



# Protection des données sensibles sur le supercalculateur Jean Zay

JSO 2022

La science ouverte  
et les données de la recherche



« Avec  
Jean Perrin,  
nous créâmes  
le Centre  
national de  
la recherche  
scientifique »  
Jean Zay,  
29 avril 1942,  
Souvenirs  
et solitude

## Sommaire

1. JEAN ZAY : une collaboration IDRIS - GENCI
2. Qu'est-ce qu'une donnée sensible ?
3. Que dit la réglementation ?
4. Que doivent faire les projets de calcul concrètement ?

# 1. JEAN ZAY

## Une collaboration IDRIS - GENCI



« Avec Jean Perria, nous créâmes le Centre national de la recherche scientifique »  
Jean Zay,  
29 avril 1942,  
Soussens et solitude

## Unité d'Appui à la Recherche du CNRS (UAR851)

### Centre National pour

- Calcul numérique intensif de très haute performance (HPC)
- Intelligence Artificielle (IA)

### Activités

- Centre de calcul intensif Tier1  
équipé de supercalculateurs parmi les plus puissants du moment
- Centre de support, de conseil et d'expertise HPC et IA
- Formations (200 à 350 stagiaires/an)
- Datacenter

## Société civile

- détenue à 49 % par l'Etat représenté par le MESRI,
- 20 % par le CEA,
- **20 % par le CNRS,**
- 10 % par les Universités représentées par la Conférence des Présidents d'Université,
- 1 % par Inria

## Missions

- Mettre à disposition des moyens de calcul et de traitement de données massives performants
- Favoriser l'usage du calcul intensif associé à l'Intelligence Artificielle au bénéfice des communautés de recherche académique et industrielle

# JEAN ZAY

## Supercalculateur GENCI hébergé et exploité par l'IDRIS

Surface au sol : 320 m<sup>2</sup>

Masse : 70 tonnes

Consommation électrique ~2400 kWh

Refroidissement eau chaude à 32-40 °C

36,85  
Pflop/s

Puissance crête  
cumulée

3152

GPU NVIDIA  
2712 V100 16-32 Go  
440 A100 80 Go

89832  
cœurs

de calcul Intel Cascade  
Lake 2.5 GHz

454 To

de mémoire

100  
Gb/s

Réseau d'interconnexion  
Omni-Path

(1 lien par nœud scalaire  
et 4 liens par nœud convergé)

2,2 Po  
450 Go/s

Full Flash

35 Po  
150 Go/s

Disques rotatifs

Jean Perrin,  
nous créames  
le Centre  
national de  
la recherche  
scientifique >

Jean Zay,  
29 avril 1942,  
Souvenirs  
solidité

GENCI

IDRIS

ard

## 2. Qu'est-ce qu'une donnée sensible ?

 **GENCI**  
Le calcul intensif au service de la connaissance

 **Hewlett Packard Enterprise**

# Jean Zay

« Avec Jean Perrin, nous créâmes le Centre national de la recherche scientifique »

Jean Zay,  
29 avril 1942,  
Souvenirs et solitude

## **Données à caractère personnel**

- **DCP**  
= **Information relative à une personne physique susceptible d'être identifiée**
  - Directement ou indirectement
  - Confidentielle ou publique
- **Anonymisation**  
= **Rendre toute identification de la personne concernée impossible**
- **Traitement de données à caractère personnel**  
= **Opération portant sur des données personnelles**

## Données à caractère personnel

- **Identité** (nom, prénom, date de naissance, lieu de naissance, ... )
- **Vie personnelle** (habitudes de vie, loisirs, situation familiale, adresse postale ou électronique, numéro de téléphone, ... )
- **Vie professionnelle** (emploi, formation, diplômes, lieu de travail, adresse électronique, numéro de téléphone, ... )
- **Informations d'ordre économique ou financier** (revenus, impôts, données bancaires, droits sociaux, situation financière, ... )
- **Données judiciaires** (casier judiciaire)
- **Données de localisation** (coordonnées GPS, géolocalisation véhicule ou téléphone, badges bâtiments, télépéages, ... )
- **Données de connexion** (adresse IP, identifiant de connexion, cookie, ... )
- **Numéro d'identification** (numéro de carte d'identité ou carte vitale, ... )
- **Données non textuelles qui permettraient d'identifier une personne** (images, sons)

