Enjeux de la diffusion de données de simulations

Franck Le Petit - ASOV

Montpellier - 6 & 7 octobre 2022

Le contexte de la Science Ouverte

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

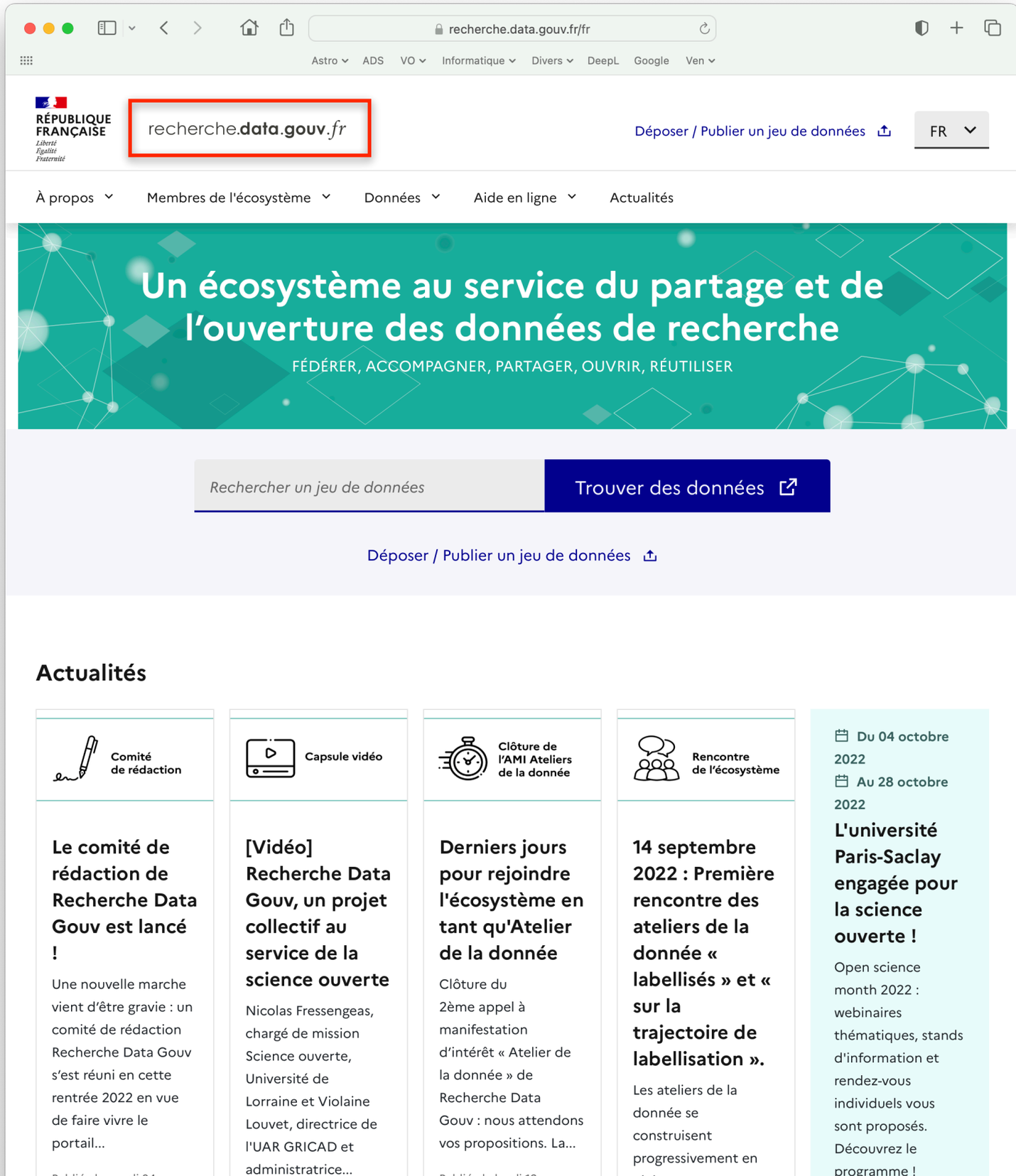
Science Ouverte

- 2016: Loi pour la République Numérique
- 2018: Plan National pour la Science Ouverte
- 2019: Feuille de route du CNRS pour la Science Ouverte

