Inserm

La science pour la santé
From science to health

OpenDose on EGI: producing reference dosimetric data for Nuclear Medicine on the European Grid Infrastructure

Gilles Mathieu (1), Maxime Chauvin (2), Axel Bonnet (3), Sorina Camarasu Pop (3), Isabelle Perseil (1), Manuel Bardiès (2)

(1) Inserm, DSI, Coordination de l'Informatique Scientifique, Paris, France
(2) CRCT, UMR 1037, INSERM, Université Toulouse III Paul Sabatier, Toulouse, France
(3) CREATIS, CNRS, INSERM, Université Claude Bernard Lyon I, INSA, Lyon, France

OpenGate meeting, December 13th 2018, Orsay

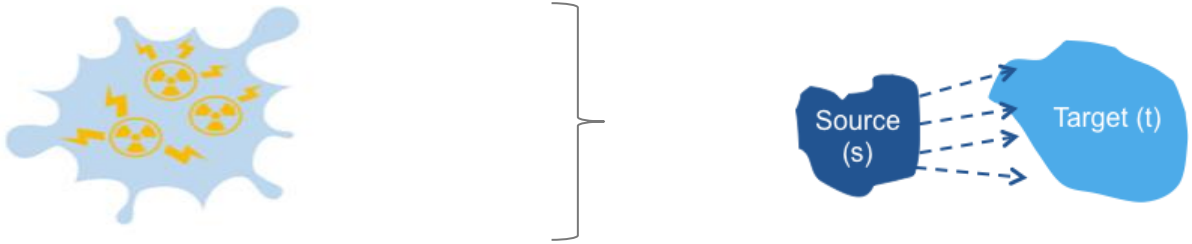
PLAN A

- Scientific view
- Technical view
- Foreseen solution
- Infrastructures
- Tools
- First results
- Conclusion

PLAN B

- Start the talk
- See how it goes
- ☺

Some context: what do we want to do?

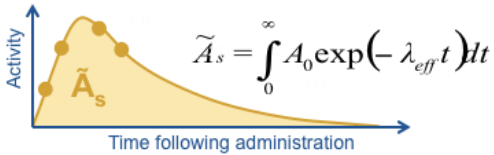


Tumoral cell

Organs

$$D_{t \leftarrow s} = \tilde{A}_s \times S_{t \leftarrow s}$$

Cumulated activity



$$\tilde{A}_s = \int_0^{\infty} A_0 \exp\left(-\lambda_{eff} t\right) dt$$

- Bio-kinetics and uptake distribution
- Patient-specific

S-value (dose factor)

$$S(r_t \leftarrow r_s) = \sum_i \frac{\Delta_i \Phi_i(r_t \leftarrow r_s)}{m_t}$$

- Nuclear decay data
- Source/target geometry

The OpenDose collaboration

Challenges

- 2 ICRP 110 reference adult phantoms (male and female) and more to come (pediatric and next generation mesh-based phantoms)
- 140 organs (19600 target/source combinations!)
- ICRP 107: ~1200 radionuclides
- MIRD RADTABS source of decay data: ~300 radionuclides

Too big for a single institution!!!

Proposal

- Collaborative work, everyone is welcome!
- Generate data with different Monte Carlo codes to cross-verify data
- Generate Specific Absorbed Fractions with associated uncertainties
- Traceable and reproducible data
- Create a free database
- Create an easily accessible website
- Compute S-values with uncertainties from SAFs

And what's the matter?

- **1 simulation = 1 Gate/Geant4 job**
  - `Gate -a [Source_ID,95][particle,gamma][energy,1][nb,10e5] main_AF.mac`
- **50 960 simulations**
  - 140 source organs, 2 particle types, 91 energy levels, 2 models
- **On a local cluster (240CPU) :**
  - 1 organ x 2 particles x 91 energies x 1 model = 1 day

More than 1 million CPU hours

Where we start from, where we go

- A computational model
  - Parametric, independent simulations
  - 50,960 simulations = 50,960 jobs
  - « Embarrassingly Parallel » problem

- A local cluster
  - 240 CPU
  - 280 days of computing

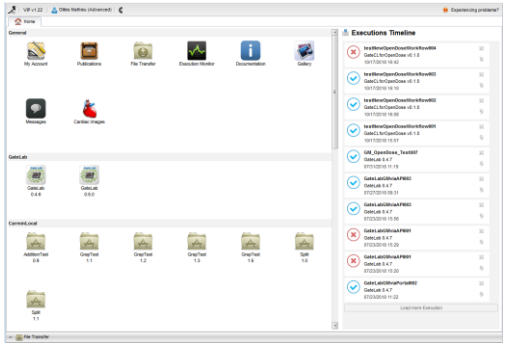
- Grid infrastructure
  - Adapted model
  - Resources ++
  - Available tools

Solution

- Infrastructure

Together, we'll be many!

