

## ***FAIR un processus inné et/ou acquis ?***

C. Calvin, G. Camus

*CEA Direction de la Recherche Fondamentale*

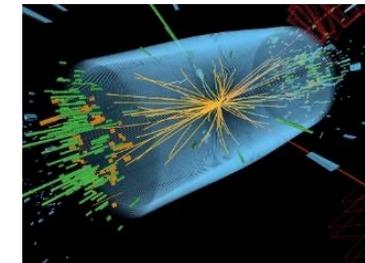
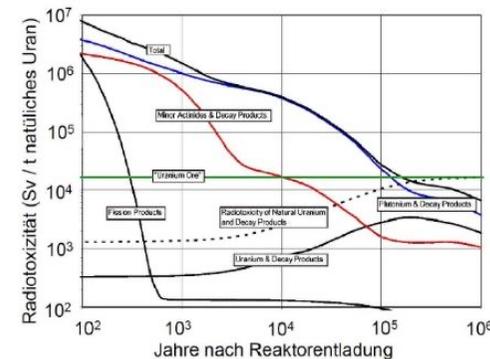
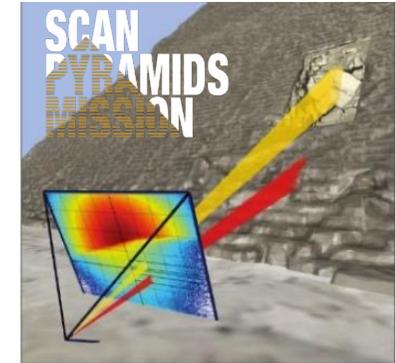
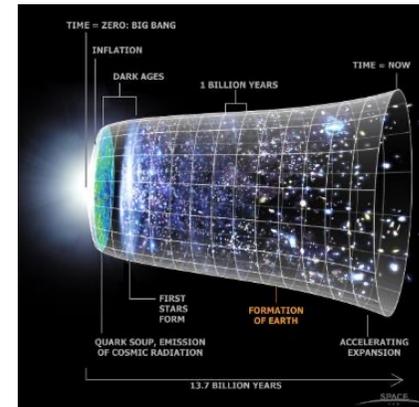
[Christophe.calvin@cea.fr](mailto:Christophe.calvin@cea.fr)

[Guigone.camus@cea.fr](mailto:Guigone.camus@cea.fr)

# Les dimensions de la donnée au CEA



- Dimension spatiale : De l'infiniment petit ( $10^{-18}$  m) à l'infiniment grand ( $10^{26}$  m)
- Dimensions temporelles
  - Age: de la yoctosecondes ( $10^{-24}$ ) à l'exasecondes ( $10^{18}$ )
  - Durée de vie : De l'instantané aux millions d'années
  - Ancienneté de l'acquisition (80 ans) et durée de conservation (infinie)

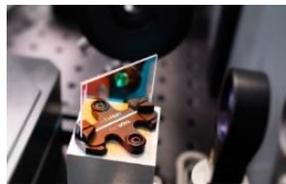
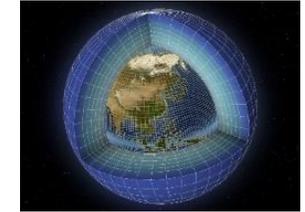
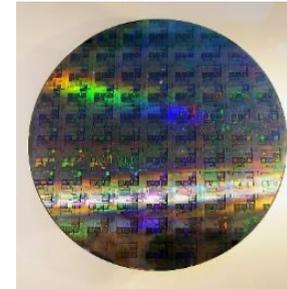


Le boson de Higgs ne vit que très peu de temps. En effet, une fois créée, lors d'une collision de particules, la célèbre particule vit moins de  $1.6 \times 10^{-22}$  seconde

# Les dimensions de la donnée au CEA



- Dimension de matière : de l'inanimé (minéral, eau ...) à l'animé (vivant humain, végétal ...)
- Dimension des outils d'acquisition et de traitement
  - Du LHC au nanopore
  - Du edge au supercalculateur
- De la donnée réelle (générée par l'homme) à la donnée artificielle (simulation, IA)
- Dimensions de partage : données ouvertes (climat) au données totalement fermées (données défense)



# Données nativement FAIR ?



Exemples	FAIR	A quel endroit du cycle	Sphère de partage
Verres nucléaires	F A i r	02 → 05	Restreint (Labo/DR)
Données exp. tokamak	F A I R	02 → 06	Restreint (communauté fusion) - International
Données du climat	F A I R	01 → 06	Open - International
Données IRM (CATI)	f a l r	01 → 03	Restreint (données propriétaires) - National
Données sismiques	F A i R/r	02 → 05	Du très restreint au très ouvert - International
Données phy. nucléaire	F A I R	02 → 06	Open - International



# Exemples

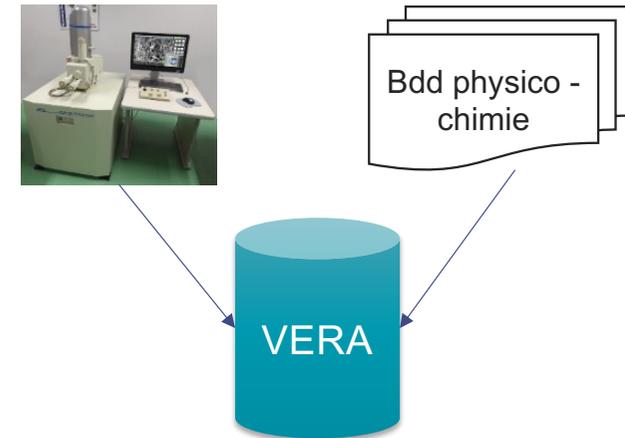
F A i r	02 → 05	Restreint (Labo/DR)
---------	---------	---------------------

