young researchers

- 14 PhD **theses** (radiation physics – chemistry – biology) including 13 supervisions/co-supervisions by IN2P3
  - Funding: Bordeaux U., Paul Sabatier U., Conseil général de l'Allier, IRSN, ANR, AUF, CSC Chine
- 17 **postdoctoral** fellows
  - Funding: IN2P3, Bordeaux U., IRSN, AVIESAN, ESA

Regular organization of international conferences, workshops and tutorials

- « **Geant4 at the Physics-Medicine-Biology frontier** » series of international conferences initiated in Bordeaux in 2005
- 4th iteration in Napoli, Italy, on Oct 24th-26th, 2022
- LP2i « **Geant4 Virtual Machine** » (C. Seznec *et al.*): easier access, popular for tutorials

Additional funding

- IN2P3 « collaborative tools »: IRL (South Korea, Japan), IEA (Greece, Lebanon, Serbia), exchanges (JINR, Czech R.), MITI
- Région Nouvelle Aquitaine, Aviesan/Inserm, ANR
- ARC Australia, CampusFrance STAR South Korea, European Space Agency (BioRad1, BioRad2, BioRad3), FNS Switzerland

Permanent recruitments

- 1 IN2P3 research engineer in 2022 @ LP2i Bordeaux: Dr Hoang Tran, now **Technical Coordinator** of the Geant4-DNA Collaboration
- Regular at other institutions: ANSTO (Australia), Campinas (Brazil), Ioannina U. (Greece), IRSN (France), KEK (Japan), NPI (Czech R.) Osaka U. (Japan)

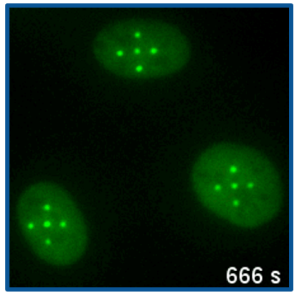
Collaboration led by IN2P3

- IN2P3 re-elected every two years

The Geant4-DNA project

A collaborative program for the next 10 years

... in line with GT10



Ph. Barberet et al. / LP2i

DSB foci
AIFIRA microbeam

Radiation physics

- Extension & improvement of **physics models**: energy / particles / target materials (e.g. solid-state, gaseous)

Radiation chemistry

- « **Mesoscopic** » approach: towards longer times, larger volumes and high dose rates (e.g. Flash RT)
- Experimental **validation** of simulation approaches (SBS, IRT, mesoscopic) with irradiation platforms (e.g. Arronax)

Radiation biology

- A library of **multiscale biological target geometries**: chromatin fibers, plasmids, bacteria, cell nuclei of different lines,...
- Integrated **simulation chains** (physics - chemistry - biology) allowing the prediction of early damage in various lines
- **Repair** models: from early to late effects
- Experimental **validation** of damage (e.g. AIFIRA, Cyrcé, ALTO…)

Computing performance

- **GPU** porting accessible in open source
- **IA**

Medical, space & environment applications

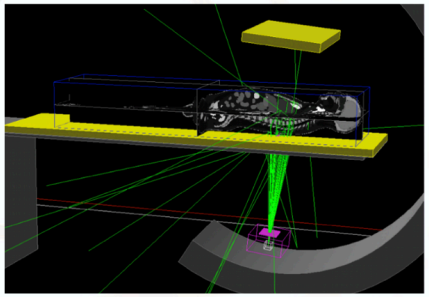
- **Innovative radiotherapies**: internal (new radionuclides, BNCT, …) and external (hadrons, Flash, VHEE, …) – towards risk prediction
- **Space radiation protection** (in coll. with ESA): from incidental spectra to biological damage (e.g. lunar habitat)
- **Environment**: effects of ionizing radiation on living organisms

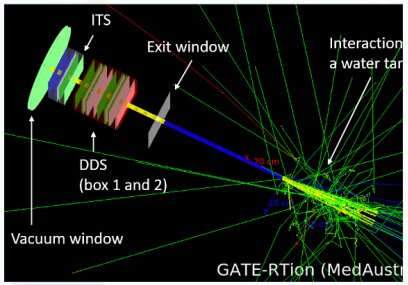
Strategy

- Encourage collaborations with **partners** (INSB, Inserm, CEA, IRSN, etc.) and collective responses to **calls** (Aviesan, Pianoforte, etc.)
- **For the first time, the Geant4, Geant4-DNA and GATE collaborations are led by IN2P3: an opportunity to strenghthen our links?**

The GATE project

a MC simulation platform for medical physics applications





An open source and open access simulation platform for **medical physics**

- **Long term development** based on Geant4 toolkit
- International collaboration composed of **25 members**: laboratories (**6 IN2P3 labs**), companies and clinics
- **Very large community of users** (more than 2000 users)
- **Newly coordinated by IN2P3/CNRS** since 2018
- Funded by regular support from institutions and international calls
- Fruitful involvement in international conferences, workshops & tutorials
  - 2 tutorials /year + Tutorials in medical physics master programs
- High rank and highly cited publications
  - 2 PMB citation prizes in 2009 and 2015 for the 2 collaboration papers

A PLATFORM TO GATHER INNOVATIVE DEVELOPMENTS FOR THE FUTURE CHALLENGES IN MULTI-SCALE SIMULATIONS

The GATE project

The collaboration

25 members, public laboratories and companies developing an open source platform

Spokesperson: Lydia Maigne (LPC)
Technical coordinator: David Sarrut (CREATIS, Lyon)

Research engineer just recruited at LPC for 3 years (IN2P3 funding)

France

- U1101 INSERM, Brest
- IJCLab – CNRS-IN2P3, Paris Saclay
- IRCM INSERM, Montpellier, France
- LPC – CNRS-IN2P3, Clermont-Ferrand
- IPHC – CNRS-IN2P3, Strasbourg
- CPPM – CNRS-IN2P3, Marseille
- UMR5515 CNRS-INSERM, CREATIS, Lyon
- IP2I, CNRS-IN2P3, Lyon
- BioMaps CEA Paris Saclay
- CRCT - U1037 INSERM, Toulouse
- LPSC – CNRS IN2P3, Grenoble

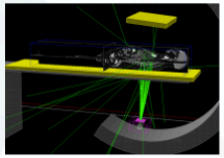
Europe

- Institute of Nuclear Physics Polish Academy of Sciences, Poland
- University of Julich, Germany
- University of applied Sciences, Aachen, Germany
- Medisip, Ghent University, Belgium
- BioemTech, Athens, Greece
- Medical University of Vienna, Wiener Neustadt, Austria
- MedAustron, Wiener Neustadt, Austria
- Christie Medical Physics & Engineering, Manchester, UK
- JPET collaboration, Poland
- Univ. of Patras, Dept of Med. Phys., Greece

Elsewhere

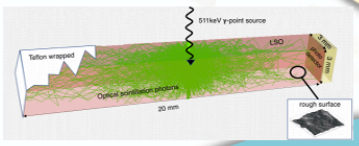
- Memorial Sloan-Kettering Cancer Center, New York, USA
- UC Davis, Davis, USA
- Sogang University, Seoul, South Korea
- QST, Chiba, Japan

The GATE project



DOSIMETRY

- **Interventional radiology**
- **Multimodal nanoparticles** (stem cell therapy on muscle regeneration)
- **Innovative internal RT (177Lu…)**
- CT scan (pediatric applications)
- **Arctherapy**
- SBRT – Interplay effect & EPID
- Database of S-values (OpenDose)
- **GATE/G4DNA micro & nanodosimetry**



PYTORCH

COMPUTING

Neural Networks for:
- Positioning of γ interactions in monolithic PET detectors
- SPECT detector ARF
- GAN for compact beam source modeling

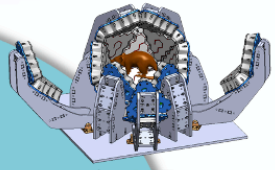
HYBRID SIMULATIONS

PYTHON data analysis

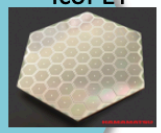
HOT Topics

IMAGING

- Models to define crystal surfaces
- **Optical simulation study on timing performance of monolithic crystals**
- **Testing of new camera and crystal designs**
- **New digitizers for SiPM (analogue and digital)**
- Compton camera
- Proton radiography
- Non-circular orbit motion in SPECT
- JPET (plastic scintillator-based PET)
- Metabolic modeling
- Digital Photon Counting commercial PET



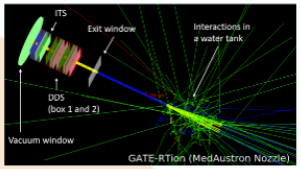
IcoPET



UTOFPET

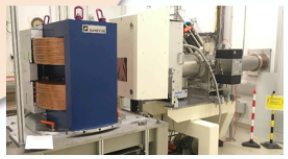
HADRONTHERAPY

- G4 Physics settings
- on-line MR-guidance for particle therapy (integration of magnetic and electric field maps)
- **Neutron dosimetry**
- **Biophysical models for the prediction of RBE ->G4DNA**
- Carbon ion dosimetry and ICRU90 recommendations
- **Range monitoring**
- **FLASH therapy**
- **BNCT & PBCT**



Magnet

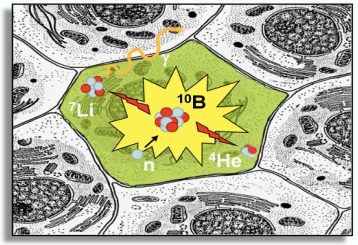
Nozzle

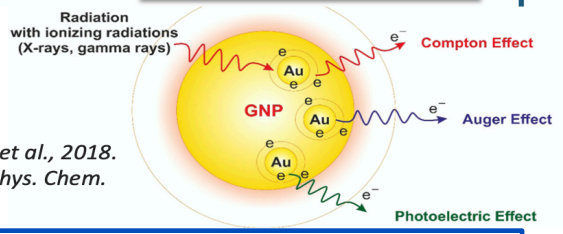


The GATE project

Highlights and future challenges in multiscale simulations

- **Highlights**
  - Validated **quality insurance** of preclinical and clinical beams
  - Reference platform for theranostic treatments
  - Integration and validation of biophysical models to tackle relative biological effectiveness in hadrontherapy
  - New developments to assess **microdosimetry** & **nanodosimetry**
- **Challenges**
  - Improve biophysical models
    - Better combine MCTS simulation outputs to macroscopic simulations
    - Use AI to train the models for different types of cells
  - Adapt or develop new biophysical models for new techniques
    - FLASH, mini-beam therapy, VHEE, IRT, NP…..
  - Combine the models to tumoral growth simulations
  - Combine to statistical analysis of clinical data





Borran et al., 2018. Rad. Phys. Chem.