- Tools



Gate/Geant4 on the grid

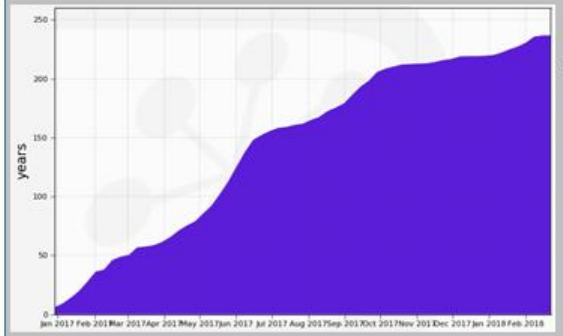
Web portal

Application as a service
File transfer to/from grid

VIRTUAL IMAGING PLATFORM

Infrastructure

Supported by EGI Infrastructure
Uses biomed VO (~65 sites in Europe and beyond)
230 cumulated CPU years utilized by VIP applications in 1 year

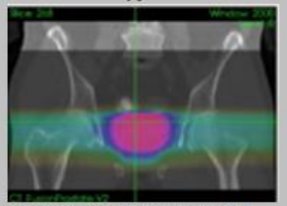


France-Grilles

DIRAC

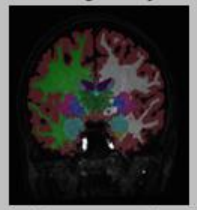
Scientific applications

Cancer therapy simulation



Prostate radiotherapy plan simulated with GATE(L. Grevillot and D. Sarrut)

Neuro-image analysis

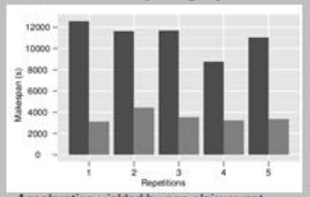


Brain tissue segmentation with Freesurfer

Image simulation

Echocardiography simulated with FIELD-II (O. Bernard *et al*)

Modeling and optimization of distributed computing systems



Acceleration yielded by non-clairvoyant task replication (R. Ferreira da Silva *et al*)

Users

1000+ registered users in January 2018
44 publications since 2011

GateLab

GateLab allows you to submit your GATE simulations on EGI from a simple web page:

- Upload your data and main macro file from the web browser.
- Select the estimated CPU time needed to complete your simulation.
- Choose the GATE release.
- After completion, results are automatically merged and ready for download.

OpenDose requires the new DoseByRegions output:

→ Need to update the GATE release to the latest version 8.1

→ Need to update the merger

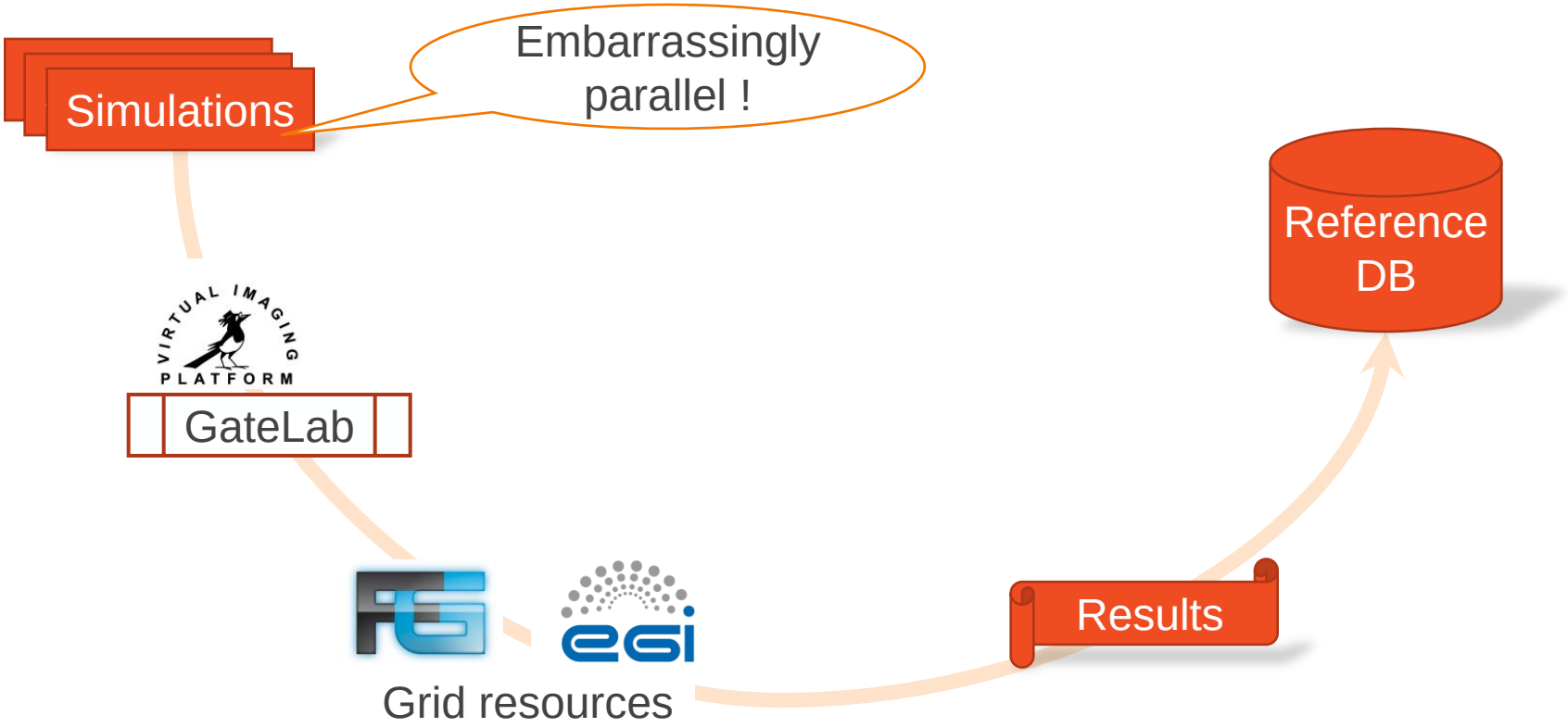
GateLab

- Update of the GATE release
  - DockerFile which build a new image with Geant4, ROOT and GATE on CentOS.
  - Script to extract the GATE binary and dependencies from the Docker image.
  - Add a env.sh to ease the setup of Geant4 and ROOT environment variables.

GATE release 8.1 is now available

- Update of the merger for the DoseByRegions output
  - Almost finished.

To summarise…



Nice. However…

- **Generate 50 960 configuration files?**

Thou shalt use parameters

→ A NEW WORKFLOW

- **Manually launch 50,960 jobs through the portal?**

Thou shalt script

→ A CLIENT CODE

Workflow description through Boutiques

Describe | Publish | Integrate | Execute

Inter-platform, command line applications.

- Polyvalent JSON format
- Linux containers

- https://github.com/boutiques

An OpenDose application using Boutiques

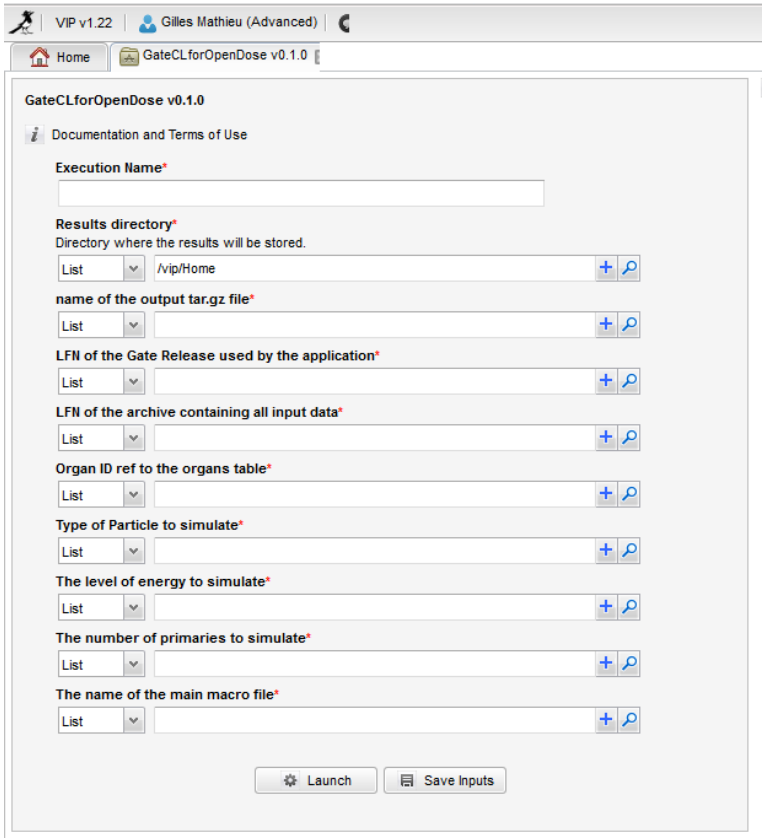
- A JSON file describing:
  - The command line itself

```
"command-line": "tar -zxf [GATERELEASE]; unzip [INDATA];

tar -zxvf gate_shared_libs.tar.gz; source ./init_env.sh;

./Gate -a [Source_ID,[ORGANID]] [particle,[PARTICLETYPE]] [energy,[ENERGY]] [nb,[NBPRIMARIES]]
[MACFILE] > output.log;

tar czf [RESULTS] ./output output.log",
```