## Données sensibles

### Catégorie particulière des données à caractère personnel

- informations qui révèlent la prétendue origine raciale ou ethnique,
- opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale,
- traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique,
- données concernant la santé,
- données concernant la vie sexuelle ou l'orientation sexuelle

### 3. Que dit la réglementation ?

 **GENCI**  
Le calcul intensif au service de la connaissance

 **Hewlett Packard Enterprise**

# Jean Zay

« Avec Jean Perrat, nous créâmes le Centre national de la recherche scientifique »

Jean Zay,  
29 avril 1942,  
Souvenirs et solitude

# Données à caractère personnel

## Que dit la réglementation ?

### Le Règlement Général de Protection des données

- Obligation générale de sécurité et de confidentialité
- Obligation d'information
- Réalisation d'analyse d'impact
- Désignation d'un délégué à la protection des données (DPO)

### s'applique à tous les organismes qui procèdent au traitement de données à caractère personnel

- Responsables de traitement
- Les sous-traitants (hébergeurs, intégrateurs de logiciels, agences de communication entre autres)

### Le traitement par un sous-traitant est régi par un contrat ou un autre acte juridique

# Données à caractère personnel

## Que dit la réglementation ?

### Sanctions

- En cas de violation du règlement, la Cnil peut prononcer des amendes administratives qui peuvent atteindre, selon la catégorie du manquement, 2 % à 4 % du chiffre d'affaires annuel mondial de l'exercice précédent.
- C'est la Cnil qui adresse un courrier au responsable de traitement lui indiquant la sanction.
- Des sanctions pénales peuvent également s'ajouter à ces sanctions administratives. C'est le cas des infractions liées à la discrimination, des infractions de mise en danger en cas de mesures de protection insuffisantes (exemple : révélation de l'adresse d'une personne).

## Données sensibles

### Que dit la réglementation ?

#### Interdit de recueillir ou d'utiliser ces données, sauf :

- si la personne concernée a donné son consentement exprès (démarche active, explicite et de préférence écrite, qui doit être libre, spécifique, et informée) ;
- si les informations sont manifestement rendues publiques par la personne concernée ;
- si elles sont nécessaires à la sauvegarde de la vie humaine ;
- si leur utilisation est justifiée par l'intérêt public et autorisé par la CNIL ;
- si elles concernent les membres ou adhérents d'une association ou d'une organisation politique, religieuse, philosophique, politique ou syndicale.

