

# L'OV dans le contexte de la préservation des données



- Le contexte scientifique exige aujourd'hui de pérenniser les données
- Des thèmes se développent
  - L'identification pérenne : DOI, ORCID
  - La provenance des données : origine des données
  - La réutilisation des données
- Réponses : standards, labellisation

# □ La labellisation DSA



Le CDS (VizieR) labellisé DSA depuis 2014

DSA: Data Seal of Approval <http://datasealofapproval.org/en/>

→ certification (mais non ISO) dédiées au centre de données de recherche.



——— Seals Acquired Around the World ———



# □ La labellisation DSA



- Nouvelles directives (en cours) issues d'un groupe de travail RDA --> une nouvelle version plus spécialisée dans la qualité des données et leur réutilisation
- 16 critères définissant le SI regroupées en 3 parties
  - organisation du SI : 5 items  
mission du centre de données, licences, plan de préservations, mécanisme d'expertises scientifiques...
  - gestion des données : 8 items  
intégrité/authenticité des données, meta\_données, description du SI (OAIS), identifications, réutilisation des données...
  - technologies/sécurités : 2 items  
OS, softwares, protection du SI...

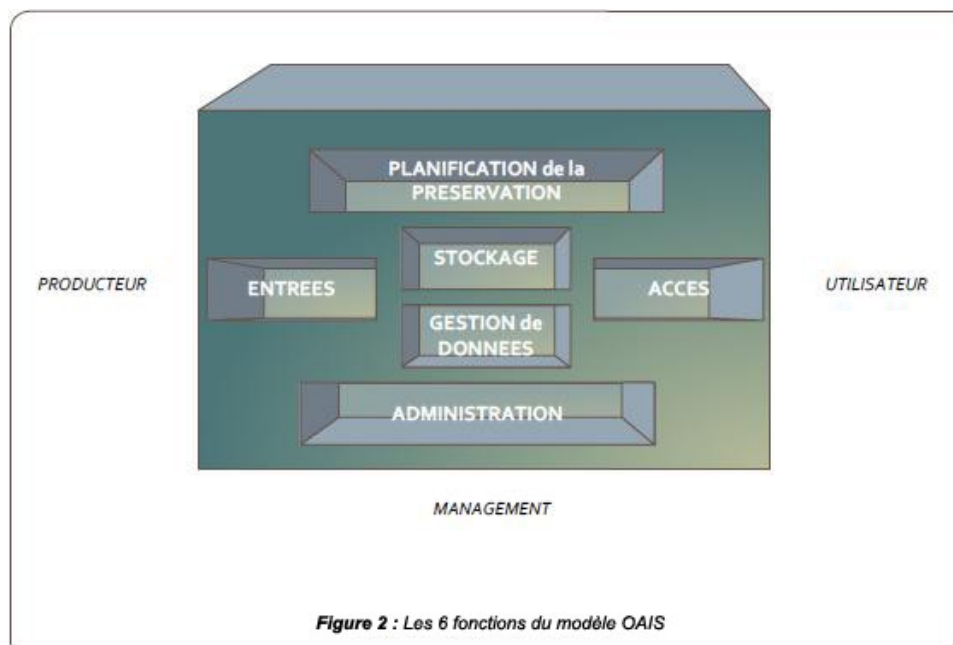


## OAIS: Open Archive Information System (norme OAIS, ISO 14721)

→ définit un cadre conceptuel et rédactionnel d'un SI. La norme inclue les descriptions des workflows, la stratégie de préservation pour le long terme et la dissémination

