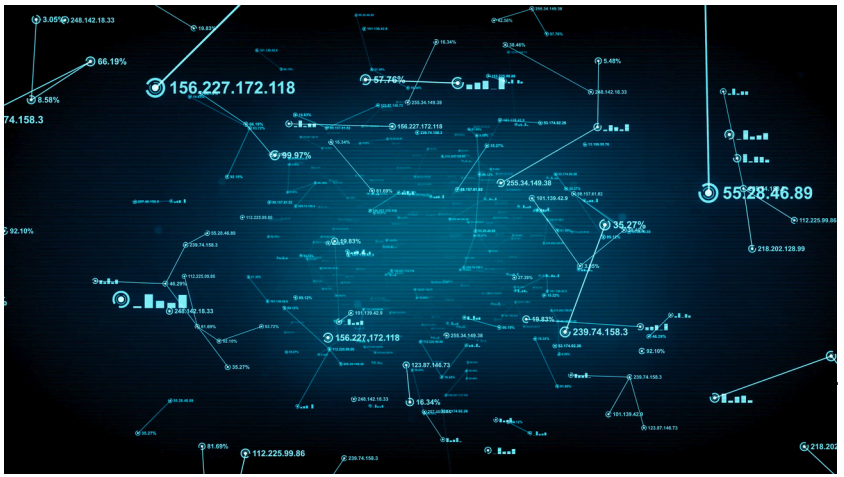
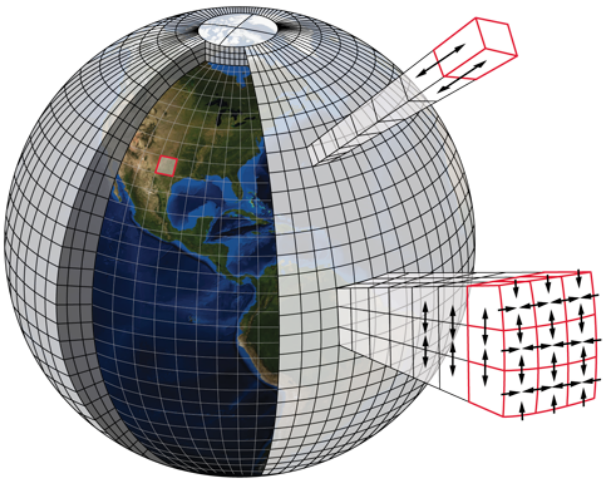
ADO2019 - Journée thématique sur les Autorités de DOnnées (8 avril 2019 / Toulouse)

- → gestion du cycle de vie des données et des méta-données
- → organisation de l'éco-système des ressources
- → émergence de la notion d'autorité sur les ontologies et les thésaurus dans un Web sémantisé
- → évolutions des usages
- → évolutions de la recherche avec la survenue de la science ouverte





pascal.dayre@irit.fr

Remerciements

- CNES
- COMET SIL, communauté d'Experts (R&T) : Dominica Ly
- Météo-France – Christophe Baehr et le CIC
- OMP et atelier permanent – Etienne Gondet
- IRIT et le CNRS
- Les intervenants

METEO FRANCE

cnes
CENTRE NATIONAL D'ÉTUDES SPATIALES

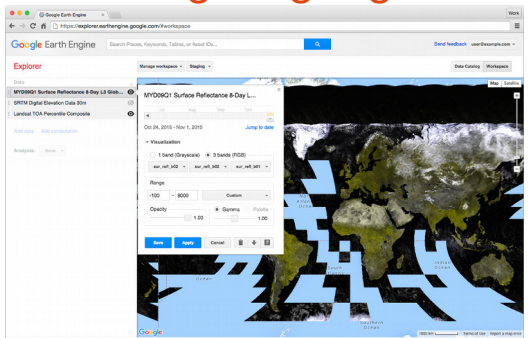
&cnrs
Depuis 80 ans, nos connaissances bâtissent de nouveaux mondes

