

Enjeux autour du partage et de l'ouverture des données de la recherche

Françoise Genova

Observatoire Astronomique de Strasbourg

RDA-France



Research Data Sharing
without barriers



AU NIVEAU POLITIQUE INTERNATIONAL, EUROPEEN ET NATIONAL

06/09/2019

Séminaire IRD - Montpellier



□ Déjà en 2004...

- Ministres de la Recherche et de la Technologies des pays de l'OCDE + Afrique du Sud, Chine, Israël, Russie
 - Declaration on Access to Research Data from Public Funding, demandent à l'OCDE de formuler des principes et directives
- *OECD Principles and Guidelines for Access to Research Data from Public Funding, 2007*
 - Openness, flexibility, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability, sustainability
- Déclaration forte des ministres de la recherche du G8 en 2013



□ Les principes du G8 (juin 2013)

- i. To the greatest extent and with the fewest constraints possible **publicly funded scientific research data should be open**, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests, whilst acknowledging the legitimate concerns of private partners.
- ii. Open scientific research data should be easily **discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards**.
- iii. To maximise the value that can be realised from data, the mechanisms for delivering open scientific research data should be efficient and cost effective, and consistent with the potential benefits.
- iv. To ensure successful adoption by scientific communities, open scientific research data principles will need to be underpinned by an appropriate policy environment, including **recognition of researchers** fulfilling these principles, and **appropriate digital infrastructure**.



□ Ils ont persévéré et approfondi

... avec entre autres...

- Octobre 2015, Berlin
 - Further progress on sharing and managing scientific data and information should be achieved, especially by **continuing engagement with community based activities such as the Research Data Alliance RDA**.
- Septembre 2017, Turin – Open Science
 - 19. We recognize that ICT developments, the digitisation and the vast availability of data, efforts to push the science frontiers, and the need to address complex economic and societal challenges, are **transforming the way in which science is performed towards Open Science paradigms**. We agree that an **international approach** can help the speed and coherence of this transition, and that it should target in particular two aspects. First, the **incentives** for the openness of the research ecosystem: the evaluation of research careers should better recognize and reward Open Science activities. Secondly, the **infrastructures** for an optimal use of research data: all researchers should be able to deposit, access and analyse scientific data across disciplines and at the global scale, and research data should adhere to the **FAIR principles** of being findable, accessible, interoperable, and reusable.



□ Au niveau européen



- C. Moedas (Commissaire à la Recherche, Science et Innovation): *Open Science, Open Innovation, Open to the World* (2016)
- European Open Science Cloud (EOSC)



□ En France

- « La loi pour une République numérique vise à favoriser l'ouverture et la circulation des données et du savoir, à garantir un environnement numérique ouvert et respectueux de la vie privée des internautes et à faciliter l'accès des citoyens au numérique. » (JO 8 octobre 2016)
- Comité pour la Science Ouverte
 - Collèges: Publications, Données de la Recherche, Europe et international, Compétences et formation
 - Groupes projets: Evaluation, Logiciels libres et Open Source, Observatoire Pratiques informationnelles, Construire la bibliodiversité
- Plan National pour la Science Ouverte (juillet 2018)



□ Le Plan National pour la Science Ouverte



- Publié en juillet 2018
 - Généraliser l'accès ouvert aux publications
 - Structurer les données de la recherche
 - S'inscrire dans une dynamique durable, européenne et internationale



□ Les mesures pour les données (1)

MESURES

- 4 📁 Rendre obligatoire la diffusion ouverte des données de recherche issues de programmes financés par appels à projets sur fonds publics.
- 5 📁 Créer la fonction d'administrateur des données et le réseau associé au sein des établissements.
- 6 📁 Créer les conditions et promouvoir l'adoption d'une politique de données ouvertes associées aux articles publiés par les chercheurs.



Accélérer

- Proposer un appel ANR Flash destiné à accélérer l'adoption des principes FAIR et l'ouverture des données de la recherche en France.
- Créer un prix des données de la recherche récompensant les équipes et projets exemplaires dans ce domaine.

□ Les mesures pour les données (2)

Coordonner

- Construire autour de l'administrateur des données un réseau de correspondants dans les établissements, pour répondre aux questions que se posent les chercheurs sur les données de la recherche.
- Dans le cadre du soutien public aux revues, recommander l'adoption d'une politique de données ouvertes associées aux articles, le développement des articles de données et des revues de données.