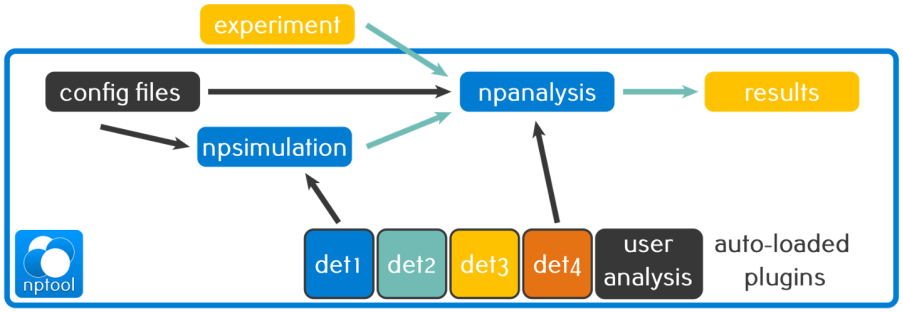
- **A new major release of GATE 10. will be published in about 1 year**
  - Complete refont of C++ classes with Python encapsulation for a more flexible user interface and a better compatibility with all analysis codes written in Python

The nptool project

nptool: key concepts

Key concepts

- A common framework for low energy nuclear physics
- Best use of Root and Geant4
- In use since 2008
- Modular design:
  - Detector and Event Generator as plugins
  - Tool box: tracking, nuclear data, scorer ,…

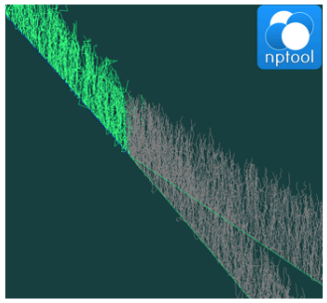


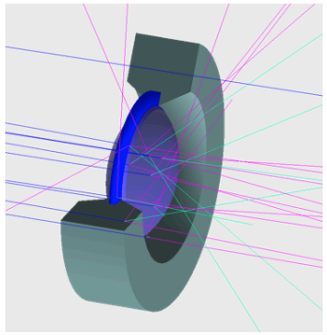
Collaboration

- Framework team: A. Matta, F. Flavigny (LPC Caen), N. de Sereville (IJCLab), V. Alcindor (TU-Darmstadt), P. Morfouace (CEA-DAM), M. Labiche(STFC)
- World wide user base: France, Germany, UK, Spain, Italy, Finland, Norway, Romania, USA, Canada, Japan, India
- Open collaboration: every body is welcome
- Ref. Article : J. Phys. G: Nucl. Part. Phys. 43 045113
- Repo: gitlab.in2p3.fr/np/nptool. Website: nptool.org

The nptool project

nptool: v3 in exploitation





A lot of detectors to choose from (~ 70)

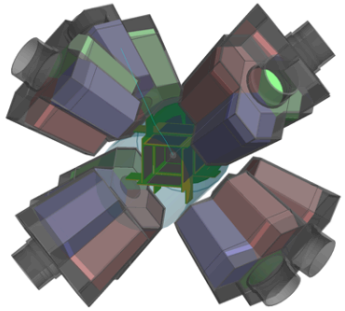
- Silicon (Hira, Share, T-Rex, Grit,…)
- HPGe (Agata, Miniball, Exogam,…)
- Gas (Actar, Minos, Tactics,…)
- Scintillator (Paris, Fatima, Nebula,…)
- Magnetic (ISS, Samurai, Swiper,…)
- Cryo Target w/ deformation

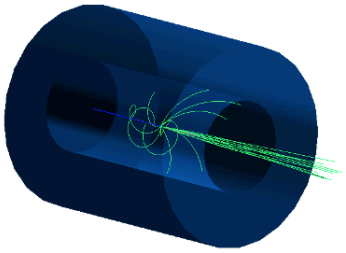
Event generator and physics

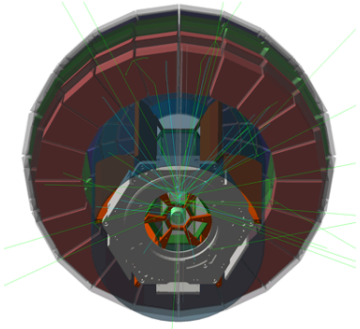
- Direct reaction with CS : transfert, coulex, knock-out
- Nuclear decay : gamma and particle decay, neutron physics, fission,…
- Drift electron simulation for gaseous detector (TPC, IC,…)

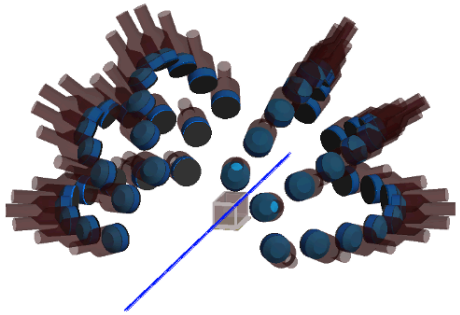
Publications

- Used for about 15 PhD thesis work
- Cited in 24 journal articles









The nptool project

nptool: v4, a pip like experience

Philosophy

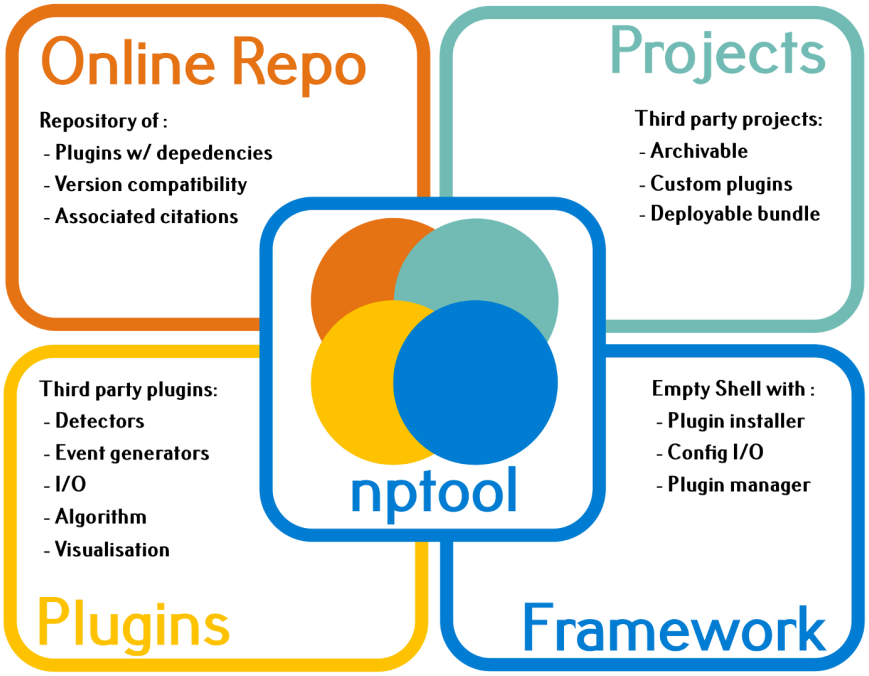
- **Separate plugin and framework**
- **Minimum lightweight installation**
- **Automatic dependencies**

For collaborations

- **Control their own repo. of plugins**
- **Easy CI/CD workflow**
- **Increase visibility with citations utilities**

For end user

- **Plugins on demand**
- **Hassle free installation**
- **Custom projects w/ duplicate plugins**
- **Open science friendly (bundle, archive and containers)**

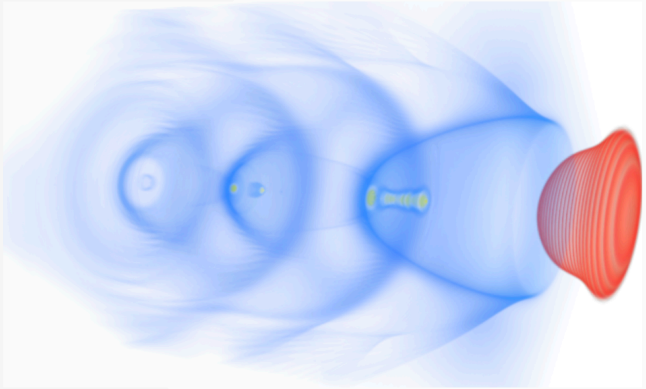


beta version under test /release in 2023

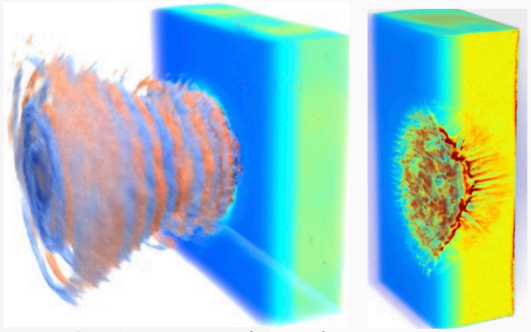
The Smilei project

Smilei is a Particle-In-Cell (PIC) code for the simulation of plasmas

Laser Plasma Interaction

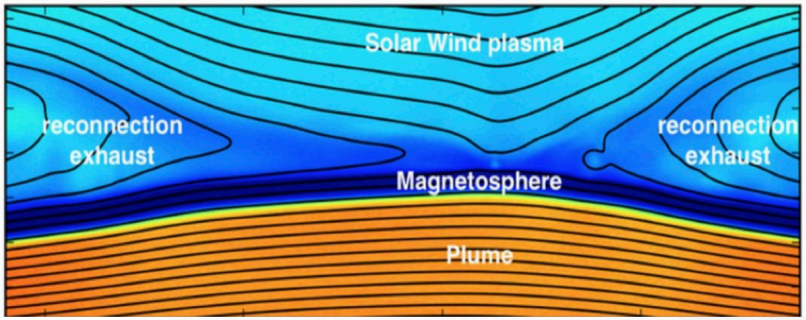


source: Massimo et al. (2020)