IRIT

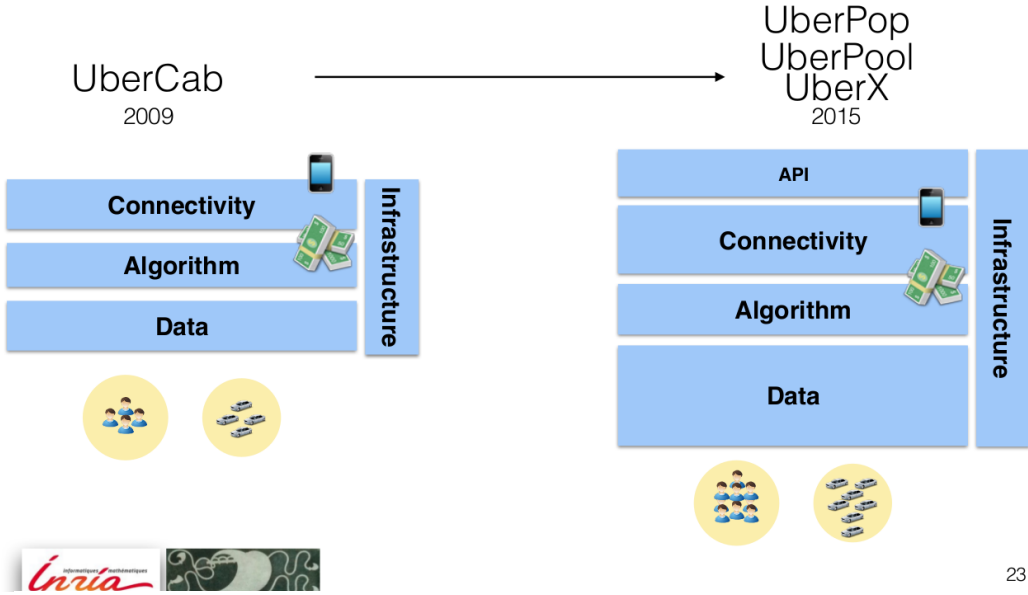
Observatoire Midi-Pyrénées
OMP

Vers une ubérisation des producteurs de données ?

- Les industriels du numérique et les plateformes d'intermédiation (Données et pouvoir : une enquête sur les plateformes d'intermédiation)
- Les données de la recherche ?
- Les données d'usage une richesse pour tous ?
- Les VRE (Les environnement de Recherche Virtualisés), la prochaine plateforme d'ubérisation des données de la recherche ?
  - Accès facilité aux catalogues de données
  - Capitalisation des codes
  - Accès simplifié aux ressources de calcul et de stockage
  - Monétisation des données d'usage ?
  - Ex : Google Earth Engine https://earthengine.google.com

