

Journées ASOV 2021

Introduction

22 & 23 mars 2021



Renouvellement de l'ASOV

Renouvellement de l'ASOV à la CSAA de juin 2020

Objectifs orientés en partie par les actions sur la Science Ouverte

- Science Ouverte
- FAIR
- EOSC (European Open Science Cloud)

Science Ouverte

- 2016: Loi pour la République Numérique
- 2018: Plan National pour la Science Ouverte
- 2019: Feuille de route du CNRS pour la Science Ouverte
- Plusieurs groupes de travail : ministère, INSU

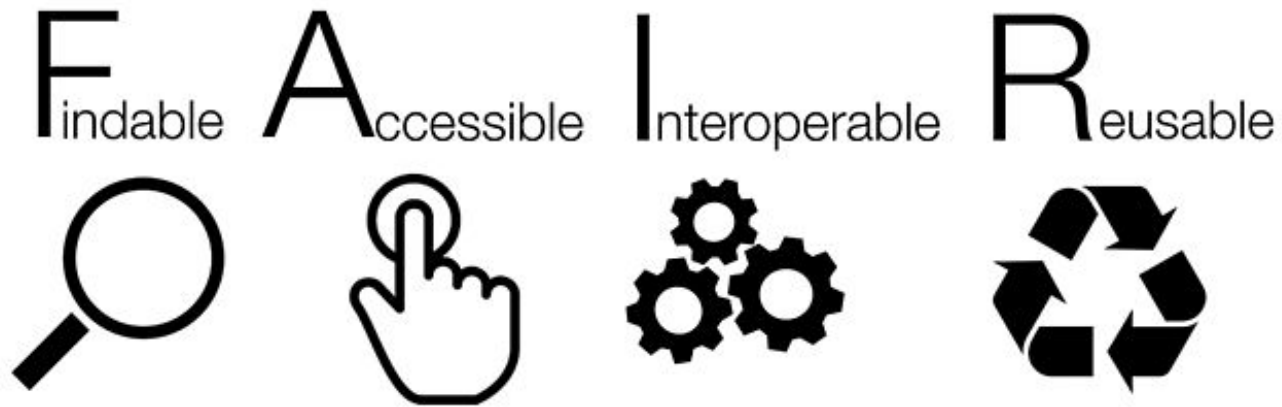
Rendre disponible les données issues de la recherche publique

- meilleur retour scientifique des données
- valider les résultats
- une recherche plus efficace



Renouvellement de l'ASOV

Données FAIR



- 2018: Commission Européenne: “Turning Fair into Reality”
- 2018 : Plan National pour la Science Ouverte (MESRI): “Notre ambition est de faire en sorte que les données produites par la recherche publique française soient progressivement structurées en conformité avec les principes FAIR.”

Impacts:

- ESFRI doivent respecter les principes FAIR
- ERC et ANR demandent des plans de gestion des données et de plus en plus à ce que les données produites soient rendues publiques



Plan d'action de la commission européenne

Renouvellement de l'ASOV

EOSC - European Open Science Cloud

Infrastructure européenne pour le partage et la découverte des données

- Partage et découverte de données
- Stockage des données
- Ressources de calcul
- Analyse de données
- etc ...

Respecte les principes FAIR

Actuellement: des programmes pilotes pour définir EOSC

Plusieurs projets européens de diffusion des données font le lien avec EOSC
Ex: ESCAPE (participation du CDS & Observatoire de Paris)



**EUROPEAN OPEN
SCIENCE CLOUD**

EUROPEAN OPEN SCIENCE CLOUD

About Services & Resources Policy Use Cases Media For providers Subscribe

The European Open Science Cloud (EOSC)

Offers 1.7 million European researchers and 70 million professionals in science and technology a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines.