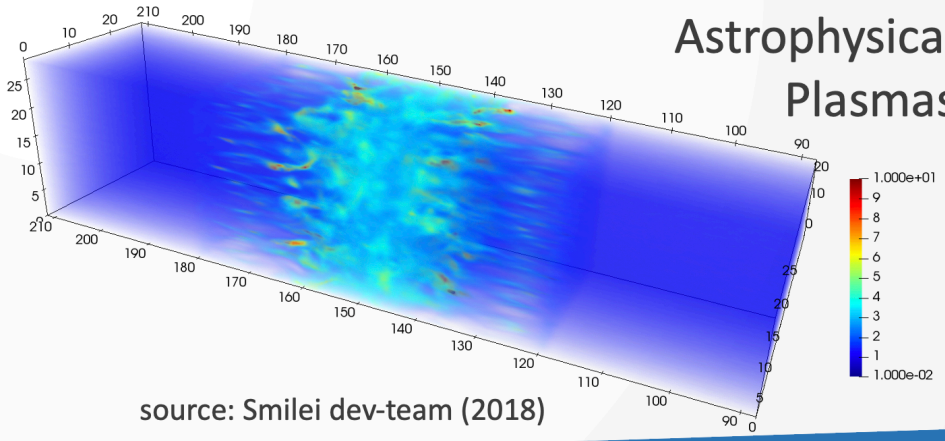
source: Smilei dev-team (2018)

Space Plasmas



source: Dargent et al. (2017)

Astrophysical Plasmas



source: Smilei dev-team (2018)

The Smilei project

Smilei is massively parallel and anchored in the HPC landscape

Integration in the French & European HPC landscapes

GENCI
Le calcul intensif au service de la connaissance

EuroHPC
Joint Undertaking

- Running on all super-computers in France (and many in Europe)
- 10s millions computing hours every year via GENCI & PRACE
- GENCI technological survey
- French Exascale Project

Special/early access to various machines

2015 IDRIS/Turing BlueGene-Q
2016 CINES/Occigen
2018 TGCC/Irene-Joliot-Curie
2019 IDRIS/Jean Zay
2021 RIKEN/Fugaku

The Smilei project

Smilei main contributors: Co-development between physicists & HPC specialists

Core dev-team

Mickael Grech*
Frederic Perez*
Tommaso Vinci*

Arnaud Beck*
Guillaume Bouchard

Mathieu Lobet*
Francesco Massimo*
Res. Engineer**

Computing centers and manufacturers

Physics Laboratories

The code

- 150 000 lines of code
- C++ 85% / Python 15%
- 300 man.months in dev. only
- 10's MCPUhours every year
- CeCILL-B licence / Github

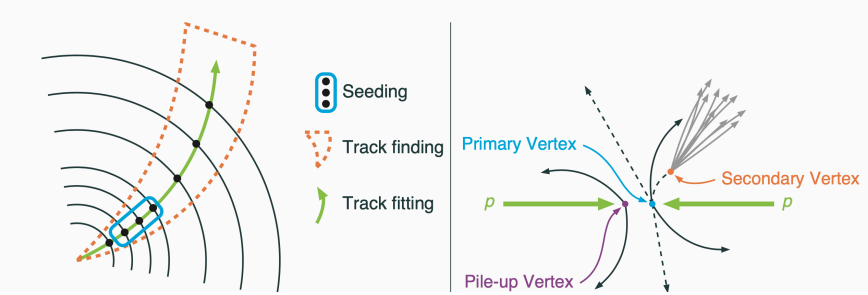
*permanent staff **Res. Engineer position to be opened (CNRS DDOR, w.s.f. INP, INSU, IN2P3)

Data Processing and analysis

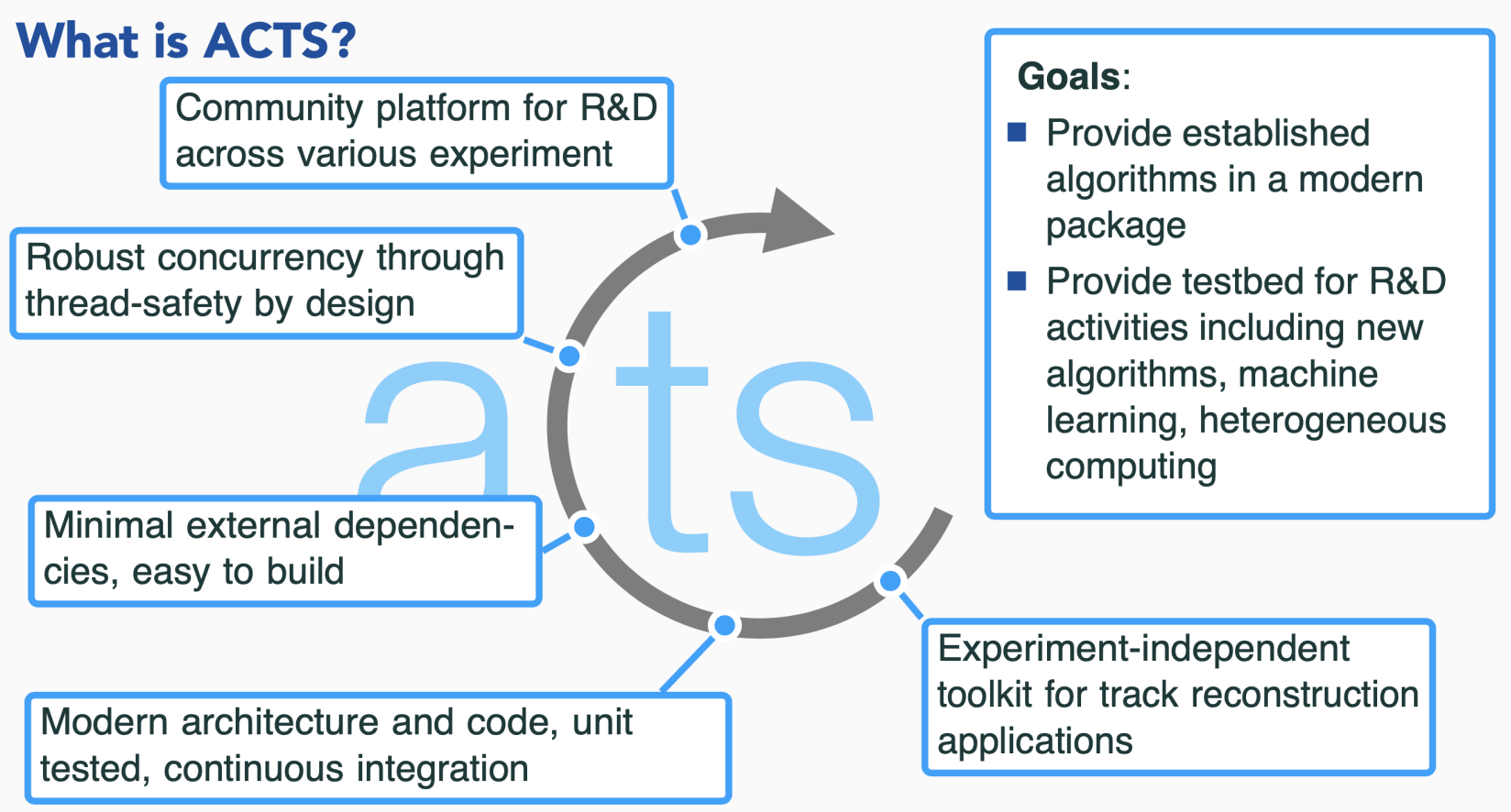
- **Toward common software solutions** for multiple experiments/ communities
- **ACTS** (A Common Tracking Software) (HEP community)
  - **Experiment-independent** toolkit for track reconstruction
- **Gammapy** (astroparticle community)
  - A python package for high-level γ-ray astronomy based on **common data formats**
- Similar effort in the nuclear physics domain, where traditionally each group was developing its own software
- **AGATA software**
  - On-line processing of data produced with the AGATA detector
- **KaliVeda**
  - Data processing and analysis for the INDRA and FAZIA detectors
- **nptool**
  - Data processing (and simulation) for low energy experiments

The ACTS project

- Scientific challenges
  - The reconstruction of charged particle trajectories is one of the most complex and CPU consuming parts of event processing in HEP experiments
  - Even more challenging with the increased number of simultaneous collisions at HL-LHC and Future Circular Collider
- A Common Tracking Software (ACTS)
  - Experiment-independent toolkit for track reconstruction
  - Originally launched by the ATLAS tracking group in the 2010s
  - Today several contributors from CERN, LBNL, CNRS/IN2P3, University of Geneva, Eotvos Lorand University in Budapest and Max-Planck Institute for Physics

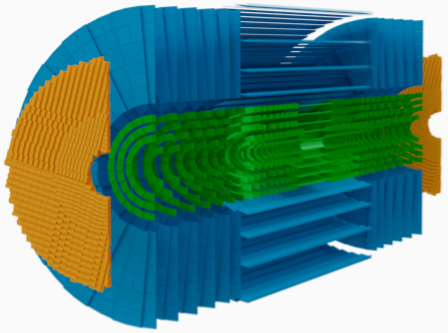


The ACTS project

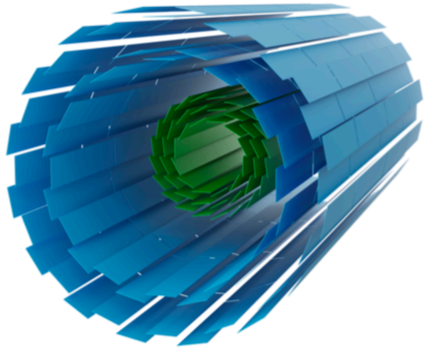


The ACTS project

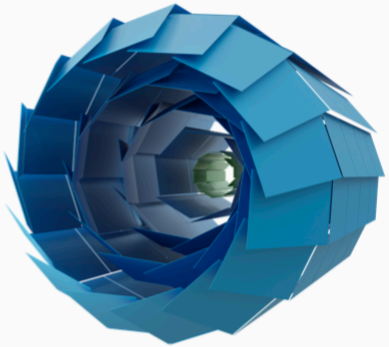
Evaluation and/or deployment by multiple experiments