### OAIS est un système ouvert (open data)

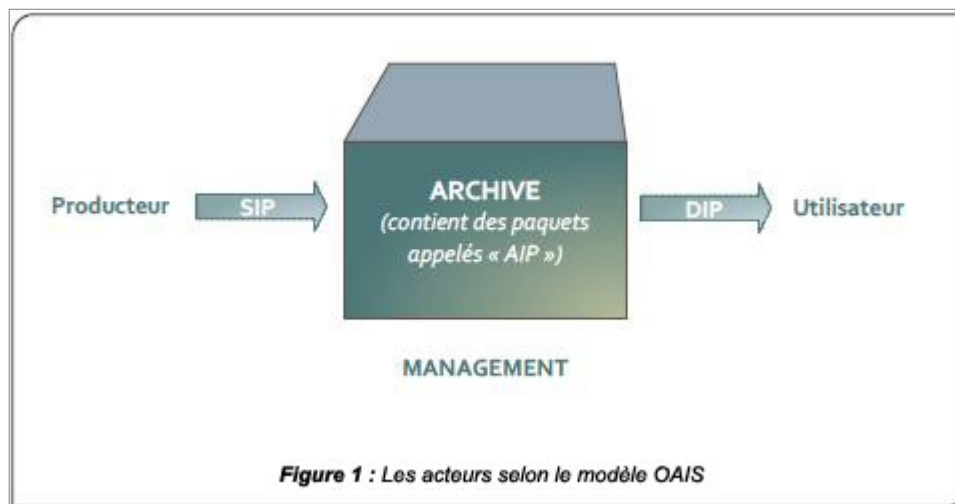
- Les modalités de workflows, le contenu sont développés dans des forums ouverts
- Les données sont accompagnées de méta-données : ouverte elles aussi



## Le rôle des métadonnées

OAIS défini 3 types de paquets :

- **SIP**: paquets recus des producteurs de données: data+meta FITS, CSV, ..
- **AIP**: paquets archives: enrichissement des meta-données, format perenne pouvant être différents du format de données des SIP
- **DIP**: informations diffusées: data+meta ex: VOTable.



## Le rôle du VO apparaît dans les 5 principes du DSA



Home Assessment Comm

Board then places the approved  
of the new DSA repository on the DSA  
ing the name of the specific repository  
if provided.

— Data Seal of Approval guidelines —

The 16 Data Seal of Approval guidelines are based  
on the following five criteria:

- The data can be found on the Internet
- The data are accessible (clear rights and licences)
- The data are in a usable format
- The data are reliable
- The data are identified in a unique and persistent way so that they can be referred to

# □ Réponse du VO au DSA



## Critère DSA 8:

« *The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.* »

→ Le VO propose des standards de représentation des données (Data models, provenance...)

ex: ObsCore pour les observations type images, spectres...



## Critère DSA 13:

« The repository enables users to discover the data and refer to them in a persistent way through proper citation. »

- Le registry VO peut être vu comme une indexation externe qui permet de découvrir les ressources du centre de données
- Les ressources du registry sont identifiées
  - identifiants VO (ivorn) ex: ivo://CDS.VizieR/II/246
  - Identifiants pérennes qui sont indexés dans les registries de l'OV
- Le registry VO est basé sur le standard OAI-PMH  
→ permet de diffuser l'information dans des "registries" plus large que l'astronomie: Eudat!

## Critère DSA 14:

*« The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data. »*

→ la définition même du VO: interopérabilité

- Le format XML VOTable est lu/interprété par des logiciels VO
- Le VO propose des protocoles standards : TAP/UWS, cone search...
- OAI\_PMH propose des services ( 'capacity' ) aux ressources du VO

# □ Réponse du VO au DSA



- Le VO correspond aux exigences DSA pour la dissémination
- Les standards du VO sont d'autant plus pertinents qu'ils s'appuient eux même sur des standards :
  - OAI\_PMH
  - VOTable sur XML
  - ADQL est basé sur SQL
- Concernant la pertinence des méta-données :
  - l'IVOA agit comme une force de proposition pour l'utilisation de datamodel
  - l'IVOA est composé d'experts issus de la discipline mais de différents laboratoires et pays ce qui renforce sa pertinence.
- Les standards du VO évoluent :



nécessite des migrations pour des ressources archivées au format VO lorsque celles-ci sont accessible directement (ex : VOTable-DML)

# LE VO dans le schéma OAIS