Rendre disponible les données issues de la recherche publique

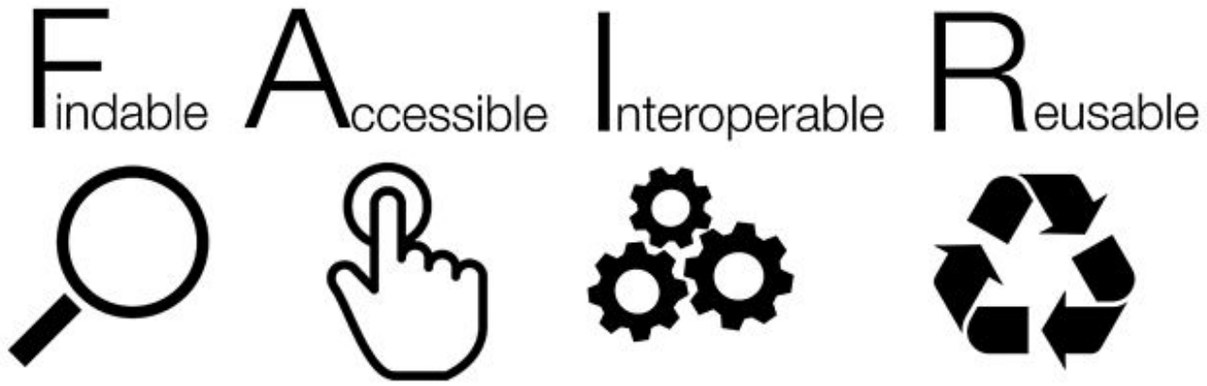
➔ meilleur retour scientifique des données
➔ valider les résultats
➔ une recherche plus efficace

Le contexte de la Science Ouverte



Le contexte de la Science Ouverte

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

Le contexte de la Science Ouverte

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)

Le contexte de la Science Ouverte

Astronomy & Astrophysics is a pioneer in data dissemination

IUE : FITS format in 1978
CDS : 1972

IUE - Launched 1978

Le contexte de la Science Ouverte

IUE - Launched 1978

Astronomy & Astrophysics is a pioneer in data dissemination

IUE : FITS format in 1978
CDS : 1972

IVOA

International Virtual Observatory Alliance

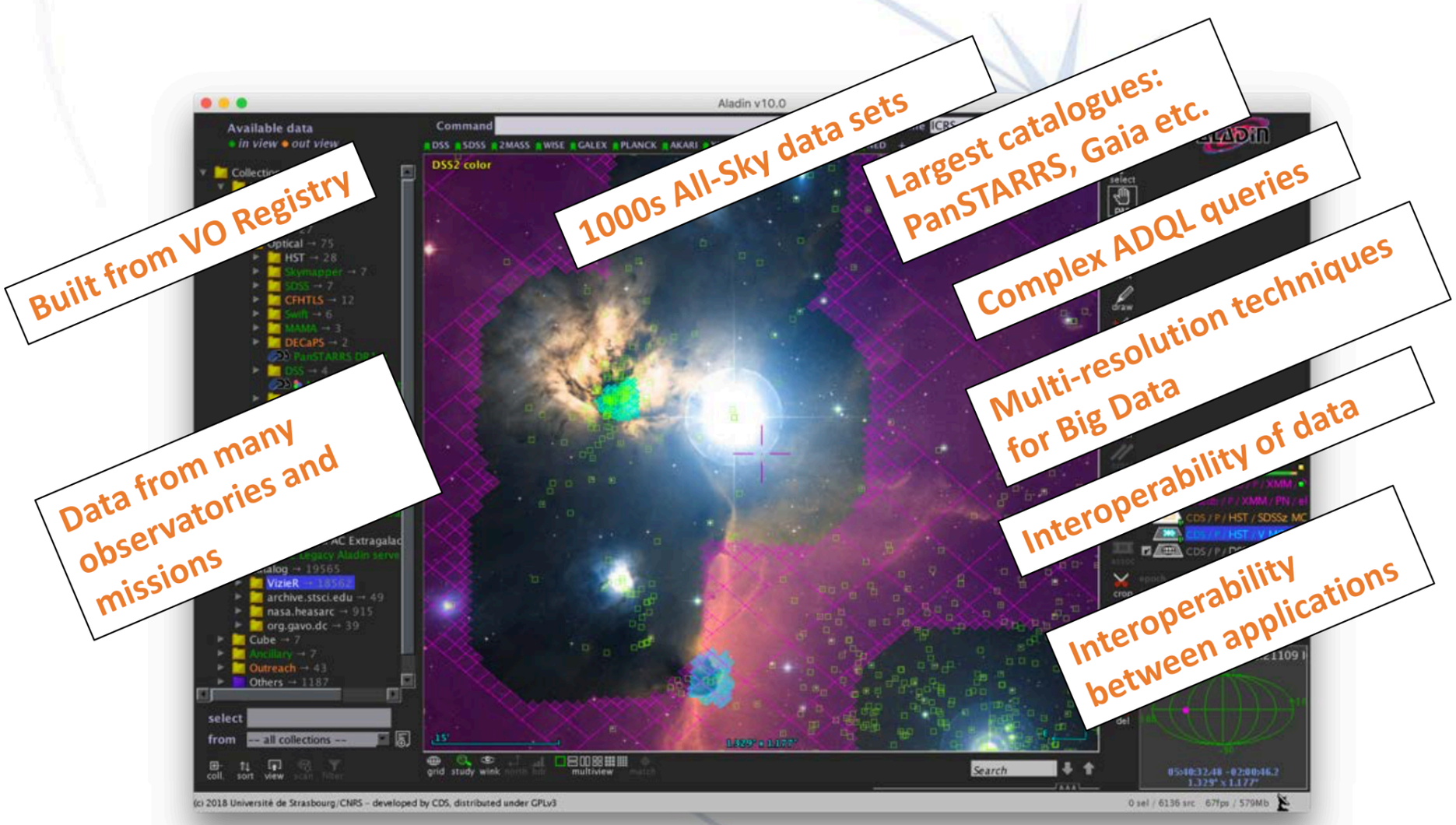
Created in **2002**
Defines standards for the diffusion of data in astronomy
Gathers the national projects
Organisation