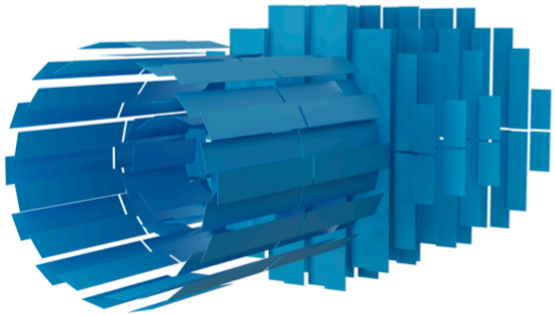
ATLAS Inner Tracker



sPHENIX

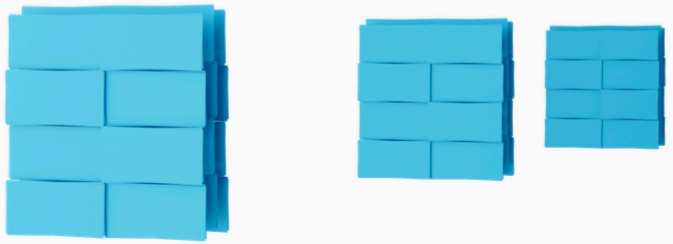


Belle II



Panda

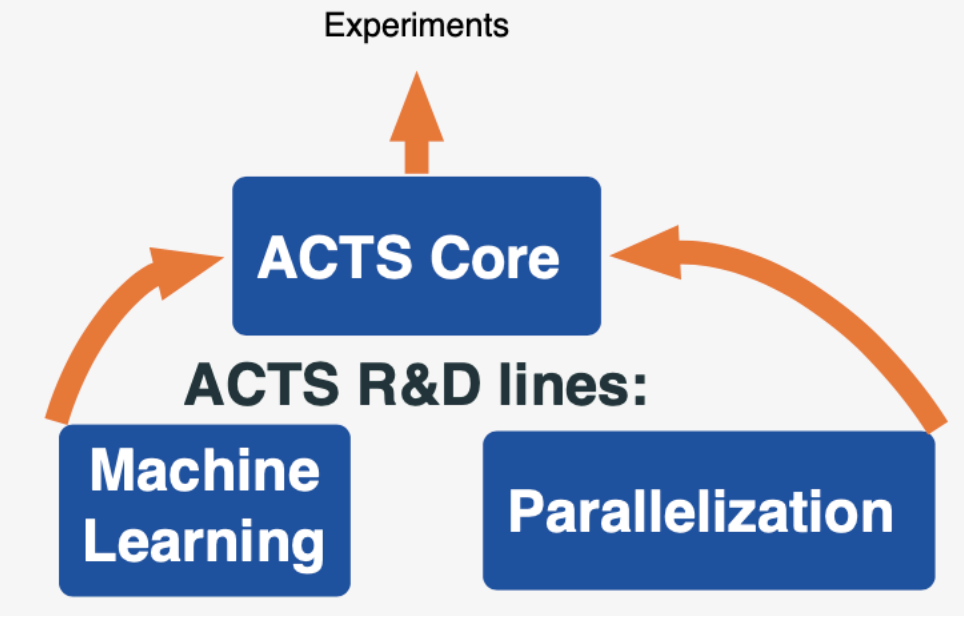
+ ALICE,
EIC



FASER

The ACTS project

- Mature project, ready for a progressive adoption within ATLAS
  - It will allow for an extensive validation
  - It will encourage its adoption by other experiments
- R&D activities



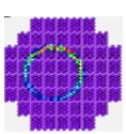
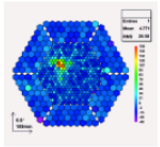
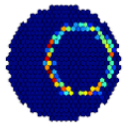
The ACTS project

- ACTS core development team ~ 10 persons, mainly at CERN and US universities
- IN2P3 contribution < 10% (1.8 FTE) in the past years, but now increasing
  - **Recent recruitment of 1 post-doc at IJCLab and 1 post-doc at LAPP**
- GPU R&D at IN2P3
  - Development of tracking algorithms suitable to run on GPU
  - The goal is to have a full reco chain demonstrator on GPU by the end of 2022
  - **Evaluation of SYCL** (GPU programming model w/o Nvidia lock-in) **(IJCLab)**
- Machine Learning R&D at IN2P3
  - **IJCLab** involved since long (**TrackML challenge**), more recently also LAPP and L2IT started to actively contributing
  - **AIDAinnova** : auto-tuning of tracking algorithm parameters, involves work on Python bindings **(CERN, IJCLab)**
  - **ATRAPP ANR** : Spatial hashing, integration of detector conditions, anomaly detection, … **(LAPP, IJCLab)**
  - Development of **Graph Neural Networks** tracking algorithms **(L2IT)**

Gammapy
A Python package for gamma-ray astronomy

Proprietary formats and tools

- All instruments have their own proprietary formats and tools
  - usually based on the open software ROOT from CERN
- Common joint analyses are impossible

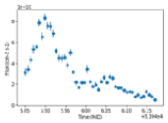
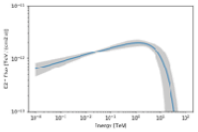
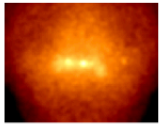




Raw data

H.E.S.S.

MAGIC

VERITAS

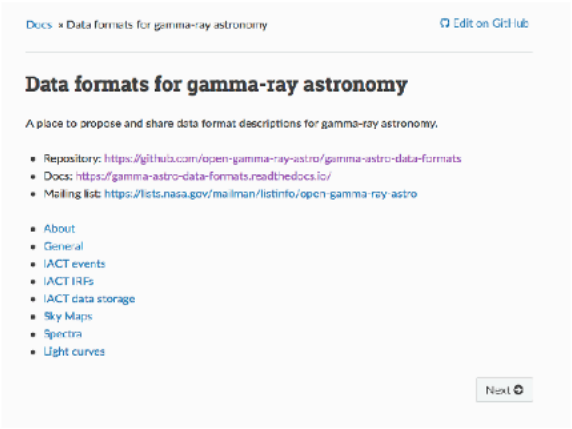
High level science products

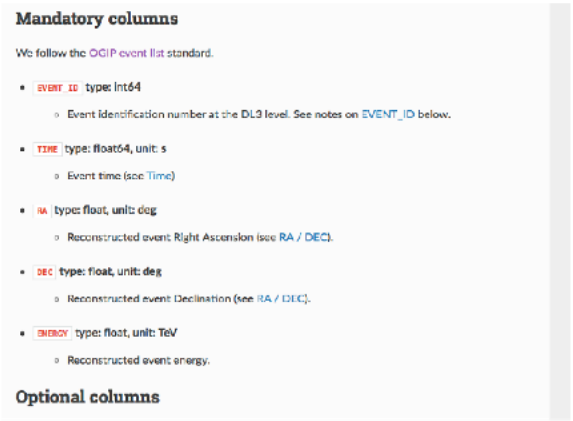
Interoperability & reusability require common open *data formats* and common open *tools*

Gammapy

Towards common data formats

- A community initiative to provide open data formats for VHE astronomy
  - g.a.d.f. (gamma-ray astro data formats) :a prototype format
  - VHE common formats initiative with people representing
    - pointing Cherenkov telescopes (e.g. CTA, Veritas, MAGIC etc)
    - drifting telescopes (e.g. HAWC, SWGO and Km3Net, IceCube)
    - Fermi-LAT





The data flow

Separating instrument specific data treatment from common use cases and methods

Reconstruction pipeline (internal) → Science Tools (general users)

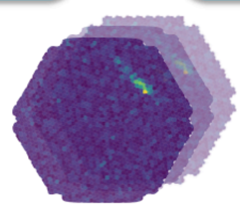
DL 0-1 — *raw - calibrated*

DL 2 — *reconstructed*

DL 3 — *filtered γ-like events*

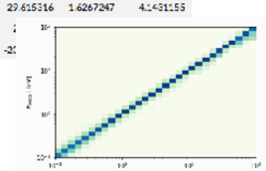
DL 4 — *science*

DL 5 — *high-level*

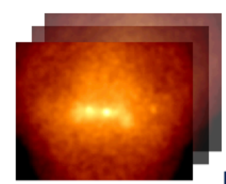


camera data

ENERGY	RA	DEC	L	B
MeV	deg	deg	deg	deg
float32	float32	float32	float32	float32
12186.642	260.45935	-33.553337	353.36273	1.7538676
25496.598	261.37506	-34.395004	353.09607	0.6520652
15621.498	259.56973	-33.409416	353.05673	2.4450684
12816.32	273.95883	-25.940391	6.45856	-4.0548879
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	-26.224436	[illegible]	[illegible]
13960.802	271.44742	29.615316	1.6267247	4.1431155
10477.372	266.3981			
13030.88	271.70128			



γ-like event lists
IRFs



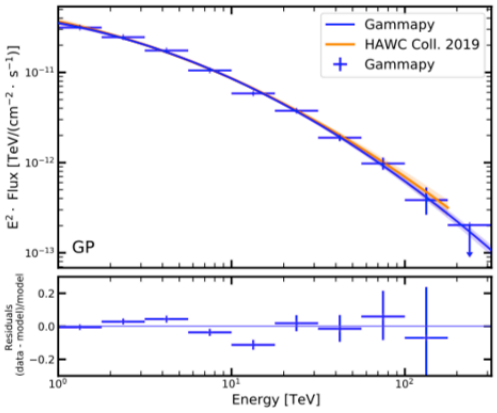
maps,
spectra,
light curves

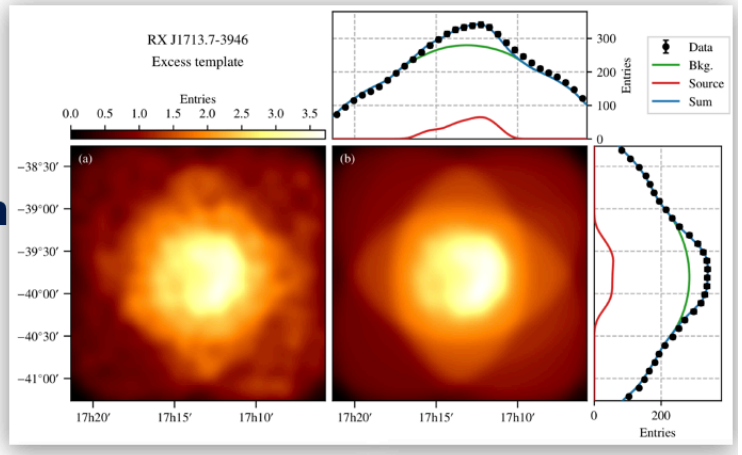
Gammapy

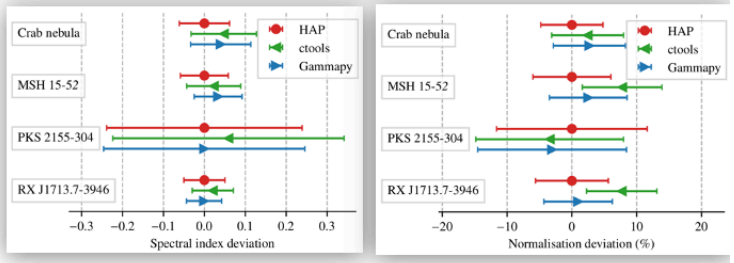
Science validation & maturity

Science usage validated on HESS DL3 test data release

- **Selected as science tool library by CTAO**
- **Now regularly used for science publications in HESS collaboration**
- **Prototype analyses with HAWC**