  - Used parameters

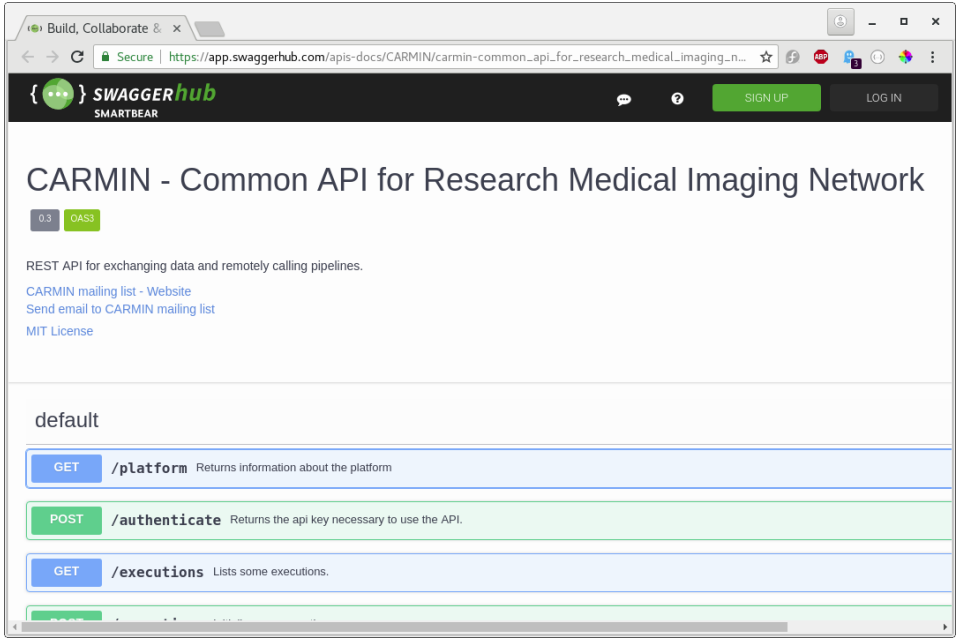
```
"inputs": [
{ "id": "gaterelease",
 "name": "LFN of the Gate Release used by the application",
 "optional": false,
 "type": "File",
 "value-key": "[GATERELEASE]"}, (…)]
```

An OpenDose application using Boutiques



The CARMIN API

- https://github.com/CARMIN-org
- https://app.swaggerhub.com/apis/CARMIN



An OpenDose client using CARMIN

```
public class OpenDoseClientClass {
  public static void main(String[] args) throws ApiException {
    /********** INIT API CLIENT ***********/
    ApiClient vipApiClient = new ApiClient();
    vipApiClient.setApiKey("thisisadummykey");
    vipApiClient.setBasePath("http://vip.creatis.insa-lyon.fr/rest");
    DefaultApi vipApi = new DefaultApi(vipApiClient);
    /********** SET PIPELINE AND LAUNCH EXECUTION ***********/
    Execution execution = new Execution();
    execution.setPipelineIdentifier("GateCLforOpenDose/0.1.0");
    Map<String, Object> inputValues = new HashMap<>();
    inputValues.put("indata", gateInputLFN);
    inputValues.put("gaterelease", gateReleaseLFN);
    inputValues.put("nbprimaries", numberOfPrimaries);
    inputValues.put("macfile", macFileName);

    // loop over all organs found in input matrix
    for (String organ : organsList) {
        inputValues.put("organid", organ);
        // loop over all particle types
        for (String particle : particlesList) {
               inputValues.put("particletype", particle);
               // loop over all energy levels
                   for (String energy : energiesList) {
                       inputValues.put("energy", energy);
                       executionName = "OpenDose_" + organ + "_" + particle + "_" + energy;
                       inputValues.put("outfilename", executionName + ".tar.gz");
                       // setup and launch execution
                       execution.setInputValues(inputValues);
                       execution.setName(executionName);
                       execution = vipApi.initExecution(execution);
                   }
        }
    }
  }
}
```