[More about EOSC](#)

ACCESS EOSC SERVICES & RESOURCES

- NETWORKING
- COMPUTE
- STORAGE
- SHARING & DISCOVERY
- DATA MANAGEMENT
- PROCESSING & ANALYSIS
- SECURITY & OPERATIONS
- TRAINING & SUPPORT

Renouvellement de l'ASOV

Défi 14 de l'INSU: "Accès ouvert aux données"

[R2.2] Il faut utiliser le pilotage national pour mettre en place les conditions pour proposer des services de l'EOSC qui soient adaptés aux besoins scientifiques de la communauté.

[R4.3] La duplication des données dans plusieurs entrepôts différents est un risque

[R6.1] Toutes les données n'ont pas vocation à être FAIRisées. Lorsqu'on réfléchit aux priorités, il faut que les données produites par les Infrastructures de Recherche et celles des SNO soient FAIR, ainsi que les données liées aux publications.

[R6.3] La certification doit prendre en compte des choix stratégiques, en particulier les besoins scientifiques et la maturité des entrepôts de données et de services.

[R7.3] Trois métiers différents doivent coopérer étroitement pour que les données et les services aient le niveau de qualité nécessaire aux besoins scientifiques : les chercheurs, les spécialistes des systèmes d'informations/ documentalistes et les informaticiens.

[R9.4] Les labellisations des services nationaux d'observation doivent prendre mieux en compte le contexte FAIR, en particulier la FAIRisation des services et la participation à la définition des pratiques FAIR de la discipline.

[R11.2] L'auto-évaluation est très utile, même si l'entrepôt ne va pas jusqu'au dépôt d'une demande de labellisation.

Conclusions du Défi 14 "Accès ouvert aux données" de la prospective transverse INSU

Caroline Bot¹ ; Françoise Genova¹ ; François André² ; Mark Allen¹ ; François Bonnarel¹ ; Aude Chambodut³ ; Sylvie Galle⁴ ; Maryvonne Gérin-Laslier⁵ ; Frédéric Huynh⁶ ; Helle Pedersen⁷ ; Véronique Stoll⁸ ; Jérôme Vergne³

¹ Observatoire Astronomique de Strasbourg ; ² Observatoire Midi-Pyrénées ; ³ Ecole et Observatoire des Sciences de la Terre ; ⁴ Institut des Géosciences et de l'Environnement ; ⁵ Laboratoire d'études du rayonnement et de la matière en astrophysique ; ⁶ Institut de Recherche pour le Développement ; ⁷ Institut des Sciences de la Terre ; ⁸ Observatoire de Paris, PSL Research University

Avril 2020

1. Introduction

Les données sont au cœur de la démarche scientifique de l'INSU. Le partage des données est l'un des axes transverses forts de l'Institut, qui a depuis longtemps une politique volontariste pour le soutenir. L'INSU est certainement en pointe dans ce domaine où il s'est impliqué bien en amont du développement des principes "FAIR" (Facile à Trouver/*Findable*, Accessible, Interopérable, Réutilisable) [1] et de l'adhésion politique au concept de Science Ouverte, tel qu'il s'exprime en particulier dans le Plan National pour la Science Ouverte publié par le Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation [2] et dans la Feuille de Route Science Ouverte du CNRS [3].

C'est donc très naturellement que l'INSU consacre aux données un des Défis de son premier exercice de prospective transverse. Le Défi 14 porte sur l'accès ouvert aux données scientifiques, et son atelier s'est tenu à Strasbourg les 20 et 21 janvier 2020. La discussion a été organisée sur la base des thèmes de réflexion proposés par l'Institut, organisés selon le découpage suivant: le contexte FAIR national, européen et international (Section 2); la gestion scientifique des données et services (Observatoire Virtuel Astronomique et Infrastructure de Recherche (IR) CDS, IR Data Terra) (Section 3); les données "hors observatoires" (données de longue traîne, données attachées aux publications, données hors SNO, données non numériques) (Section 4); l'harmonisation des modèles de données et des métadonnées (description des données, formats, etc) (Section 5); et la certification des entrepôts de données et de services (en particulier *CoreTrustSeal*) (Section 6). La Section 7 identifie l'apport du Défi 14 aux questions transverses identifiées par l'INSU: lien avec les objectifs du développement durable et la transition énergétique et écologique, outils du futur (nouveaux métiers, ...), communication et diffusion de la connaissance, sciences participatives, liens avec l'industrie. Les grands thèmes identifiés et les recommandations sont résumés Section 8.