Olivera-Nieto et al (2021)

Mohrmann et al. (2019)

Gammapy

Gammapy project organisation

Coordination committee

Project managers
Bruno Khélifi at APC & ECAP

Lead developers
Régis Terrier at APC & MPI-K

Developers
About 10 regular contributors
More than 80 contributors in total

More formal structure of the project introduced to ensure development planning and improve interactions with experiments/observatories

About 30 FTE in total and 12.3 FTE at IN2P3 from 2015 to 2022
2 permanent staff at APC in the roles of Project Manager (B. Khélifi) and Lead developer (R. Terrier)

Data processing software for AGATA

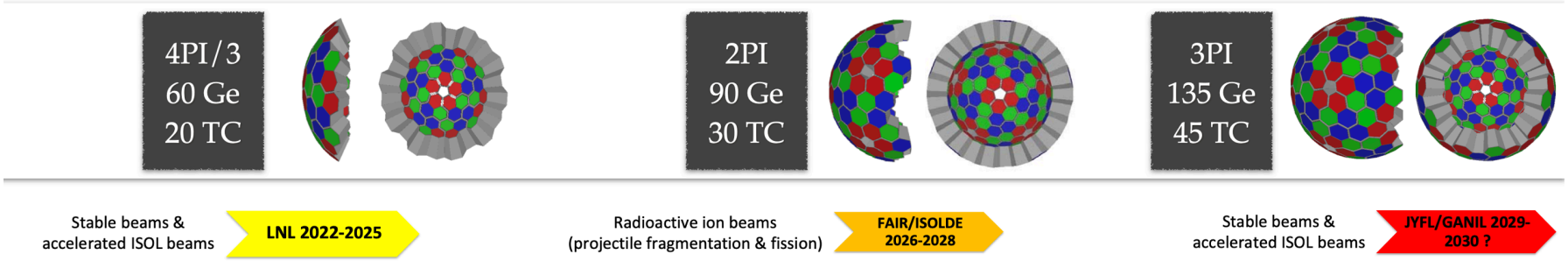
AGATA
Advanced Gamma-ray Tracking Array

Ultimate goal: tracking of gamma-rays in a Germanium 4Pi ball

AGATA is a nomad array built by an European collaboration (~12 countries)

MOU
Phase 1
2010-22

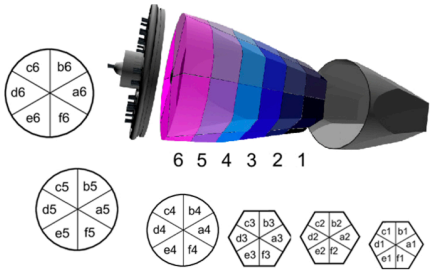
MOU - Phase 2 - 2022-30



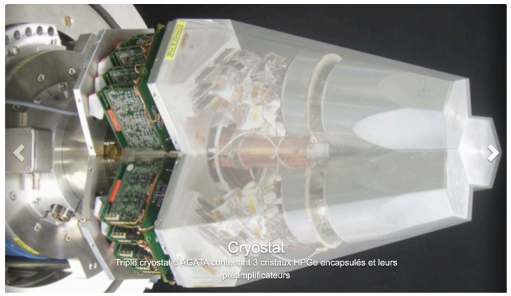
Data processing software for AGATA

AGATA : INGREDIENTS

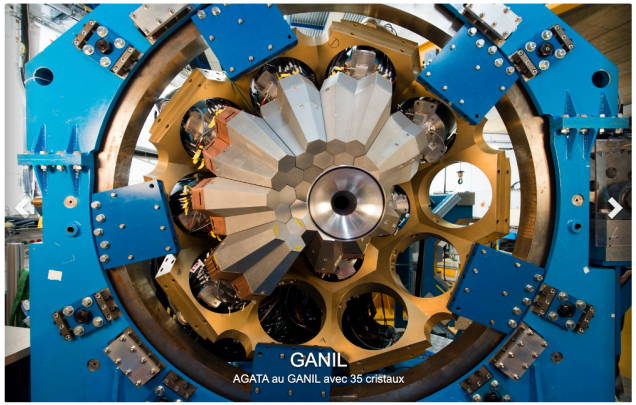
2 key algorithms PSA (Pulse Shape Analysis) and Tracking



High Purity Germanium crystal
36 segments + 1 core digitized signals



Cluster composed of 3 capsules



Assembly of clusters around the target

Global level data pipeline

N crystals

Event
hits of all crystals

Builder

ancillaries

Merger

Charged particles
Neutrons
Recoils etc …

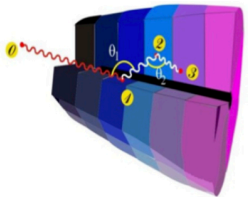
γ Tracking

Tracked gamma-rays

Local level data pipeline

37 signals

PSA

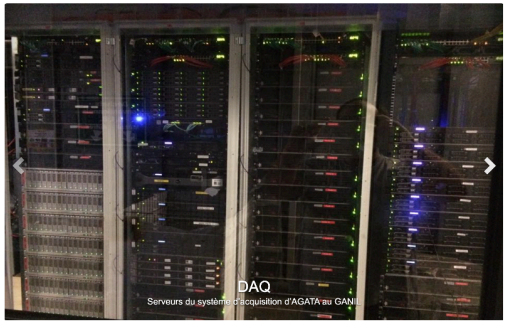


List of hits for one crystal

Data processing software for AGATA

AGATA : Computing (+R&D) @ IN2P3

The 2 key algorithms are continuously improved
Same for the harware and software infrastructure to run them



In charge of the Online Computing infrastructure

Services and computing nodes : ~ 70 servers
CEPH storage ~ 500 Tb, ~ 130 Tb (RAID mode)

Maintenance / upgrades / adaptation

Provide the online workflow management

DCOD written in ADA, developed @ IJCLab
Topology manager

Maintenance / upgrades / adaptation / R&D

IN2P3

Involved at many levels in the data pipelines

Code tracking dev. @ IJCLab
Online monitoring dev. @ IP2I
Production of root trees dev. @ IP2I
…

Maintenance / upgrades / R&D

IN2P3 involved since 2009 in AGATA

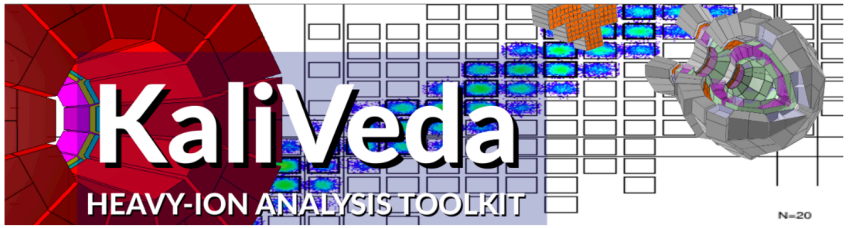
IN2P3 effort to software activitiy is ~ 5 FTE

Some key words on the IN2P3 activities

C++ / ADA / modern cmake / git / continuous integration / docker
code profiling / multithread / simulations / grid for storage …

R&D : Machine Learning (PSA/Tracking) / GPU / REDIS /
HPC / Time series databases …

KaliVeda Heavy-Ion Analysis Toolkit



C++ toolkit based on ROOT for analysis & simulation of Fermi energy heavy-ion collisions

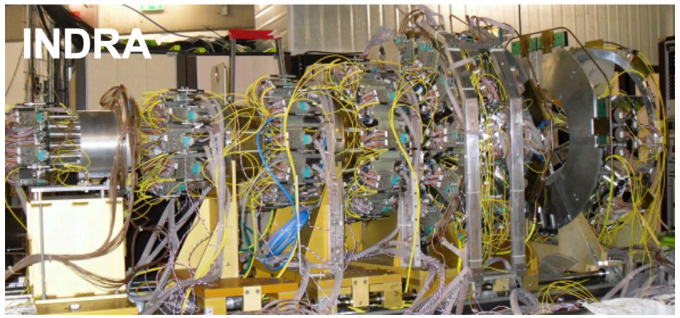
Main developers/maintainers:
John Frankland (CRHC GANIL Caen)
Diego Gruyer (CRCN LPC Caen)
Eric Bonnet (CRCN SUBATECH Nantes)

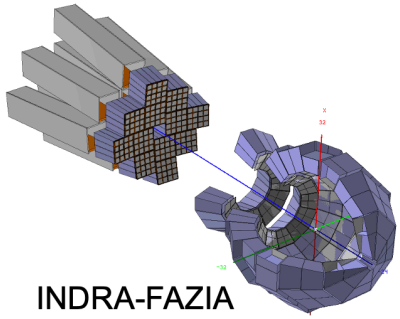
. Developed since 2002 to provide a unified environment for INDRA data reduction & analysis

- Since 2014, also used for the FAZIA array (and the INDRA-FAZIA coupling since 2019)
- Basis for all slow control GUI used during INDRA & FAZIA experiments

. Main features:

- Simulation & calibration of charged particle (ions) energy losses in detectors of different types
- Many tools for Z & A identification of ions using either ΔE-E or PSD techniques
- Single environment for analysis of all INDRA data since 1993 at CC-IN2P3 (batch) or on multi-core PC (PROOFLite)





INDRA-FAZIA experiment
E818 geometry (2022)

Workload, data and metadata management

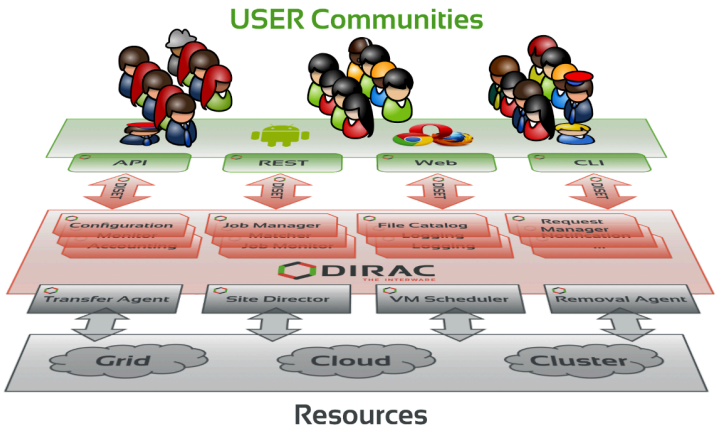
- The usage of large computing infrastructures and the production of large amount of complex scientific data requires **adavanced Management softwares**
- Solutions developed within LHC experiments, today adopted by several communities in HEP and beyond
- **DIRAC**
  - Workload Management (i.e. of computing tasks) in distributed and heterogenous computing environments (grids, clusters, …)
- **AMI**
  - A generic ecosystem for the handling of scientific metadata

The DIRAC project

DIRAC
THE INTERWARE

What's DIRAC?