## • Verres nucléaires

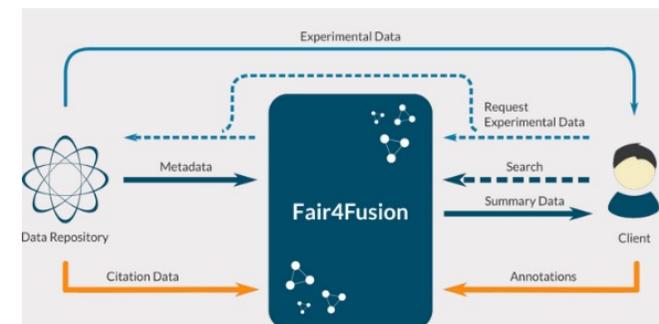
- La base de données VERA, composée à 90% de données physico-chimiques relatives aux verres (données brutes et données validées) utilisés pour le conditionnement des déchets nucléaires ultimes par vitrification

## • Données expérimentales Tokamak

- Déployer un catalogue centralisé rassemblant les métadonnées de toutes les expériences fusion européennes. Ces données sont rendues interopérables par l'utilisation du standard de données fusion « IMAS<sup>[3]</sup> » développé par ITER Organization
- Rendre possible l'accès aux données complètes des expériences via les sélections réalisées sur ce catalogue
- Rendre les données réutilisables en documentant mieux leur provenance, leur cycle de vie, en leur associant un identifiant unique et persistant, et en donnant la possibilité de les annoter et d'établir des liens avec les publications qui les utilisent
- Rendre à terme, les données accessibles à tous, dans l'esprit de la Science Ouverte.



F A I R	02 → 06	Restreint (communauté fusion) - International
---------	---------	---





# Exemples

FAIR

01 → 06

Open - International

- **Données du climat (CLIMERI)**

- Données issues des exercices internationaux d'intercomparaison de modèles climatiques
- Accord à l'échelle internationale sur quelles données (grandeurs physiques, échantillonnage spatio-temporel) et quels formats
- Liens entre modèles et résultats des modélisations



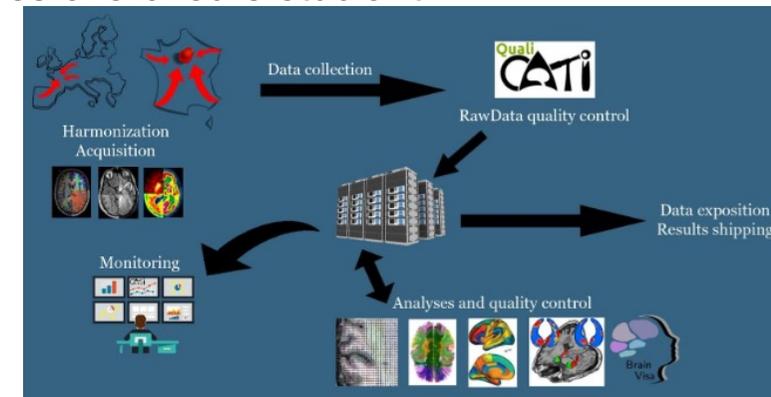
- **Données d'IRM (CATI)**

- Cette infrastructure a pour objectif de faire émerger des protocoles standardisés d'acquisition et d'analyse d'images, le tout pour obtenir un parc des données aux caractéristiques harmonisées sur les centres d'imagerie partenaires, avec un contrôle qualité systématique. À cette fin, ce centre est mis à disposition de l'ensemble de la communauté des cliniciens et des chercheurs étudiant la **maladie d'Alzheimer** et autres démences neuro-dégénératives.

f a i r

01 → 03

Restreint (données propriétaires) - National



# Exemples

F A i R/r

02 → 05

Du très restreint au très ouvert -  
International

## • Données sismiques

- Le Département Analyse, Surveillance, Environnement (DASE) du CEA/DAM est responsable de l'alerte aux essais nucléaires étrangers, aux forts séismes et aux tsunamis auprès des autorités françaises
- Diversité des données : marégraphiques, sismiques, acoustiques, satellitaires, météorologiques, internet, GPS, etc
- FDSN (Federation of Digital Seismic Networks) permet d'attribuer un code d'identification unique au flux de données sismologiques au format SEED

## • Données de physique nucléaire

- Caractéristiques des différents éléments impliqués dans une réaction nucléaire :
  - Structure et décroissance nucléaires, sections efficaces des processus de collision fondamentaux, (neutron, noyau, électrons, atomes et molécules)
  - Format international permettant de décrire les différentes données (structure, décroissance, section efficace ...) ENDF
- Ces données sont évaluées par différents pays (EU – USA – JPN) et sont disponibles au même format (donc interoperables)
- Accessible via la Nuclear Data Bank de la NEA (Nuclear Energy Agency)

F A I R

02 → 06

Open - International





# Données nativement FAIR ?

- Dans les exemples précédents, 2 seulement ont des données FAIR dès leur création :
  - Données IRM : images issues de différents IRM d'hôpitaux français et collectées au sein d'une infra de données et de services (CATI)
    - Effort de standardisation des protocoles d'acquisition des images pour maximiser l'interopérabilité des données
    - Nécessite cependant un travail important de curation des données pour les rendre totalement comparables
  - Données du climat : il s'agit de données issues des modélisations climatiques (CMIP 6 – GIEC)
    - Effort international de standardisation des données à produire avant le lancement de CMIP → toutes les données produites l'ont été au "format" décidé
- **Quelles sont les données qui peuvent être nativement FAIR ?**
  - Les données issues du réel (expérimentales, observationnelles)
  - Les données issues d'un modèle (données simulées)