2. Le Contexte FAIR national, européen et international

L'accès ouvert aux données scientifiques est un sujet "chaud" au niveau international, avec de nombreuses injonctions politiques, et au niveau européen, avec la devise "Open Science, Open Innovation, Open to the World" de la Commission Européenne et le fort soutien qu'elle apporte au développement de

Renouvellement de l'ASOV

Politique Scientifique Top - Down pour **promouvoir la Science Ouverte et les concepts FAIR**

Science Ouverte

FAIR

EOSC

Observatoires Virtuels

Communauté A&A en avance et souvent montrée comme un exemple

Cependant: La notion de données FAIR est un concept qui nécessite une implémentation

→ Il est important de s'assurer que cette implémentation soit compatible avec l'Observatoire Virtuel

RDA : Research Data Alliance. **Approche bottom-up** pilotée par la communauté qui propose les activités qui sont utiles à ses membres.

Inscriptions sur les listes RDA

- FAIRisation des données <https://listes.services.cnrs.fr/www/subscribe/rda-france-fair>
- Certification <https://listes.services.cnrs.fr/www/subscribe/rda-france-certification>

Renouvellement de l'ASOV

Standards OV dans les prochaines années

Besoin de faire évoluer les standards: couvrir d'avantage de cas scientifiques, répondre aux besoins des nouveaux instruments

1 - Evolution des standards pour couvrir plus de cas d'utilisation

2 - Ouverture à et consolidation des domaines émergents

- hautes énergies
- radio-astronomie
- planétologie
- physique solaire
- ondes gravitationnelles

3 - Diffuser les standards existants et renforcer certaines actions

Exemple - simulations :

- Prospective INSU-AA : **AS Numérique - Simulations numériques & HPDA**
- Défi INSU sur le HPC : besoin de **diffuser des simulations**
- Mise en place de l'**ANO7** sur les **simulations numériques**

HPDA & Big Data : Pas de HPDA sans des données accessibles

- IVOA : Groupe KDD
- Certains standards OV prennent en compte les données volumineuses
- Quadriennal précédent: action MADICS MAESTRO (astro-statistique)
- 2020 : action **MADICS BigData4Astro** (CDS + laboratoires INS2I)



IVOA Provenance Data Model Version 1.0

IVOA Recommendation 2020-04-11

Working group

DM

This version

<http://www.ivoa.net/documents/ProvenanceDM/20200411>

Latest version

<http://www.ivoa.net/documents/ProvenanceDM>

Previous versions

PR-ProvenanceDM-1.0-20190719.pdf

PR-ProvenanceDM-1.0-20181015.pdf

WD-ProvenanceDM-1.0-20180530.pdf

WD-ProvenanceDM-1.0-20170921.pdf

WD-ProvenanceDM-1.0-20161121.pdf

ProvDM-0.2-20160428.pdf

ProvDM-0.1-20141008.pdf

Author(s)

Mathieu Servillat, Kristin Riebe, Catherine Boisson, François Bonnarel, Anastasia Galkin, Mireille Louys, Markus Nullmeier, Nicolas Renault-Tinacci, Michèle Sanguillon, Ole Streicher

Editor(s)