- A software framework for distributed computing
- A **complete** solution to one (or more) user community
- Builds a layer between users and resources



- Started as an LHCb project, experiment-agnostic in 2009
- Developed by communities, for communities
  - Open source (GPL3+), GitHub hosted
  - Python 3 (python 2.7 kept for current production release)
  - Publicly documented, active assistance forum, yearly users workshops, open developers meetings and hackathons
- The DIRAC consortium as representing body

The DIRAC project

DIRAC
THE INTERWARE

Installations and communities

A *framework* shared by multiple experiments/projects, both inside HEP, astronomy, and life science

Experiment agnostic
Extensible
Flexible

The DIRAC project

IN2P3 Master Project

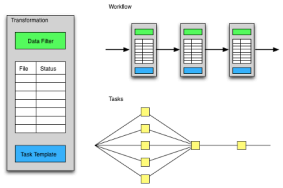
- The project started in 2017
  - Scientific Coordinator: A.Tsaregorodtsev, **CPPM**
  - Technical Coordinator: J.Bregeon, **LPSC**
- Participants
  - IN2P3: **CPPM, LUPM, LPSC, IPHC, CC/IN2P3**
    - **12 research engineers for 2.3 FTE**
  - CNRS/INSERM/INSA/U.Lyon/U.St.Etienne: **CREATIS**
- Goals
  - Federate the French actors around DIRAC
  - R&D for the needs of new scientific communities and technology advances

The DIRAC project

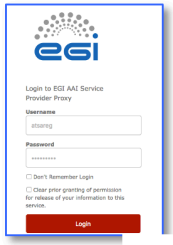
DIRAC
THE INTERWARE

Selected developments

- Production System for high level workflow management of large scientific communities
- Integration of various computing resources: cloud, HPC
- Support for OAuth2/OIDC based AAI systems (EGI Check-In, WLCG IAM)
- Application of Go language for distributed computing systems



openstack



slurm
workload manager

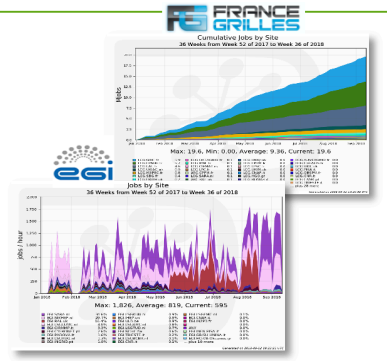
GO

The DIRAC project

DIRAC
THE INTERWARE

Collaboration

- EGI Workload Manager Service
  - Hosted at CC/IN2P3
  - ~10M jobs per year
  - ~40 VOs, 700 registered users
- DIRAC user's support
  - Support for DIRAC services: IHEP, JINR, …
  - CTA, HESS, biomed, Eiscat 3D, WeNMR …
- DIRAC service in the ESCAPE Project catalogue
  - Concordia for running containerized Corsika applications
- Coordination of the DIRAC Consortium
  - Membres: CNRS, CERN, IHEP, KEK, Imperial College, University of Montpellier
  - Managing software releases, web site
  - DIRAC Users' workshops



ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

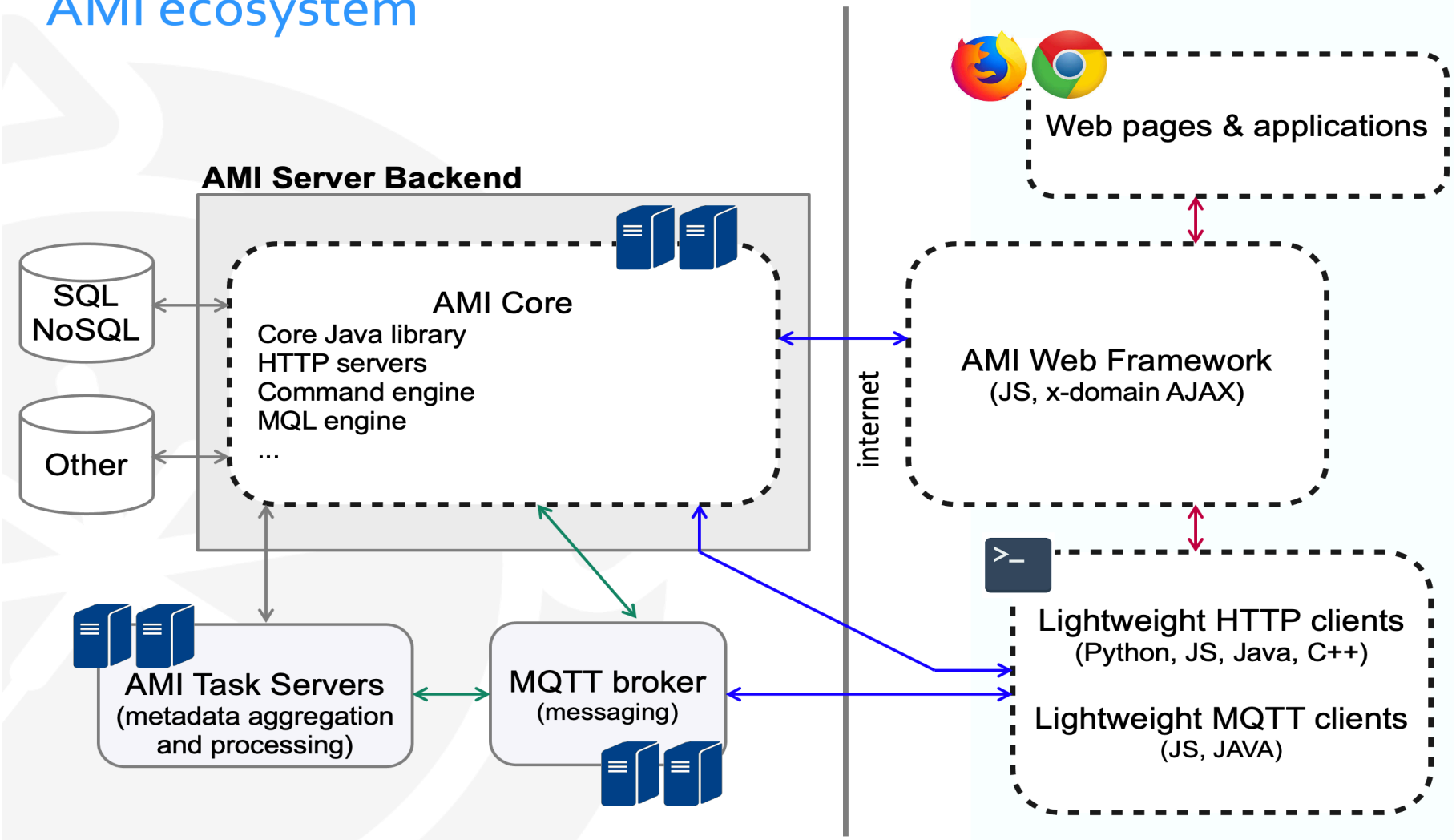
The AMI project

What is AMI?

- **AMI (ATLAS Metadata Interface) is a generic ecosystem for scientific metadata**
  - Complete solution from end-user applications down to DataBases including
    - Primitives for metadata extraction and processing
    - High level tools for selecting data by metadata criteria
    - Users base management…. and more
- AMI originally designed for ATLAS
  - Manage metadata for all ATLAS datasets and ATLAS software configurations **(~500M files, 260Gb per year, ~5K users)**
  - **Recent major re-write after ~20 years**
- Latest version entirely generic
  - Connect to any DB, deployable with docker
  - Used in other experiments : Nika2,… ?
  - **Designed for : Scalability, evolutivity and maintainability**

The AMI project

AMI ecosystem



The AMI project

Summary on AMI

- AMI (ATLAS Metadata Interface) is a generic ecosystem for scientific metadata
  - Complete solution from end-user applications down to DataBases

- **Actively developed at LPSC**
  - 3 FTE: J. Fulachier, F. Lambert, J. Odier (engineers) M. Jaume (Student)
  - Resources just enough for development and maintenance

- **Plans**

- Continue the development effort to cope with the needs of ATLAS for the Run3 and HL-LHC
  - Welcoming new users
  - Attract new developers

