



# L'ÉCOSYSTÈME DES DONNÉES DE LA RECHERCHE À L'INSU

## PAYSAGE ET STRATÉGIE

AUDE CHAMBODUT - CMI "SCIENCE OUVERTE"  
MARYVONNE GERIN LASLIER - DAS "IR / TGIR"  
JEAN-PIERRE VILOTTE - CMI "DONNÉES ET CALCUL"

**Science Ouverte 2020, où en sommes-nous ?**  
**16 nov. 2020**

# DONNÉES DE LA RECHERCHE À L'INSU

ASTRONOMIE  
ASTROPHYSIQUE



TERRE  
SOLIDE



OCÉAN  
ATMOSPHERE



SURFACES &  
INTERFACES  
CONTINENTALES



- Les données au cœur des pratiques de recherches et des applications scientifiques de l'INSU
- Culture de partage/qualité des données (observations globales et continues)
- Organisation collective bien structurée reposant sur les Observatoires des Sciences de l'Univers (OSU)

# DONNÉES DE LA RECHERCHE À L'INSU

## ➤ Grandes hétérogénéités

- Domaines Scientifiques
- Origines (systèmes d'observations : terre, mer, air, espace ; données in silico issues de modèles et simulations, ...)
- Financements (public, industrie et privé, science citoyenne, mixte)
- Niveaux (données primaires, secondaires, produits de données, ...)
- Natures (numériques, physiques)
- Fréquences (ensembles finis, séries de données en cours d'acquisition, ...)
- Volumes (issus de grands instruments jusqu'aux données de longue traîne)

## ➤ Données « dynamiques »

- observations de systèmes non stationnaires (donnée non acquise = donnée perdue)
- par utilisation :
  - diversité des usages et transformation d'information tout au long du cycle d'utilisation des données (logistiques des données, « *bufferisation* »)
  - données produites par INSU  $\subset$  données utilisées par INSU
  - aspect nécessairement trans-institutionnel, distribué du local au national vers l'Europe et l'International



# STRATÉGIE SCIENCE OUVERTE DE L'INSU VERS L'INTERNATIONAL

- Production de données internationalisée (ESFRI, grand programmes,...)



- Intégration aux outils de logistique des données à l'international



- Participation active à la définition des grandes orientations scientifiques européennes et internationales



# LES DÉFIS DE LA PROSPECTIVE TRANSVERSE INSU (2019-2020)

## 17 défis

Origine de la Terre dans l'Univers	Signatures d'habitabilité des planètes et exoplanètes	Identification et compréhension des grandes crises -> OTELo	Interactions entre crises environnementales et écosystèmes
Modélisation intégrée du système Terre pour l'étude des risques environnementaux	Interactions entre cycles longs et cycles courts pour la mise en place des ressources	Rôle du vivant sur les cycles et sur la mise en place des ressources	Milieux polaires
Zone intertropicale	Continuum Terre-Mer	Environnements urbains	Nano-satellites
Nouveaux capteurs environnementaux	Instrumentation en milieux extrêmes	De la production de mesures brutes à la distribution de données scientifiques	Accès ouvert aux données scientifiques

Vers l'exascale:  
converge HPC-  
HPDA

## 6 sujets transverses

6 EAU PROPRE ET ASSAINISSEMENT

7 ÉNERGIE PROPRE ET D'UN COÛT ABORDABLE

11 VILLES ET COMMUNAUTÉS DURABLES

13 MESURES RELATIVES À LA LUTTE CONTRE LES CHANGEMENTS CLIMATIQUES

Intégration aux ODD et liens avec la transition énergétique et écologique

Outils du futur

Compétences du futur

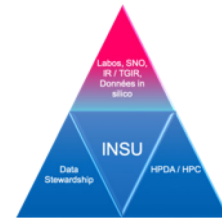
Communication, diffusion de la connaissance