Structurer

- Généraliser la mise en place de plans de gestion des données dans les appels à projets de recherche
- Développer des centres de données thématiques et disciplinaires.
- Développer un service générique d'accueil et de diffusion des données simples.
- Engager un processus de certification des infrastructures de données.

Organiser

- Soutenir la *Research data alliance* (RDA) et créer le chapitre français de l'alliance (RDA France).
- Soutenir *Software heritage*, la bibliothèque des codes sources



□ Infrastructures de Recherche ESFRI et feuille de route nationale, projets

- Infrastructures de recherche
 - Certaines sont des infrastructures de données, toutes produisent des données
 - Questions sur les 'e-needs' et la politique des données dans les questionnaires ESFRI et feuille de route nationale (depuis la mise à jour 2016)
- Data Management Plans pour les projets



□ Parmi les avancées récentes

- The FAIR Guiding Principles for scientific data management – Findable (Faciles à trouver), Accessible, Interoperable, Reusable (2016)
<http://www.nature.com/articles/sdata201618>
 - Cités par les Ministres du G7 dès 2017!
 - Déjà présents par ex. en astronomie depuis très longtemps
 - Expert Group de la Commission Européenne 2017-2018
- L'émergence rapide de la Research Data Alliance (RDA)
<https://www.rd-alliance.org>
- GO FAIR – Implementation Networks
- Et aussi DOIs, ORCID, ... pour faciliter la publication et la citation des données



□ Les principes FAIR et leur mise en oeuvre



Rapport et plan d'action de l'Expert Group FAIR de la Commission Européenne (2018)

- Concepts
- Research culture
- Technical ecosystem
- Skills
- Metrics
- Sustainable and strategic funding
- Recommandations et actions



□ La Research Data Alliance



- Fondée en mars 2013 par le gouvernement australien, la CE et la NSF (USA)
 - 8800 membres de 137 pays
 - Travail « bottom-up » sur tous les aspects du partage des données scientifiques, technologiques et sociologiques
- <http://rd-alliance.org>
- Forum de discussion neutre et international
 - Recommandations et « produits »





Vision

Les acteurs de la **recherche** et de **l'innovation** partagent librement les données, quels que soient les technologies, les disciplines et les pays, afin de répondre aux grands défis de société.

Mission

La RDA construit **des passerelles à la fois sociales et techniques** pour permettre un libre partage des données.

LA RESEARCH DATA ALLIANCE

www.rd-alliance.org

Construire des passerelles sociales et techniques permettant un libre partage des données

32 REALISATIONS PHARES

Dont 4 spécifications techniques ICT

75 CAS d'UTILISATION

Dans différents domaines, organisations et pays

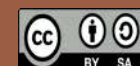
87 GROUPES QUI TRAVAILLENT SUR L'INTEROPERABILITE GLOBALE DES DONNEES

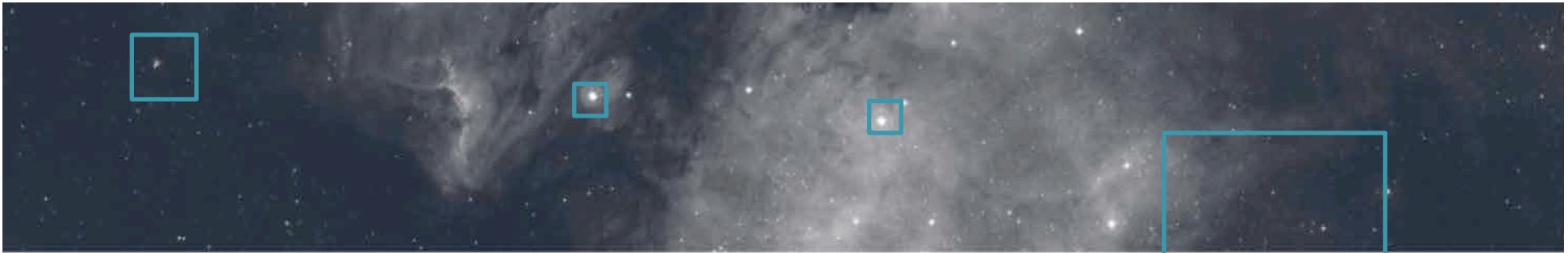
dont 32 GROUPES DE TRAVAIL & 55 GROUPES D'INTERET

8,810 MEMBRES INDIVIDUELS issus de 137 PAYS

68% recherche & académique
14,1% administration publique
12,7% entreprise & industrie

50 MEMBRES ORGANISATIONNELS & 9 MEMBRES AFFILIES





LE PARTAGE DES DONNEES AU SERVICE DE LA RECHERCHE

06/09/2019

Séminaire IRD - Montpellier



□ Pas seulement un sujet politique!