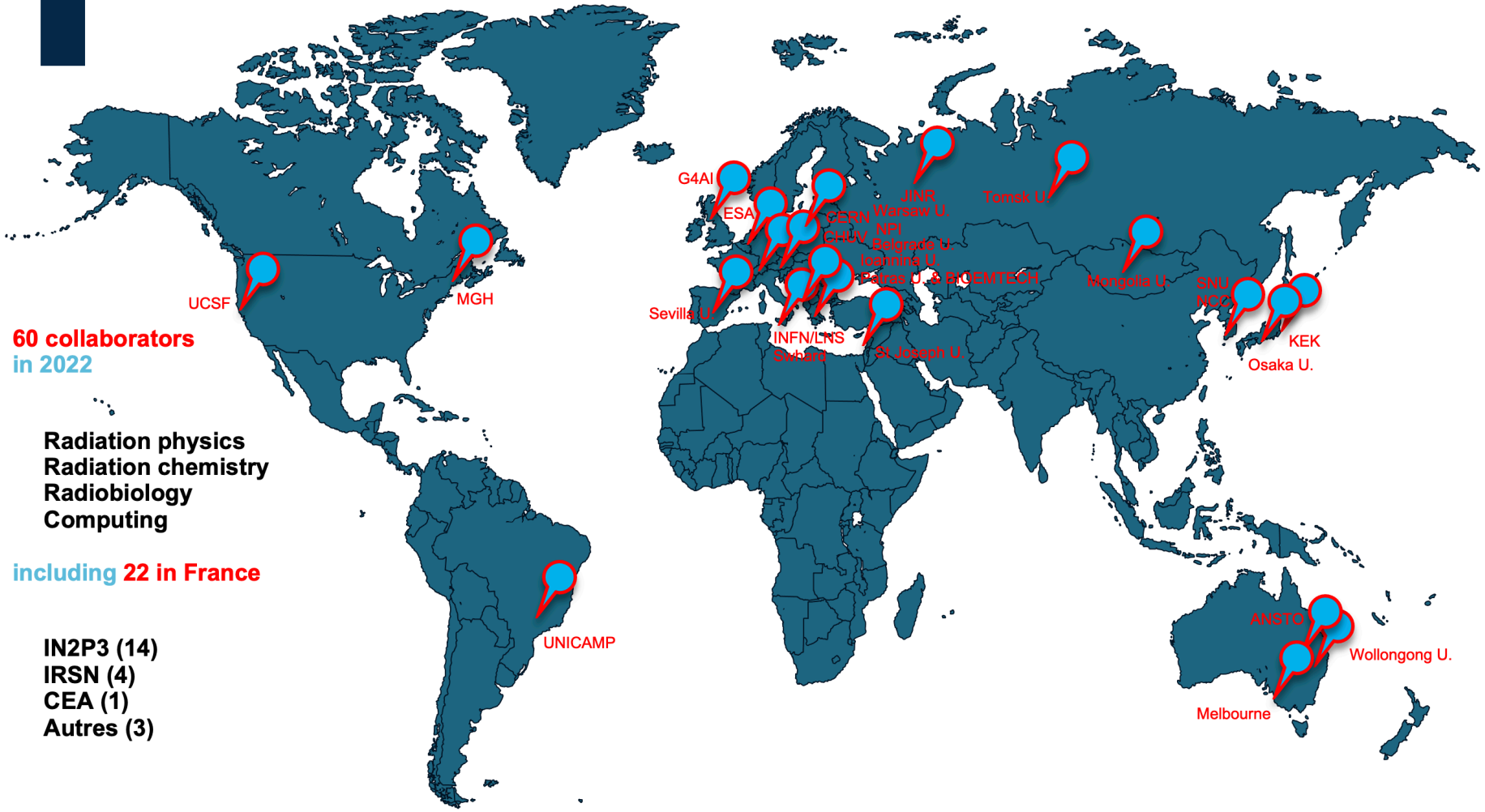
Conclusions

- Strong involvement of IN2P3 staff in several key projects in collaboration with our international partners
- Management responsibility in several projects
- Good synergies among projects around Geant4
- People very active also in dissemination and training actions
- Very large and worldwide-spread users communities
- IN2P3 staff also involved in several R&D activities
- R&D topics are common to several projects : machine learning, GPU, exploring new hardware, …
- Importance of the recruitment of permanent staff to effectively contribute to complex projects and to ensure the long-term development
- Importance of collaborations with “computing experts” (computer scientists, computing centers, manufactures, …)

Backup

The Geant4-DNA project

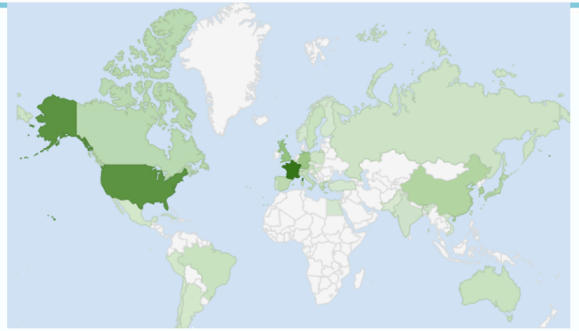
An international collaboration



- **60 collaborators in 2022**
  - Radiation physics
  - Radiation chemistry
  - Radiobiology
  - Computing
- **including 22 in France**
  - IN2P3 (14)
  - IRSN (4)
  - CEA (1)
  - Autres (3)

The GATE project

A large community of users



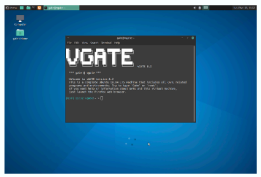
Around 2000 registered users

- Large communities in France and USA
- Increasing community: Canada, UK, Australia, Germany, China, Japan

Open source and open access platform available on Github

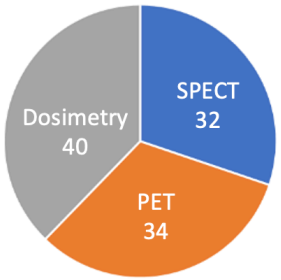
The source code: https://github.com/OpenGATE/Gate
The examples: https://github.com/OpenGATE/GateContrib
The tools for analysis: https://github.com/OpenGATE/gatetools

GATE version 9.2
1 release/year



Trainings & Workshops

- 1 workshop / conference of interest: IEEE NSS-MIC, AAPM, MCMA, PTCOG…
- Trainings for beginners: 1/year and @ companies
- Advanced trainings: 1/year (Python data analysis…)



Papers – Period: 2015-Now (PubMed search)

The GATE project

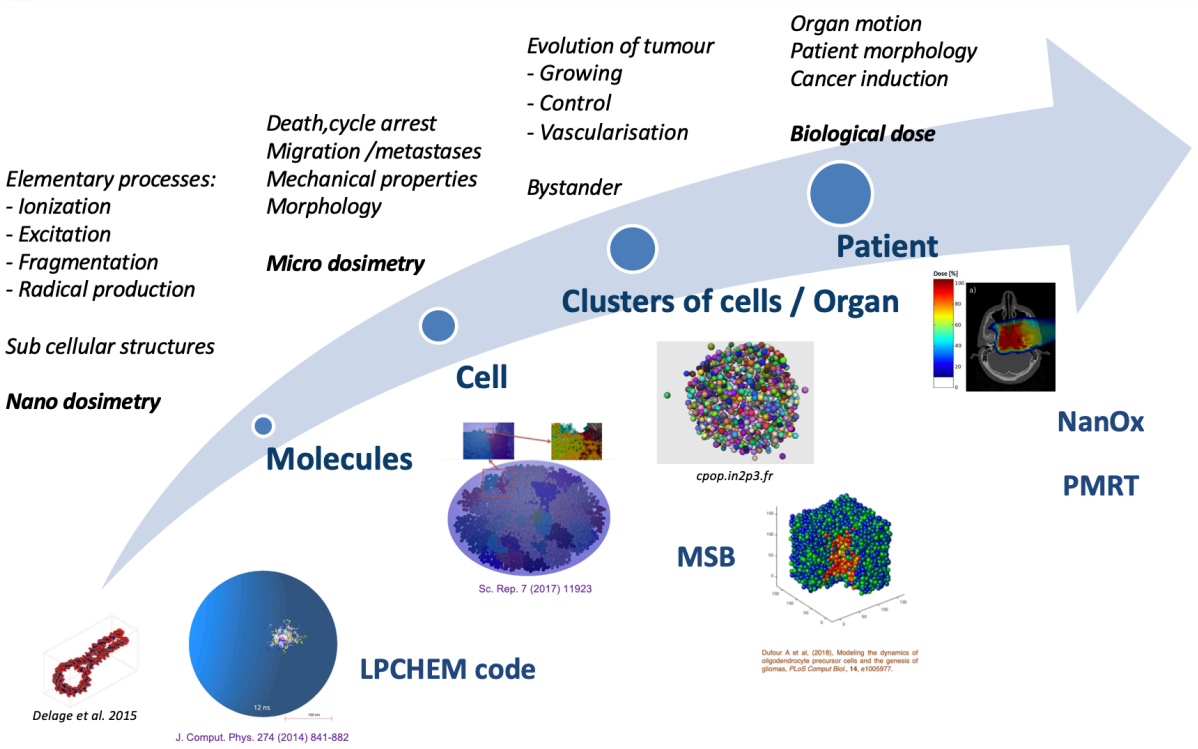
Multi-scale simulations

GEANT4-DNA

The **unique** open source and open access simulation toolkit for micro/nano dosimetry and radiation biology

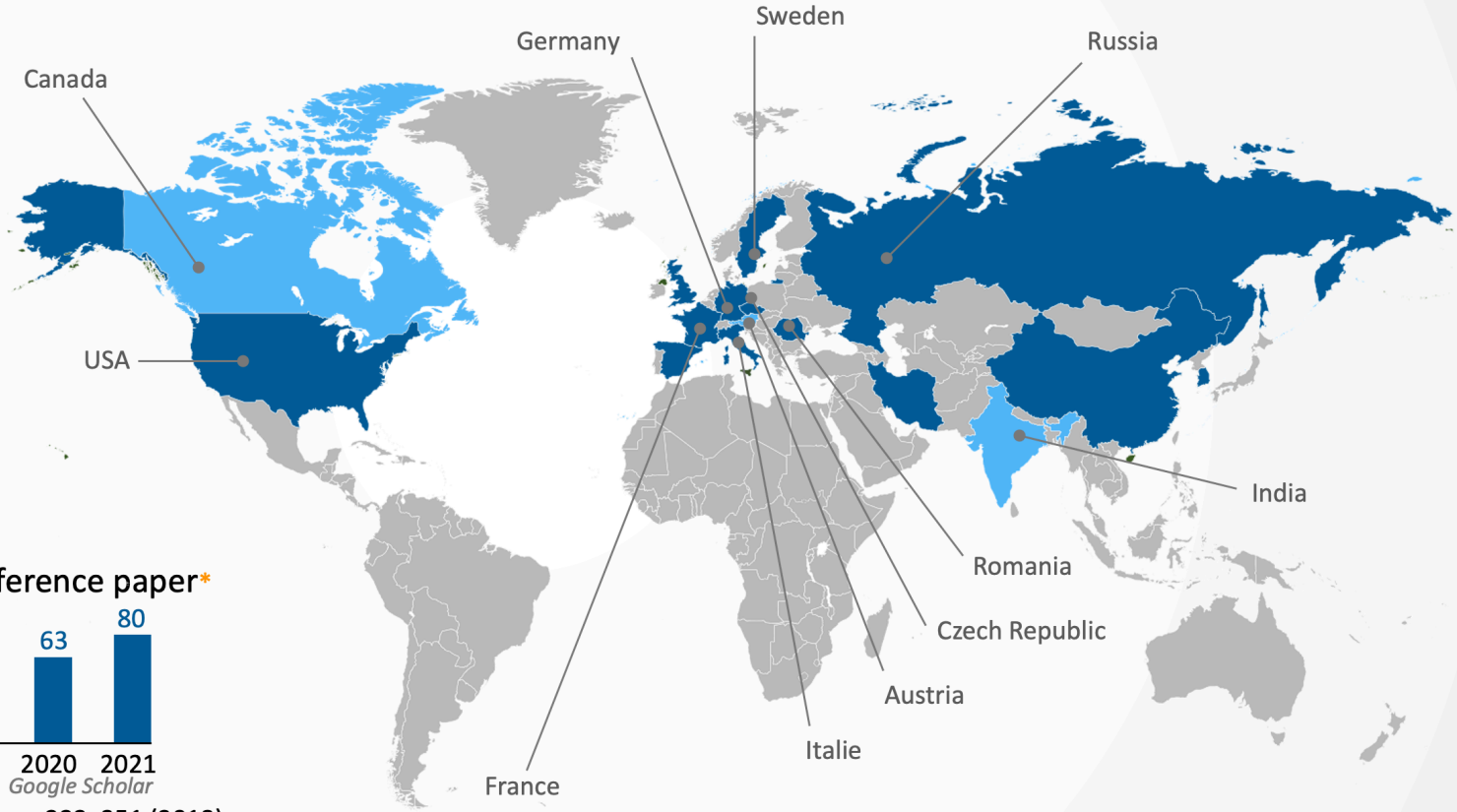
- **Long term development** fully included in Geant4 releases
- International collaboration composed of 42 collaborators
- **Coordinated by IN2P3/CNRS** since 2008
- Funded by regular support from institutions and international calls
- Fruitful involvement in international conferences & tutorials
  - Geant4 International User Conference at the Physics-Medicine-Biology frontier » series of conferences initiated by IN2P3 in 2005
  - Annual international tutorials (17)
- High rank and highly cited publications (104 since 2007)