Initialisation

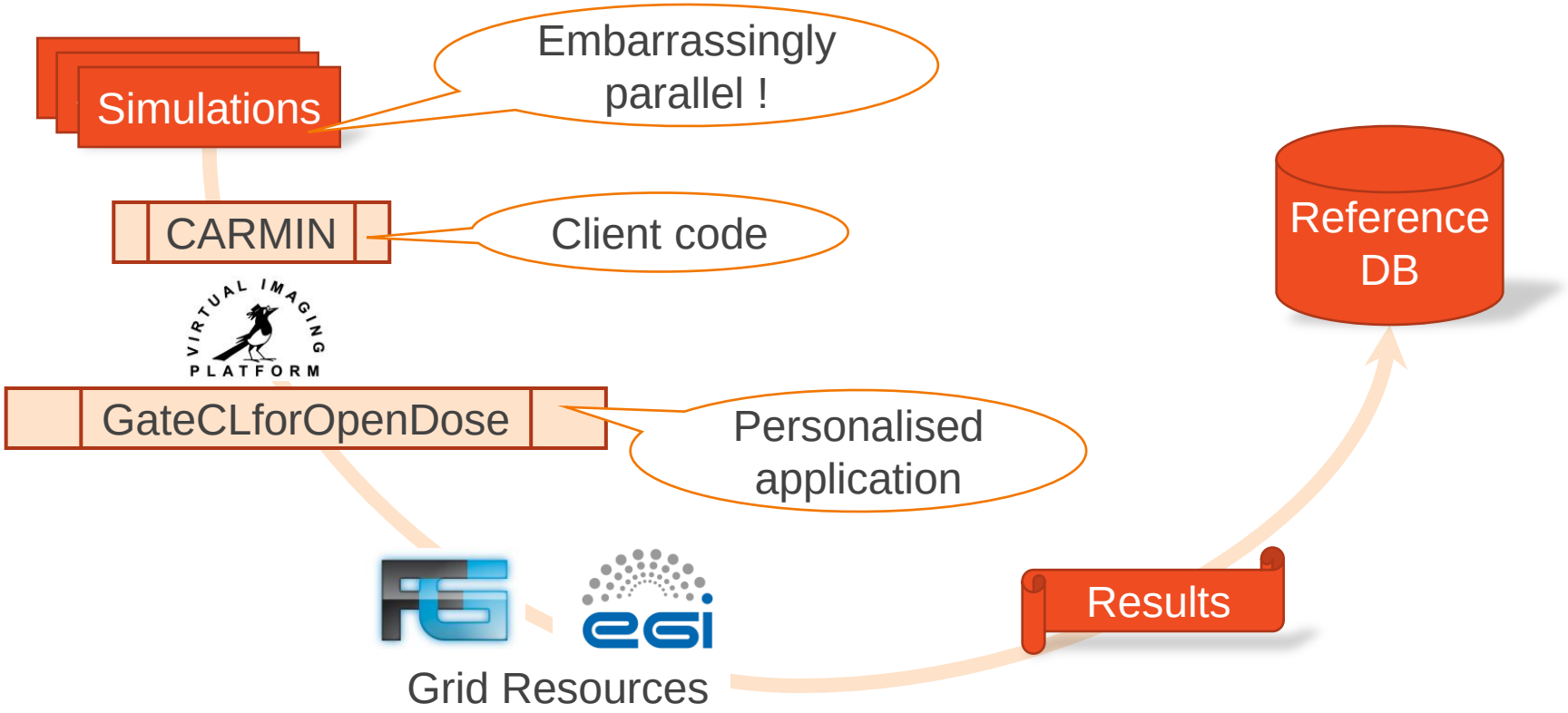
Application definition

Passing parameters

loops

1 organ
+ 1 particle
+ 1 energy level
= 1 submitted job

To summarise… (v2.0)



Where are we now?

- Global workflow analysis

- Submission tests through GateLab

- Draft client code

- Dedicated application deployed and working on VIP

- Production scale tests

- Needed adaptations!