- Un changement de paradigme dans la manière dont on 'fait de la science' – méthodes d'acquisition et disponibilité des données
Découvrir les données, y accéder, les réutiliser, les combiner
- Toutes les disciplines sont concernées
- Mais elles ne sont pas toutes dans le même état d'avancement sur le sujet
- De nombreuses disciplines travaillent sur le partage de leurs données – l'astronomie a été pionnière



□ On met souvent en avant l'utilisation interdisciplinaire des données...

- ... en particulier pour les « Grands Défis »,
- ou la réutilisation en dehors du domaine de la recherche
- Mais elles requièrent des piliers disciplinaires solides
- Le partage des données nécessite un travail au niveau des disciplines – il faut connaître les données pour les rendre réutilisables (formats, métadonnées, ...)



□ Panel interdisciplinaire à l'International Data Week (Denver, sept. 2016)

- Astronomie/sciences de la planète/sciences humaines/linguistique/sciences des matériaux/cristallographie
- Points communs
 - **Pilotage par les besoins scientifiques**
 - **Définir la partie disciplinaire des standards d'interopérabilité** est difficile mais indispensable
 - Partage des données, mais aussi des outils
 - Motivation/encouragements ('incentives')
 - Les aspects sociologiques sont plus importants que les aspects techniques
 - On peut partager beaucoup des méthodes et techniques
- Le système de gouvernance mis en place est plus divers et dépend de la discipline, de sa culture et de son organisation – chaque discipline a ses propres pratiques

Genova et al. *Data Science Journal* 2017



□ Il y a de nombreux types de piliers et de ponts!



Inspiré de J. Bicaregui, exposé oct. 2012

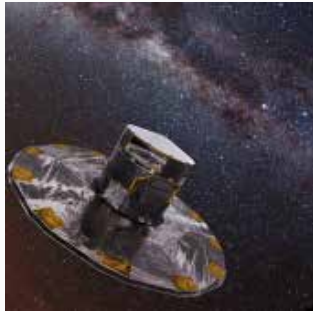




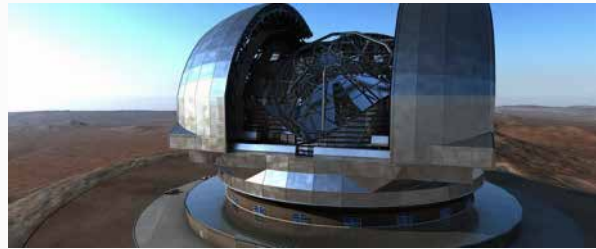
L'EXEMPLE DE L'ASTRONOMIE



□ Les infrastructures de recherche en astronomie



Gaia



ELT



SKA



Herschel



Planck

CTA



+ LES DONNEES



□ Standards...



- FITS – 1981: Données **et** métadonnées > **R**
- Observatoire Virtuel
 - Standards et outils pour trouver les données, y accéder et les interopérer (**FAI**)
 - Les standards sont définis par l'International Virtual Observatory Alliance (IVOA) depuis 2002
 - Beaucoup de travail incluant les grands projets
 - Ca marche!
- Un système ouvert
 - Les grands observatoires rendent les données disponibles après la période propriétaire
 - Tout le monde peut déclarer un service dans l'OV ou créer un outil qui utilisent les standards
 - Données et outils interopérables



One view of the VO from an application:

Built from VO Registry

Data from many observatories and missions

1000s All-Sky data sets

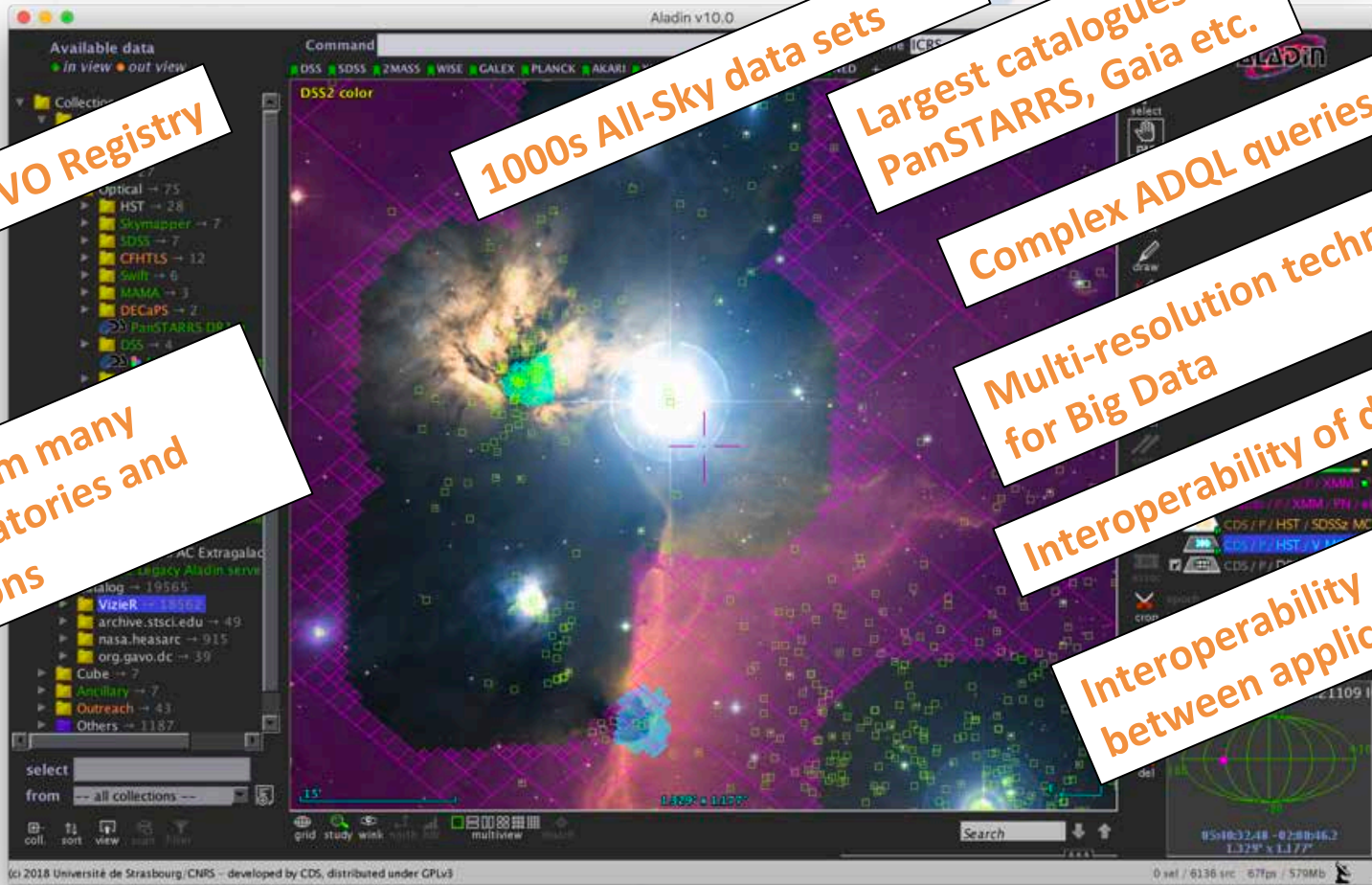
Largest catalogues: PanSTARRS, Gaia etc.

Complex ADQL queries

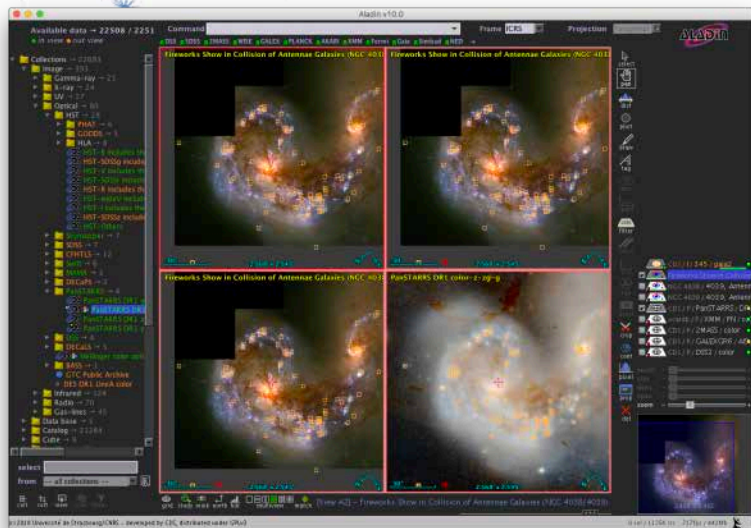
Multi-resolution techniques for Big Data