DEVELOPMENT ACCESSIBLE TO OTHER TOOLKITS, PARTICULARLY TO GATE

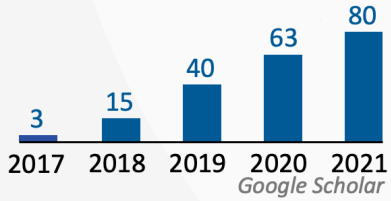


The Smilei project

Smilei's user community is international & steadily growing



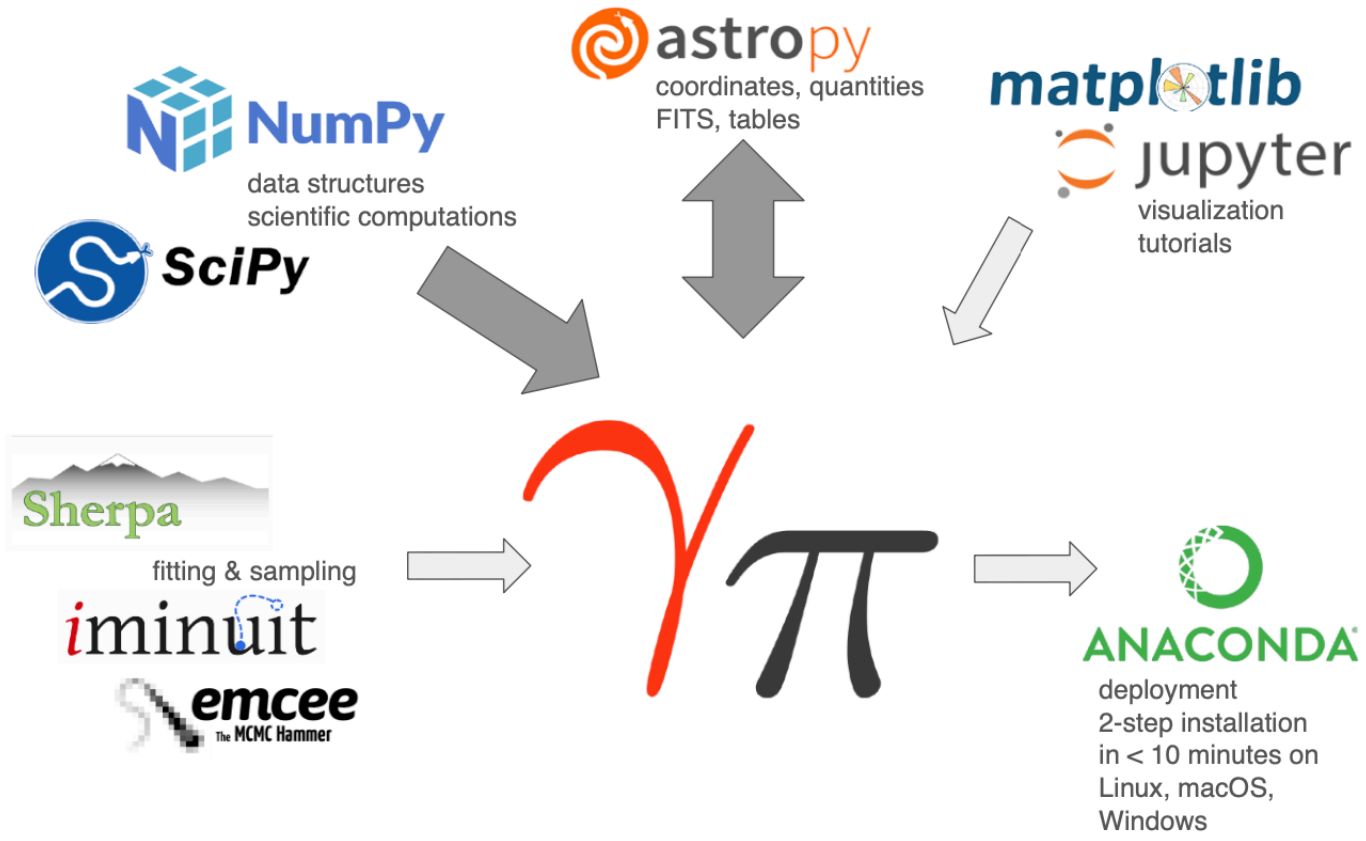
200+ citations for Smilei reference paper*



*Dérouillat et al., Comp. Phys. Comm. 222, 351 (2018)

Gammapy

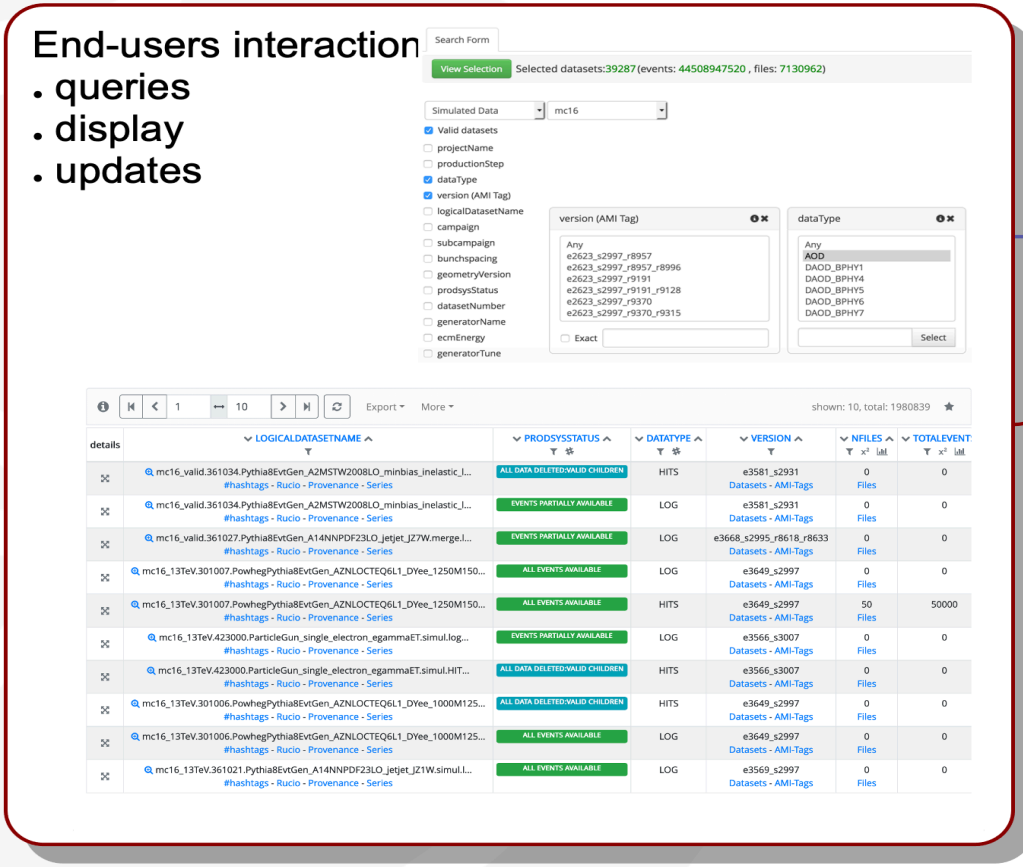
Gammapy in the Python ecosystem



The AMI project

AMI ecosystem

End-users interaction
- queries
- display
- updates



internet

Web pages & applications

AMI Web Framework
(JS, x-domain AJAX)

The AMI project

Major AMI feature : MQL

Metadata Query Language

- Query language dedicated to metadata
  - Simpler than SQL
  - Independent of underlying DB organisation
- Unique implementation in AMI
  - AMI has **reflexion** capability (automatic discovery of existing BD structures)
  - Internal translation from MQL → SQL

```
SELECT * WHERE (`AMISTATUS` = 'VALID')
               AND
               [`DATASET_KEYWORDS`.`KEYWORD` = 'stau']
               AND
               [`KEYWORD` = 'stop']
```

MQL to SQL

```
SELECT * FROM `ATLAS_AMI_MC16_02`.`DATASET`
WHERE
  (
    `ATLAS_AMI_MC16_02`.`DATASET`.`AMISTATUS` = 'VALID'
  )
  AND (
    `ATLAS_AMI_MC16_02`.`DATASET`.`IDENTIFIER` IN (
      SELECT
        `ATLAS_AMI_MC16_02`.`DATASET`.`IDENTIFIER`
      FROM
        `ATLAS_AMI_MC16_02`.`DATASET_KEYWORDS`,
        `ATLAS_AMI_MC16_02`.`DATASET`
      WHERE
        (
          `ATLAS_AMI_MC16_02`.`DATASET_KEYWORDS`.`KEYWORD` = 'stau'
        )
        AND `ATLAS_AMI_MC16_02`.`DATASET_KEYWORDS`.`DATASETFK` = `ATLAS_AMI_MC16_02`.`DATASET`.`IDENTIFIER`
    )
  )
  AND (
    `ATLAS_AMI_MC16_02`.`DATASET`.`IDENTIFIER` IN (
      SELECT
        `ATLAS_AMI_MC16_02`.`DATASET`.`IDENTIFIER`
      FROM
        `ATLAS_AMI_MC16_02`.`DATASET_KEYWORDS`,
        `ATLAS_AMI_MC16_02`.`DATASET`
      WHERE
        (
          `ATLAS_AMI_MC16_02`.`DATASET_KEYWORDS`.`KEYWORD` = 'stop'
        )
        AND `ATLAS_AMI_MC16_02`.`DATASET_KEYWORDS`.`DATASETFK` = `ATLAS_AMI_MC16_02`.`DATASET`.`IDENTIFIER`
    )
  )
```