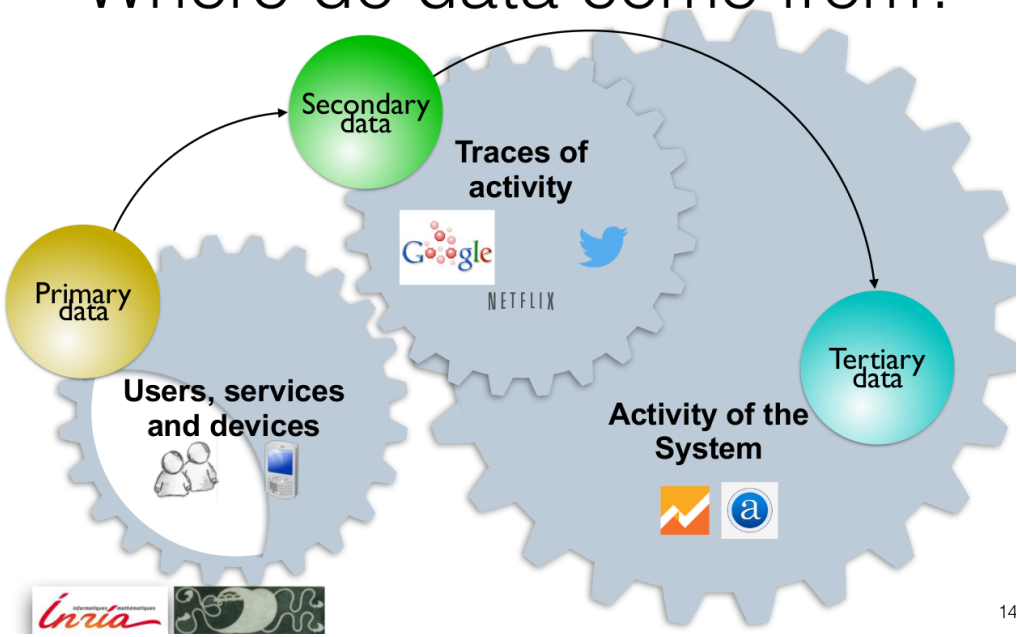


From Uber to Uberization



23

Where do data come from?



14

Aurélien Faravelon, Stéphane Frénot, Stéphane Grumbach (INRIA)
https://webcast.in2p3.fr/video/donnees_et_pouvoir_une_enquete_sur_les_plateformes_d_intermediation

Les producteurs et le cycle de vie des ressources numériques

Quels positionnements et rôles des producteurs de données dans l'éco-système des données et des services ?

Un rôle tout le long du cycle de vie de la donnée (itératif)

- Identifier les ressources
- **Modéliser et organiser un domaine**
- Faciliter l'accès (standards, protocoles, référentiels)
- Contextualiser et Aligner les données avec d'autres disciplines
- **Accompagner l'utilisation des données** pour des usages adaptés
- Aider à la curation des données en fonction de l'objectif d'usage
- **Maintenir et corriger les données avec les retours des consommateurs/utilisateurs**
- Enrichir des données au fil des usages
- **Assurer la qualité des données**
- Des prérogatives et une responsabilité juridique ?
- ...

Un éco-système de ressources et de services :

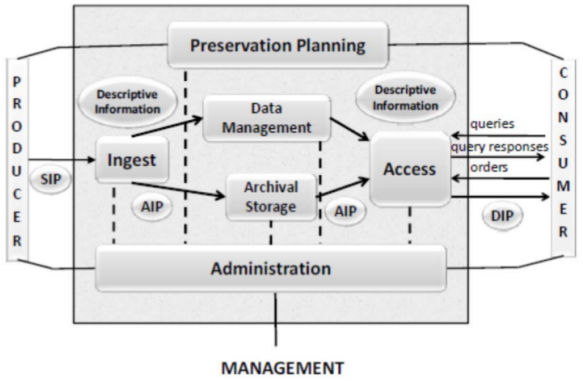
- **Le processus de la recherche, un processus itératif (science reproductible)**
- **Le web et le modèle proconsomateur**
- Le web des données un cadre global pour l'interopérabilité des ressources:
  - RDF – Ressource Description Framework
  - RDFS – Ressource Description Framework Scheme
- Des éco-systèmes toujours plus intégrés (EOSC, plateforme big data -ordonnanceurs de tâches et de données-)
- Une production de données et des ressources toujours plus répartie ? (cf. données géographiques et l'OGC)
- **Solutions centralisées versus décentralisées**

Evolution du rôle des producteurs de données vers la notion d'autorité de données

(complémentarité ? Nouvelles prérogatives ?)

- gestion du cycle de vie des données et des méta-données
- organisation de l'éco-système des ressources
- évolutions des usages