Mathieu Servillat

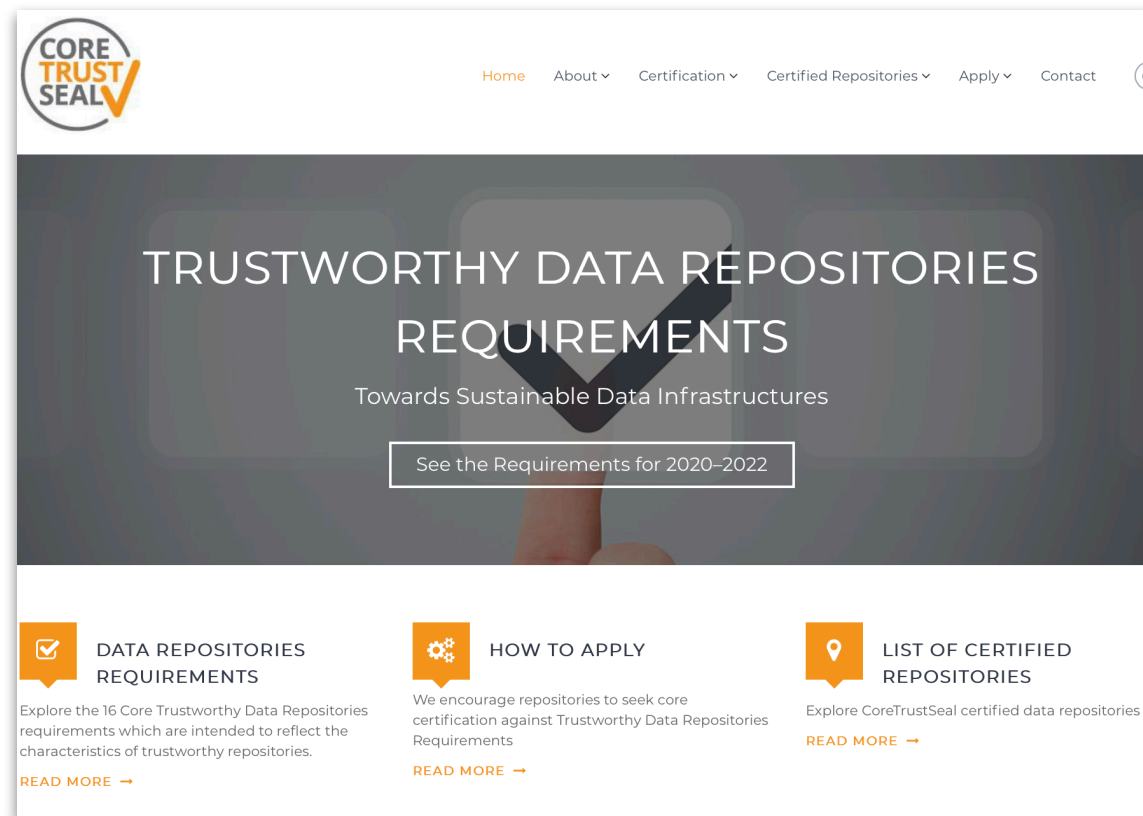
Renouvellement de l'ASOV

Certification des entrepôts de données de confiance

De nombreux services de diffusion de données dans tous les OSUs

Besoin de professionnaliser les méthodes et les procédures pour garantir la pérennité des données et des services sur le long terme

- Engager un processus de certification des entrepôts de données CoreTrustSeal
- Défi 14 de l'INSU : chercher à obtenir ce label pour des centres de données INSU



- Formations générales organisées par la RDA (Françoise)
- **2020: Réunion pour les services A&A**

RDA : Enquêtes en cours (deadline fin mars)

- **Enquête sur le paysage de la certification:**

<https://framaforms.org/entrepots-et-services-de-donnees-interesses-par-la-demarche-de-certification-1590503416>

- **Questionnaire sur les critères:**

<https://framaforms.org/criteres-de-certification-1600678445>

Renouvellement de l'ASOV

Expertise technique sur la gestion et la diffusion des données dans les OSUs

Les technologies évoluent, les standards OV évoluent

→ il est nécessaire de **maintenir un haut niveau de compétences en favorisant les échanges**