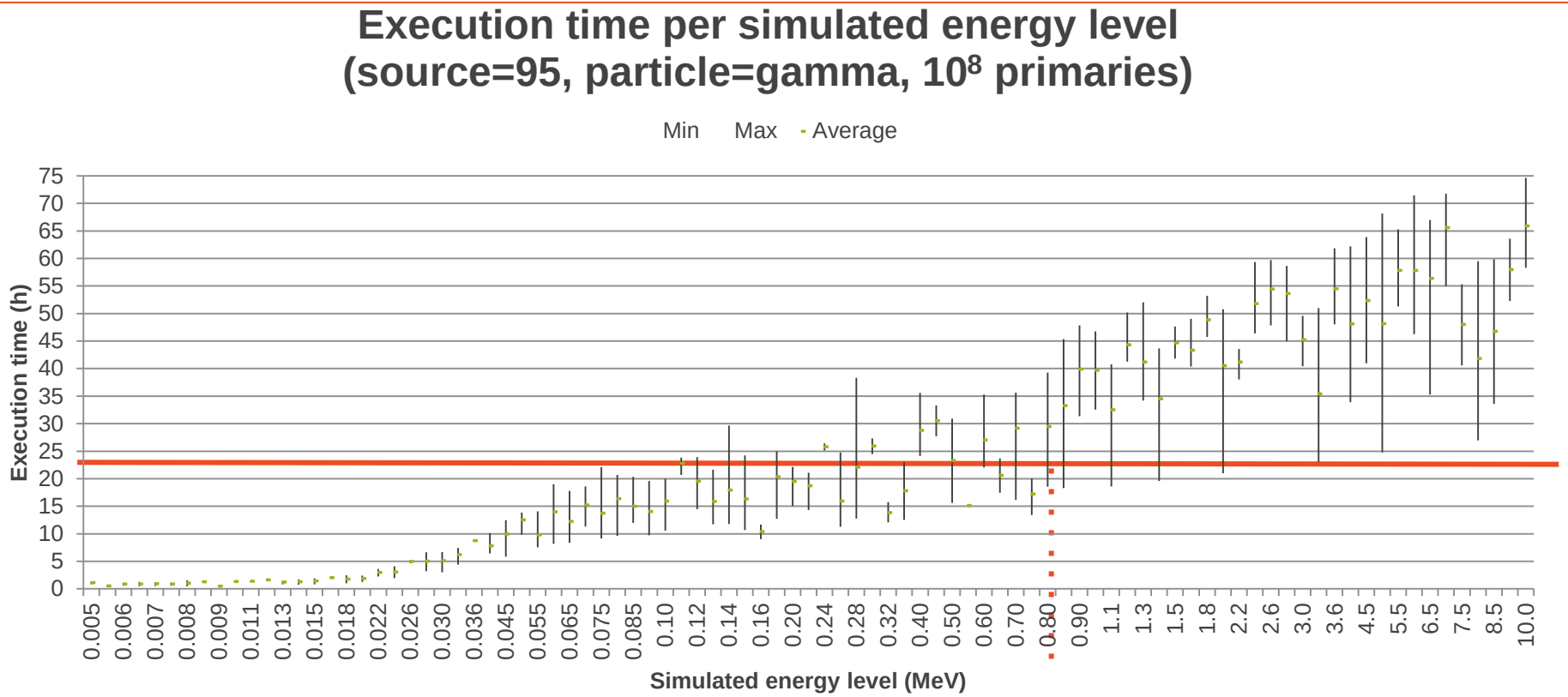
- Production

- Results gathering

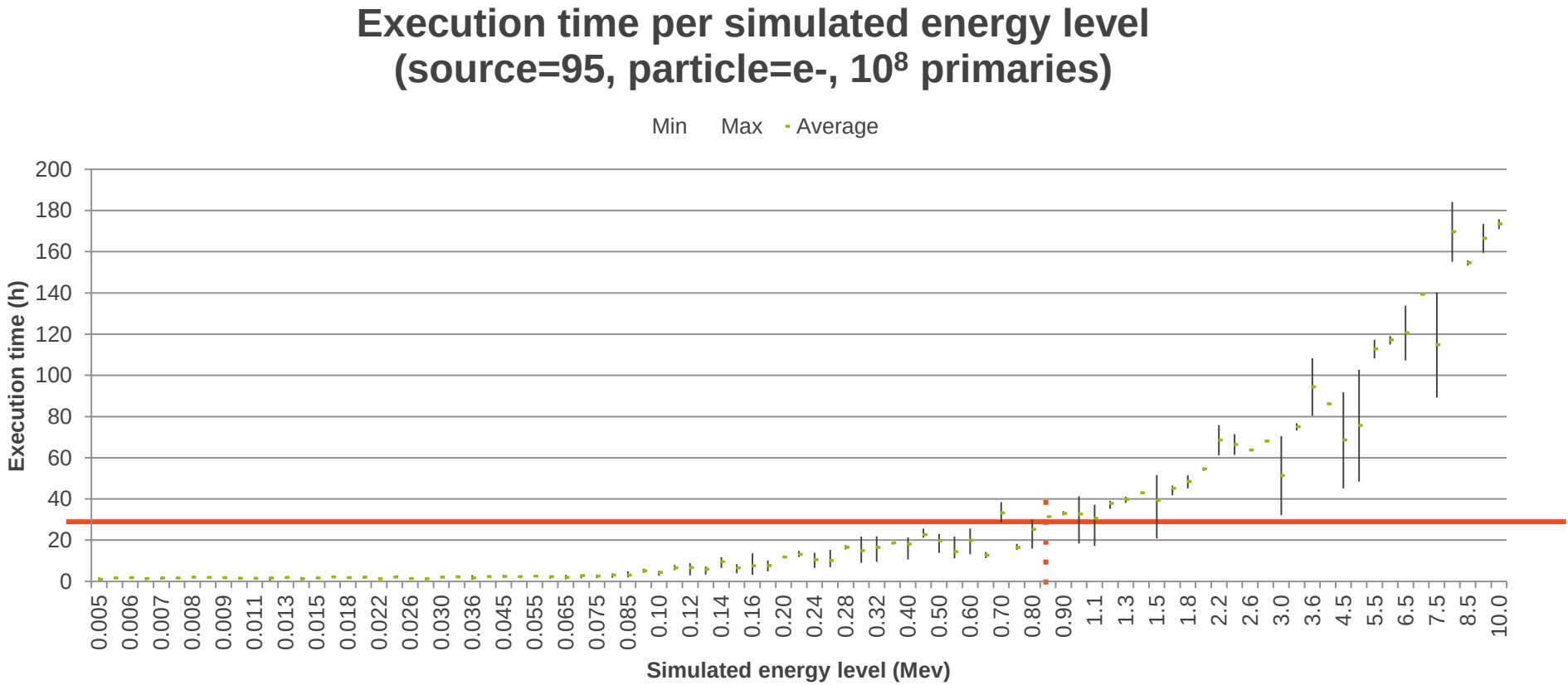
Needed adaptations ?

- Long jobs (>24h) don’t run well on the grid
- GOOD NEWS: GateLab has a mechanism to split jobs
- BAD NEWS: GateCLforOpenDose doesn’t

A quick look on execution times



A quick look on execution times



Houston, we have a problem

- Only 2/3 of the overseen production can run smoothly in current situation
  - Simulations with energy level < 0.8Mev
- We need to either:
  - Implement a split mechanism in our application
  - Modify GateLab
- WORK IN PROGRESS! Keep tuned ;-)

Meanwhile…

- We have started with REAL production anyway
  - All source organs, e- and gamma
  - Male and Female model

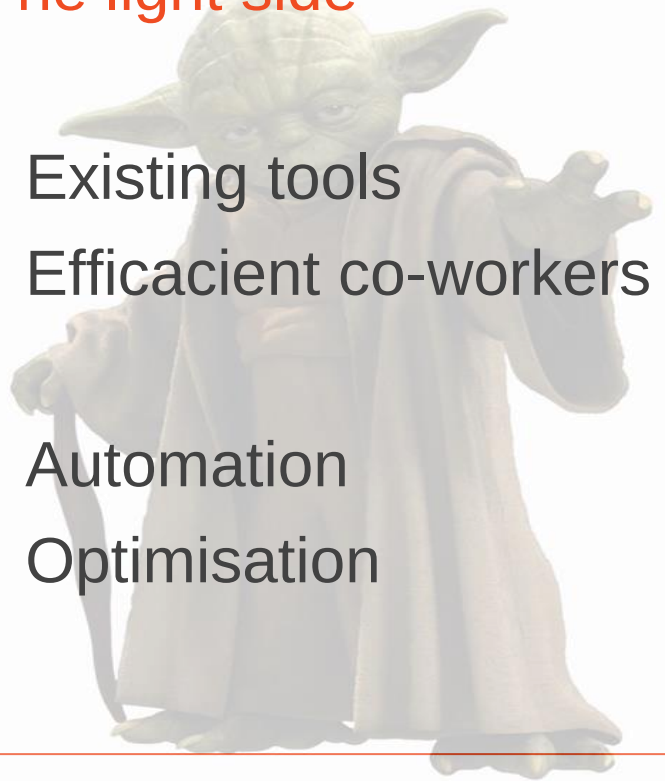
- Submission started Tuesday afternoon (11.Dec.)
  - Around 2000 simulations finished, and counting