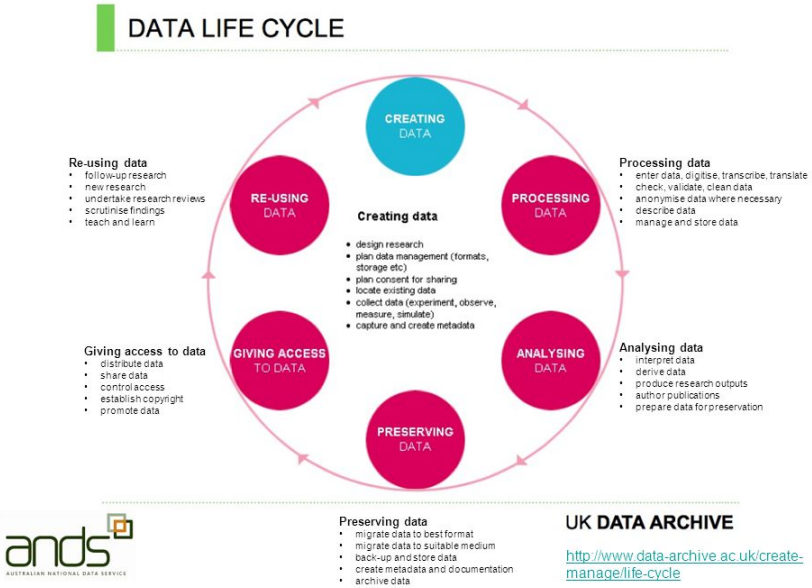
Les entités fonctionnelles des OAIS



SIP: Submission Information Package
AIP: Archival Information Package
DIP: Dissemination Information Package

Un cycle de vie itératif

Un éco-système plus complexe à organiser ?



Programme

- **09:00 - 09:05** Mot d'accueil - **Philippe Dandin**, directeur adjoint de la recherche de Météo-France
- **09:05 - 09:20** Introduction - **Pascal Dayre** CNRS/IRIT
- **09:20 - 10:00** Le rôle et les missions d'une autorité de données et ses services auprès des réseaux et des communautés. Retour d'expérience de la BNF - **Iris Berbain** - BNF / Service Référentiels au Département des métadonnées de la BnF
- **10:00 - 10:20** FAIR, une opportunité pour améliorer le cycle de vie des données - **Daniel Jacob** - INRA
- **10:20 - 10:40** Présentation de deux autorités administratives indépendantes liées aux données, la CNIL et la CADA, leurs missions et les conséquence sur l'ouverture des données publiques - **Nathalie Gandon** – DPO-INRA
- **10:40 - 11:00** Pause café
- **11:00 - 11:20** Gouvernance et cycle de vie des données au CNES - **Benoit Chausserie-Lapree** - CNES
- **11:20 - 11:35** Les données FAIR : retour d'expérience de FabSpace 2.0 - **Josiane Mothe** - IRIT
- **11:35 - 12:30** Les portails web de ressources sémantiques (ontologies, vocabulaires, thésaurus, etc), outils nécessaires des autorités de données. Présentation de solutions mises en oeuvre dans AgroPortal et le SIFR BioPortal - **Clément Jonquet** - LIRMM
- **12:30 - 13:45** Déjeuner/Buffet
- **13:45 - 13:50** L'initiative régionale DataNoos - **Nathalie Aussenac-Gilles** - IRIT / DR CNRS
- **13:50 - 14:30** L'état d'avance de EOSC, les règles de participation, la gouvernance, l'apport pour l'approche FAIR et la science ouverte et reproductible - **Philippe Segers** - GENCI / Head of European HPC projects
- **14:30 - 14:45** Quelles conditions pour que Météo-France conserve ses prérogatives sur les données météorologiques ? Quel peut être son positionnement dans l'éco-système des données et des services ? - **Philippe Dandin**, directeur adjoint de la recherche de Météo-France avec une introduction de **Christophe Baehr**, Météo-France
- **14:45 - 15:15** Les autorités de données, quelles contributions au cycle de vie de la donnée, à la construction de l'éco-système, à l'interopérabilité des données et à l'interdisciplinarité - Table ronde et réponses aux questions des intervenants
- **15:15 - 15:30** Pause café
- **15:30 - 16:30** Capitalisation et intégration sémantique de données de phénotypage - **Pascal Neveu** - INRA
- **16:30 - 17:00** Entrepôt de données IRD: un service en ligne pour l'ouverture et le partage des données scientifiques au Sud - **Jean-Christophe Desconnets** - IRD
- **17:00 - 17:15** Les Knowledge Centre : des écosystèmes de partages de données - **Daniel Galarreta** - CNES
- **17:15 - 17:30** L'accompagnement des chercheurs dans le cadre de la RGPD et des données ouvertes - **Nathalie Gandon** – DPO-INRA