Sciences participatives

Liens avec l'industrie

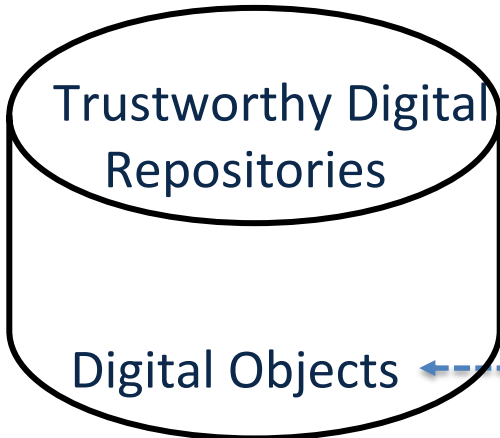


# DE LA DONNÉE EN ENTREPÔTS ...



Users / Customers  
Open Science, Open Data

↓ *Rely on*



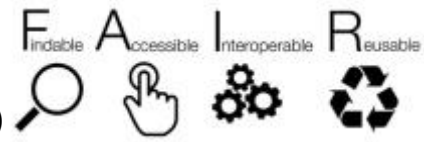
↔ Data Sharing Policies & Directives



↔ TRUST repository (Long-term)



↔ FAIR data (Snapshot)



↕ *Contribute and use*

Community / stakeholders



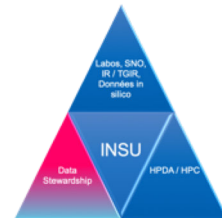
Pôles de données

TGIR/IR

OSU

ANO / SNO

# ... AUX PLATEFORMES THÉMATIQUES...



## Astronomie/ Astrophysique

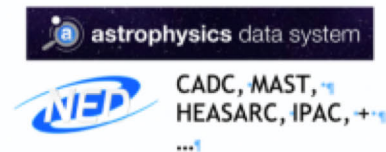
- Intégration / agrégation des SNO et projets
- Structuration par les IR/TGIR, les structures européennes (ESFRIs, GOFAIR, RDA, EOSC,...) et internationales



Ground and Space  
Observatories,  
Instruments and Missions



### Astronomy Data Centres



### Virtual Observatory

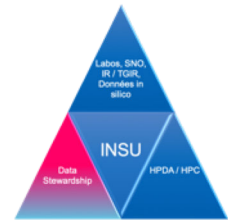


### Data e-Infrastructures





# ... AUX PLATEFORMES THÉMATIQUES...



Terre solide

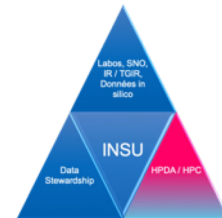
Océan-Atmosphère

Surfaces et interfaces continentales

- Intégration/agrégation des SNO et projets
- Structuration par les IR/TGIR, les structures européennes (ESFRIs, GOFAIR, RDA, EOSC,...) et internationales
- Plateforme nationale de services
  - Inter-organisme,
  - Données et produits multi-sources,
  - Services transversaux de découverte, d'accès et de gestion, des Outils,
  - Participation à la construction de l'espace européen & international.

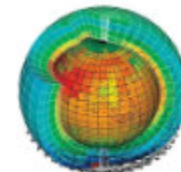
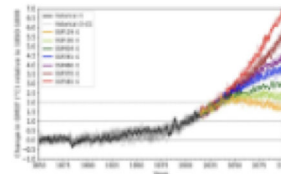


# ... JUSQU'AUX MOYENS DE CALCULS : MODÉLISATIONS, SIMULATIONS ET ANALYSES DE DONNÉES (HPC, HPDA)



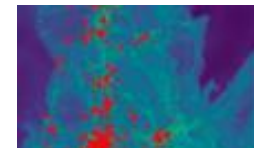
## Le Calcul intensif (HPC, Cloud) est un nœud essentiel de l'écosystème