- Etre **pro-actif pour que les jeunes recrutés (scientifiques & ingénieurs) acquièrent l'expertise OV**
- Pérenniser & **soutenir les journées semi hackathon**
- Elargir la base des experts en OV : **équipes travaillant sur les grands instruments**

ASOV - Prochain quinquennal

- Poursuivre le **soutien aux équipes pour le développement des standards OV** et l'acquisition d'**expertises** sur la diffusion de données et l'**interopérabilité**
- Soutenir les **thématiques émergentes**: radioastronomie, hautes énergies, solaire, ondes gravitationnelles
- **Soutenir les jeunes** scientifiques & développeurs
- Renforcer les journées **semi hackathon pour les développeurs** et favoriser les collaborations
- Elargir les échanges d'expertise vers les ingénieurs travaillant sur les **grands instruments**
- Faire le **lien entre la communauté** et les initiatives nationales & européennes: **RDA & EOSC**
- Partage d'expertise pour l'obtention du **CoreTrustSeal** pour les CERs qui sont prêts
- Renforcer la **formation** par exemple pour les modèles & **simulations**
- Poursuivre l'**ouverture au Big Data** en lien avec MADICS

Slack : ovfrance.slack.com

2020 & 2021

Initiative du **CNES** pour mettre en place des “**pôles**” de diffusion de données en **planétologie**

- petits corps
- surfaces planétaires

Héliophysique : définition de standards
(en lien avec ESCAPE)

RDA : Réunion plénière RDA 20-23 avril 2021

Réunions InterOp

InterOps 2020 en visio-conférence

InterOps de 2021 seront aussi en visio-conférence

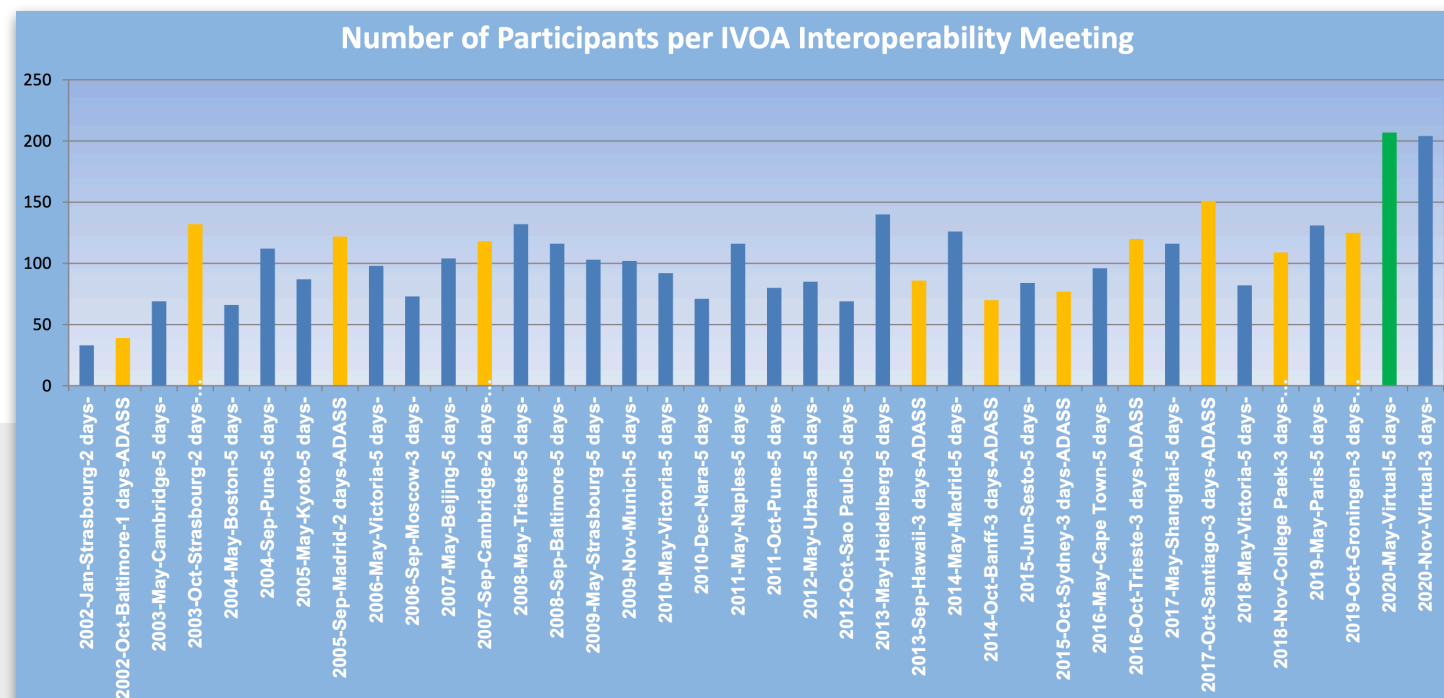
- InterOp de printemps : **24 - 28 mai 2021**
- InterOp d'automne : ~ novembre