## Données de santé Que dit la réglementation ?

### Article L.1111-8 du code de la santé publique

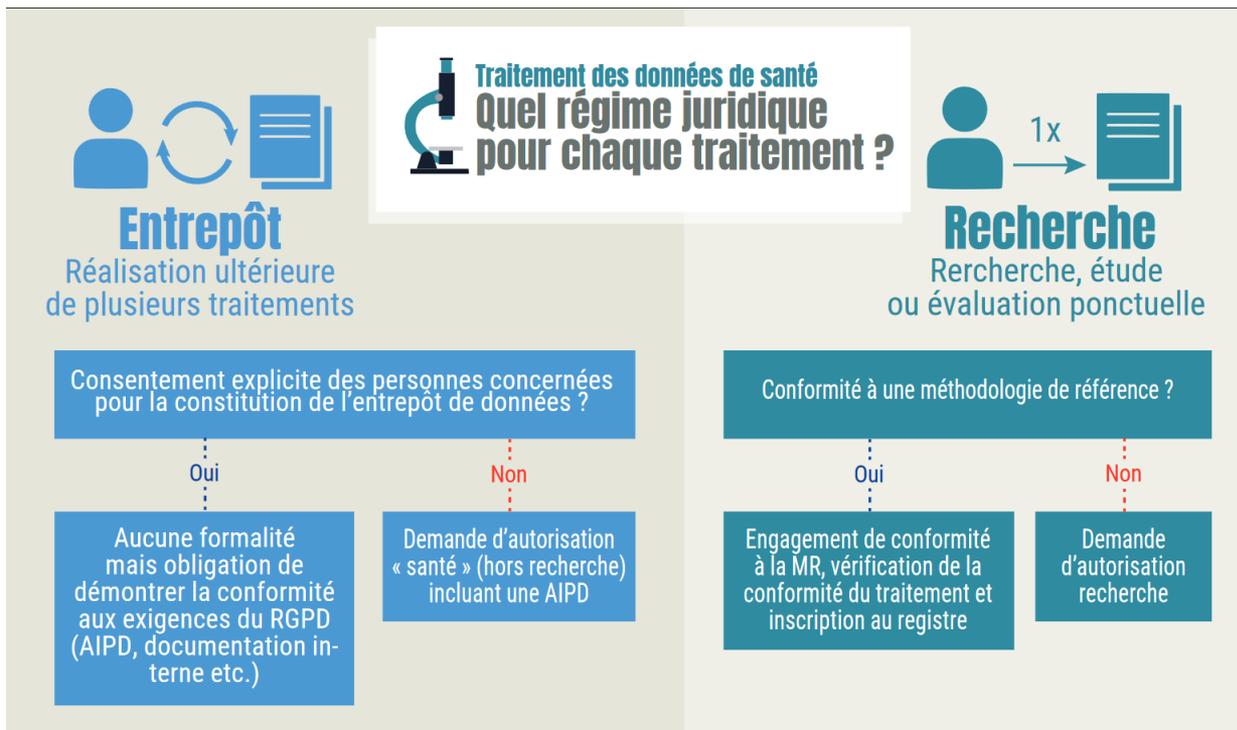
- « Toute personne physique ou morale qui héberge des données de santé à caractère personnel **recueillies à l'occasion d'activités de prévention, de diagnostic, de soins ou de suivi social et médico-social** pour le compte de personnes physiques ou morales à l'origine de la production ou du recueil de ces données ou pour le compte du patient lui-même, **doit être agréée ou certifiée** à cet effet »
- Obligation de recourir à un hébergeur certifié

### Et pour la recherche ?

- Données acquises à des fins de recherche
- Données pseudonymisées (ne permettant pas de relier une donnée à une personne)
- Alors, pas d'obligation à recourir à un hébergeur certifié

# Données de santé

## Que dit la réglementation ?



## 4. Que doivent faire les projets de calcul concrètement ?

 GENCI  
Le calcul intensif au service de la connaissance

 Hewlett Packard  
Enterprise

# Jean Zay

« Avec Jean Perrin, nous créâmes le Centre national de la recherche scientifique »  
Jean Zay,  
29 avril 1942,  
Souvenir et solitude

# **Demandez un accès aux moyens de calcul à GENCI**

**<https://www.edari.fr/>**

## **Votre projet va-t-il traiter des données à caractère personnel, dont des données sensibles comme notamment des données de santé ?**

- GENCI sera le sous-traitant
- Le CNRS (via l'IDRIS) sera le sous-traitant de GENCI

### **1. Identifiez le responsable de traitement**

→ à défaut le responsable du projet

### **2. Déclarez les données à caractère personnel traitées**

- Données détenues par un hôpital ou un médecin qui permettraient d'identifier une personne (nom, prénom, date de naissance, lieu de naissance, ? ) ?
- Données sensibles de santé (état de santé, données génétiques, données biométriques, etc.) ?
- Données acquises à des fins de recherche et pseudonymisées (sans aucune table de correspondance sur le centre)  
NON → obligation de recourir à un hébergeur certifié HDS

### **3. Formalisation de la sous-traitance**

- DPD + Responsable de traitement

## Respectez le RGPD

### En cas de demande de droit d'accès, de rectification, d'opposition, d'effacement

- VOUS gérez vos données

### En cas de problème de sécurité sur les données

- GENCI doit prévenir le responsable de traitement sous 72h
- Organisation de la gestion de crise entre GENCI, le responsable de traitement et l'IDRIS

# MERCI



**guillaume.harry@idris.fr**



**Campus Universitaire d'Orsay  
Rue John Von Neumann  
Bâtiment 506  
BP 167  
91403 ORSAY CEDEX**



**01 69 35 84 38**

**06 21 89 04 99**



[www.cnrs.fr](http://www.cnrs.fr)