- Données « *in silico* » produites et les outils associés (codes, *workflows*, librairies) : une partie intrinsèque des données de la recherche à l'INSU
- Besoins croissants d'assimilation de données pour l'IA « boostée » par le HPC
- Convergence HPC / Cloud – HPDA:
  - pour la gestion des données massives
  - pour permettre l'analyse, l'interprétation et la réduction intelligente des données
- Science ouverte & services : produits «à la demande», gestion & distribution d'objets digitaux



## Éléments de stratégie INSU sur le HPC

- **Accompagner et soutenir** les communautés HPC (simulation et modélisation numériques) pour l'utilisation scientifique **des architectures HPC Exascale et l'évolution des codes**
- HPC/Cloud, un nœud essentiel de l'écosystème des données → « **logistique des données** »
- Repenser le modèle des infrastructures centralisées (HPC, HPDA) afin de prendre en compte la **diversité des applications** (modélisations, simulations et analyses de données)
  - Ex: IDRIS/Jean- Zay
  - Ex: Architecture type DOA fédérant un ensemble d'infrastructures (données, HPC, Cloud)



# UNE ORGANISATION DISTRIBUÉE DE L'ÉCOSYSTÈME



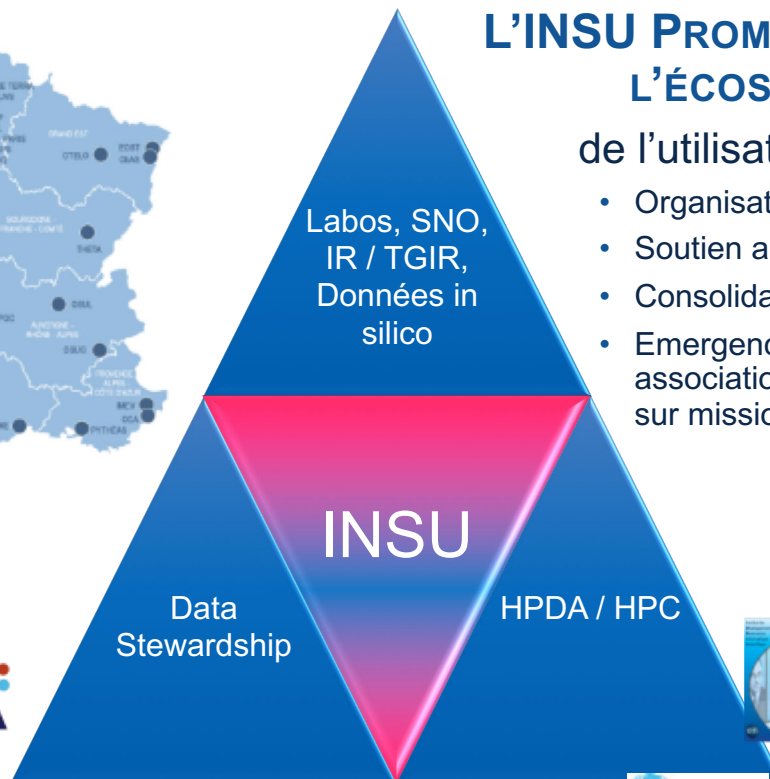
## L'INSU PROMOTEUR ET MOTEUR POUR L'ÉCOSYSTÈME NATIONAL EN SDU

de l'utilisation de la donnée

- Organisation des structures opérationnelles
- Soutien au bon niveau en moyens et compétences spécifiques
- Consolidation des pôles et plateformes données et logiciels
- Emergence d'un pôle d'expertise national : réseau métiers, association IT/CH dans une logique projet, partages de temps sur missions nationale...

du HPC en simulation et à l'interface avec les données

- Mission permanente transverse à l'INSU sur le HPC dans sa dimension calcul/exa
- Organisation de la co-construction des enjeux HPC / Cloud / IA / HPDA, entre communautés observations et modèles



UNIV, TIER-2, 3IA



**Merci pour votre attention !**



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

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