En 2020 :

- augmentation du nombre de participants aux InterOps
- provenance est devenu une recommandation IVOA en mai 2020

Actuellement, plusieurs travaux importants en cours

- Données temporelles : évolution de certains standards
- Semantics 2.0 en RFC
- Protocoles d'accès : TAP Provenance, EPN-TAP, accès aux listes de raies, ...
- Model for ANnotating Generic Objects (Mango)
- Identification (Science platforms)
- Réunion Theory + Grid & Webservices prévue en 2021



Planning des journées

Lundi 22 mars

- 09h30-10h00, Tour des participants - Introduction - Franck Le Petit
- 10h00-10h30, **IVOA : les modèles de données** - Laurent Michel
- 10h30-11h00, **L'Observatoire Virtuel et EOSC** - Retours du projet européen **ESCAPE**, Mark Allen
- 11h00-11h15, Pause
- 11h15-11h45, **Données radio à l'IVOA** et dans ESCAPE - François Bonnarel
- 11h45-12h15, IVOA : Modèle de données **Provenance en pratique** - Mathieu Servillat
- 12h15-12h35, **Artemix** - Nicolas Moreau
- 12h35-14h00, Déjeuner
- 14h00-14h30, **Base exoplanètes** - Pierre-Yves Martin / Françoise Roques
- 14h30-15h00, TFCat – **Time-Frequency Catalogue** - Baptiste Cecconi
- 15h00-15h30, VizieR : **Modèle de données 'riche' et applications de Provenance** - Gilles Landais
- 15h30-15h45, Pause
- 15h45-16h15, **SSHADÉ** : Nouvelle base de données de 'liste de bande des solides' - Bernard Schmitt
- 16h15-16h45, OV-GSO : activités à **OV-GSO et accès aux bases de spectroscopie** - Jean-Michel Glorian
- 16h45-17h15, IDOC-SO5 : Services à **IDOC et certification** - Gilles Poulleau
- 17h15-17h30, Action **MaDICS BigData4Astro** - André Schaaff

Mardi 23 mars

- 09h30-10h00, RDA : **Certification des entrepôts de données** - Françoise Génova
- 10h00-10h30, CDPP : services du **CDPP, certification et données théoriques** - Nicolas André
- 10h30-11h00, **Données théoriques dans l'OV**, Hervé Wozniak
- 11h00-11h15, Pause
- 11h15-11h45, **POLLUX et l'OV** - Ana Palacios
- 11h45-12h15, ISMDB : **Implémentation de SimDM** - Franck Le Petit
- 12h15-12h45, RDA : **Data citation** - Carlo-Maria Zwolf