Interoperability of data

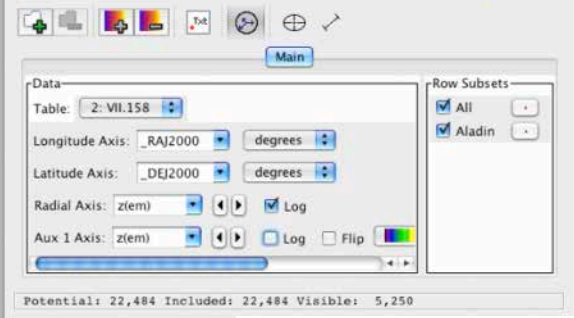
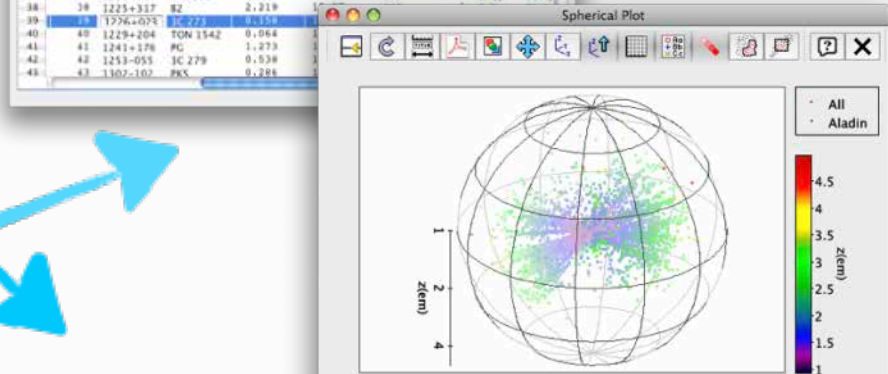
Interoperability between applications