A first feedback

The light side

- Existing tools
- Efficacient co-workers

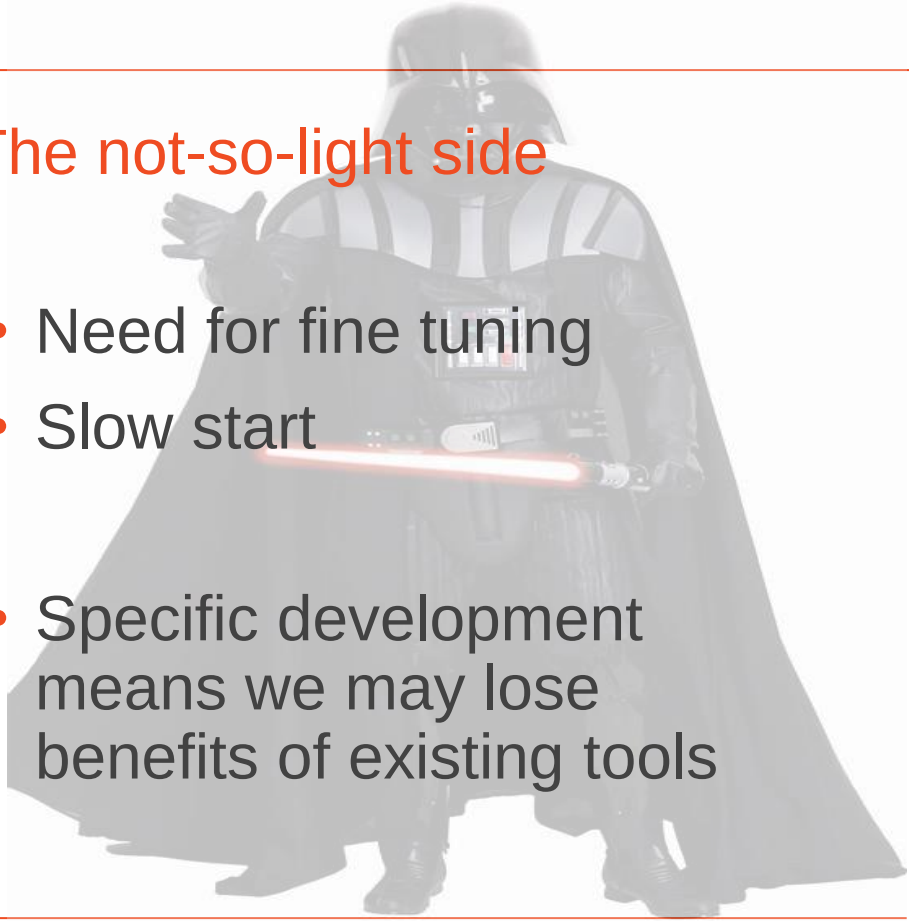
- Automation
- Optimisation



The not-so-light side

- Need for fine tuning
- Slow start

- Specific development means we may lose benefits of existing tools



To the future, and beyond

- **Industrialisation**
  - Optimise code and submission
  - Use more/different models
  - Ensure reproducibility et traceability

- **Scientific exploration**
  - Re-use submission framework
  - Test other MC softwares

- **Technical Exploration**
  - Other tools?
  - Other infrastructures?

Computing@OpenDose in a nutshell

Radioisotope

Tumoral cell

Models

Holy Grail!

Reference DB

Simulations

Nice tools

Grid

Results

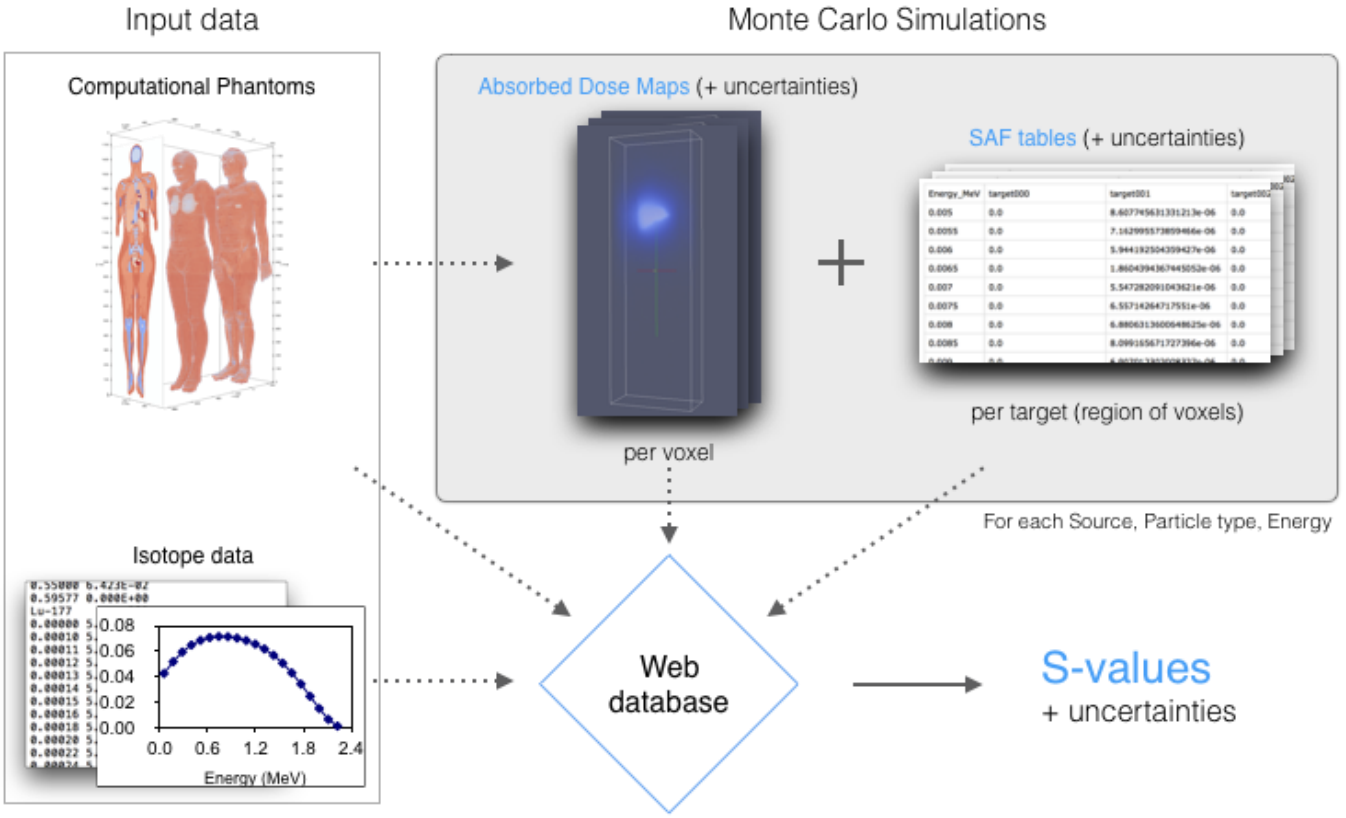
Inserm

La science pour la santé
From science to health

Annexes

Informations, links & references for an offline reading

The OpenDose collaboration



More on OpenDose

- MCMA2017 talk by M.Chauvin
  - http://people.na.infn.it/~mettivie/MCMA%20presentation/18%20Aula%20Magna/9_45_Chauvin.pdf

- MRTDosimetry2018 talk by M.Chauvin
  - http://mrtdosimetry-empir.eu/wp-content/uploads/2018/10/13_MRTdosimetryWorkshop_Chauvin.pdf

France Grilles

- Une offre de services sur une infrastructure informatique distribuée pour le calcul et le stockage de données scientifiques
- Une équipe d'animation au service de l'ensemble des acteurs qui coordonne les opérations de cette infrastructure basée sur les technologies de grille et de cloud informatiques
- Une communauté de plus d'un millier d'utilisateurs dynamiques, prêts à partager leurs connaissances et à s'entraider
- Le représentant de la France au niveau international au sein d'EGI
- Et surtout et avant tout un réseau résolument humain !
- www.france-grilles.fr

EGI

- www.egi.eu

Boutiques

- Décrire, publier, intégrer et exécuter des applications en ligne de commande inter-plateformes.
  - faciliter le portage d'applications
  - import et échange d'applications
  - science ouverte et reproductible

- Format JSON polyvalent pour décrire la ligne de commande, les entrées et les sorties
- Utilisation de conteneurs Linux pour faciliter l'installation et le partage d'applications
- https://github.com/boutiques

Portage des jobs OpenDose sur la grille

- Code client pour utiliser VIP via CARMIN
  - https://github.com/CISI-INSERM/OpenDose-Computing/tree/master/VIPclient

- Descripteur JSON pour portage de la ligne de commande OpenDose dans VIP via Boutiques
  - https://github.com/CISI-INSERM/OpenDose-Computing/tree/master/Boutiques

Crédits

- *CRCT, UMR 1037, INSERM, Université Toulouse III Paul Sabatier, Toulouse, France*
  - Maxime Chauvin, Manuel Bardiès
- *CREATIS, CNRS, INSERM, Université Claude Bernard Lyon I, INSA, Lyon, France*
  - Axel Bonnet, Sorina Camarasu Pop
- *INSERM, DSI, Coordination de l'Informatique Scientifique, Paris, France*
  - Gilles Mathieu, Isabelle Perseil