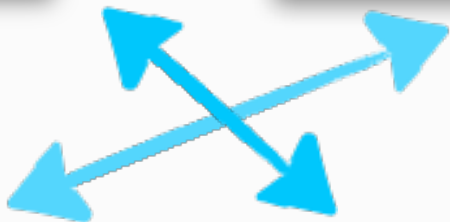
Interoperable applications and services



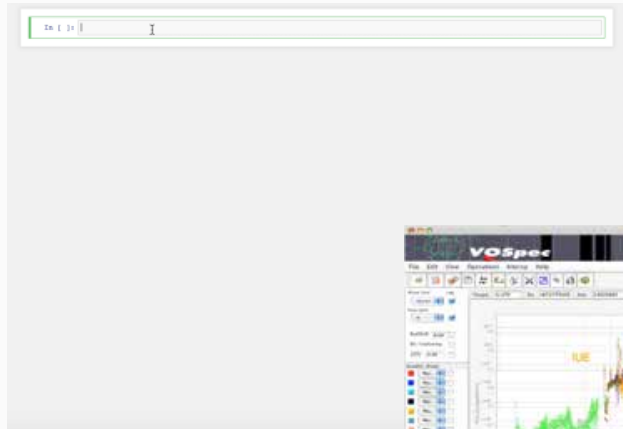
Aladin



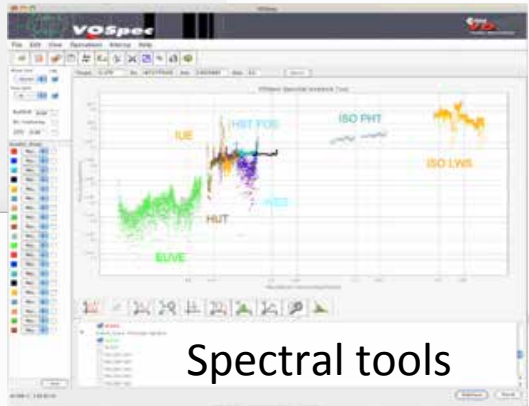
TOPCAT



Your apps & programs



Notebooks

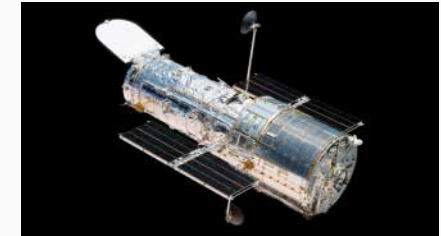
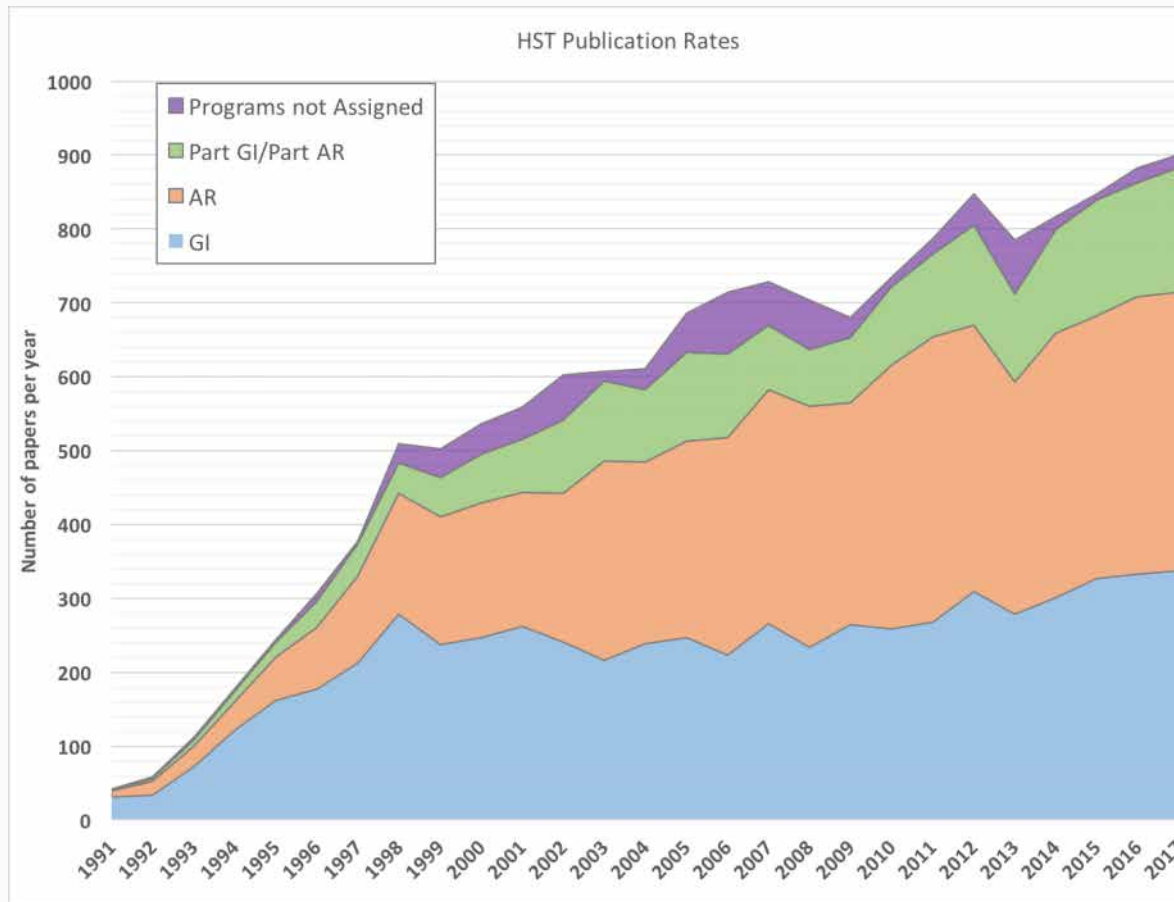


Spectral tools





L'impact du partage des données Publications avec les données du HST



Les deux

Archive

« Guest Observers »





CONCLUSIONS



□ Toutes les disciplines sont concernées

- Instruments et infrastructures « numériques », par ex. en SHS, dans les feuilles de route nationale (Huma-Num, PROGEDO) et ESFRI (DARIAH, CLARIN, CESSDA)

- IR Data Terra

- AERIS
- Form@ter
- ODATIS
- Theia



Merci à F. Huynh



□ Conclusions

- Parmi les éléments essentiels:
 - La confiance des utilisateurs (qualité des données)
 - La pertinence vis-à-vis des besoins de la recherche
- Beaucoup de travail sous-jacent
 - Chercheurs, informaticiens, curateurs du contenu (une évolution du métier de documentaliste)
 - Accord au niveau international sur les standards
 - Essentiel, mais demande du temps et des efforts



□ Conclusions

- Un sujet en pleine évolution
- Une révolution dans la manière de faire de la science
- Nous sommes tous concernés!
- Le contexte change très vite
- Des initiatives internationales en support
- La confiance (... certification) et la qualité des données sont essentielles
- Partage des données, logiciels, procédés, savoir-faire, etc