- Executive board
- Technical coordination board
- Working groups & Interest groups

Two meetings per year (InterOps)

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

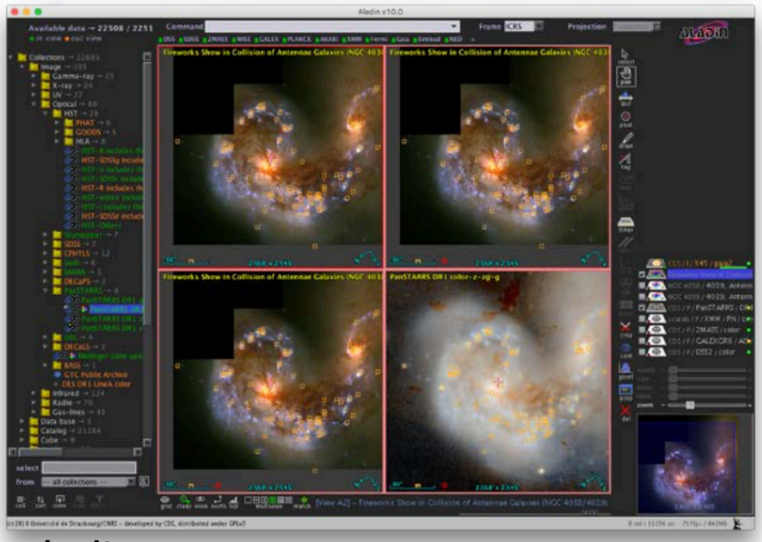
One view of the VO from an application:



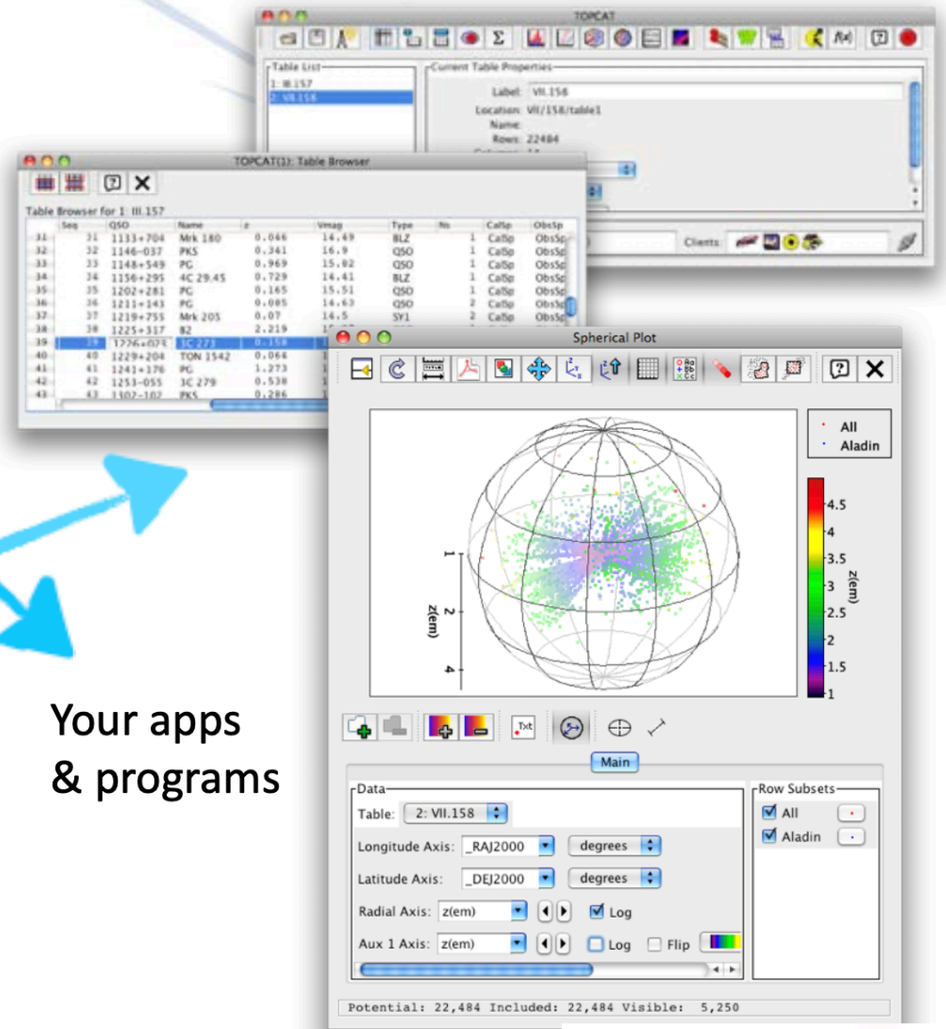
Funded by the European Union’s
Horizon 2020 - Grant N° 824064

L’Observatoire Virtuel

Interoperable applications and services

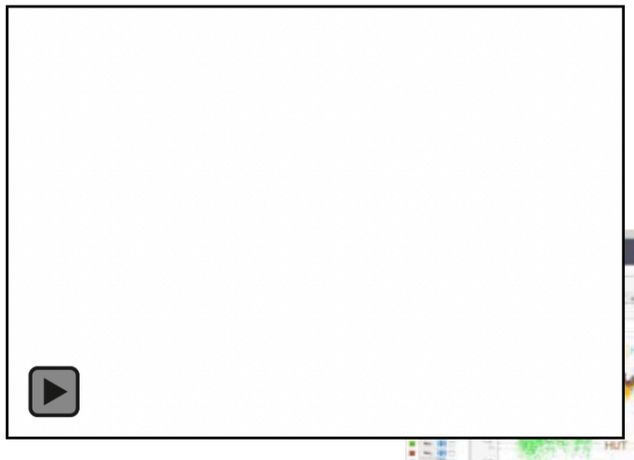


Aladin

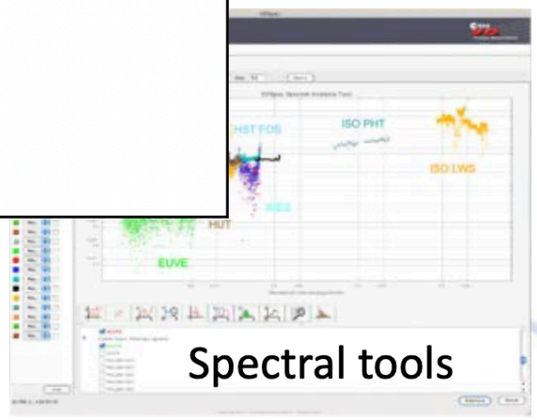


TOPCAT

Your apps
& programs



Notebooks



Spectral tools

Broadcast

Funded by the European Union’s Horizon 2020 - Grant N° 824064

IVOA - VO-Theory

Theory I.G. : 2004

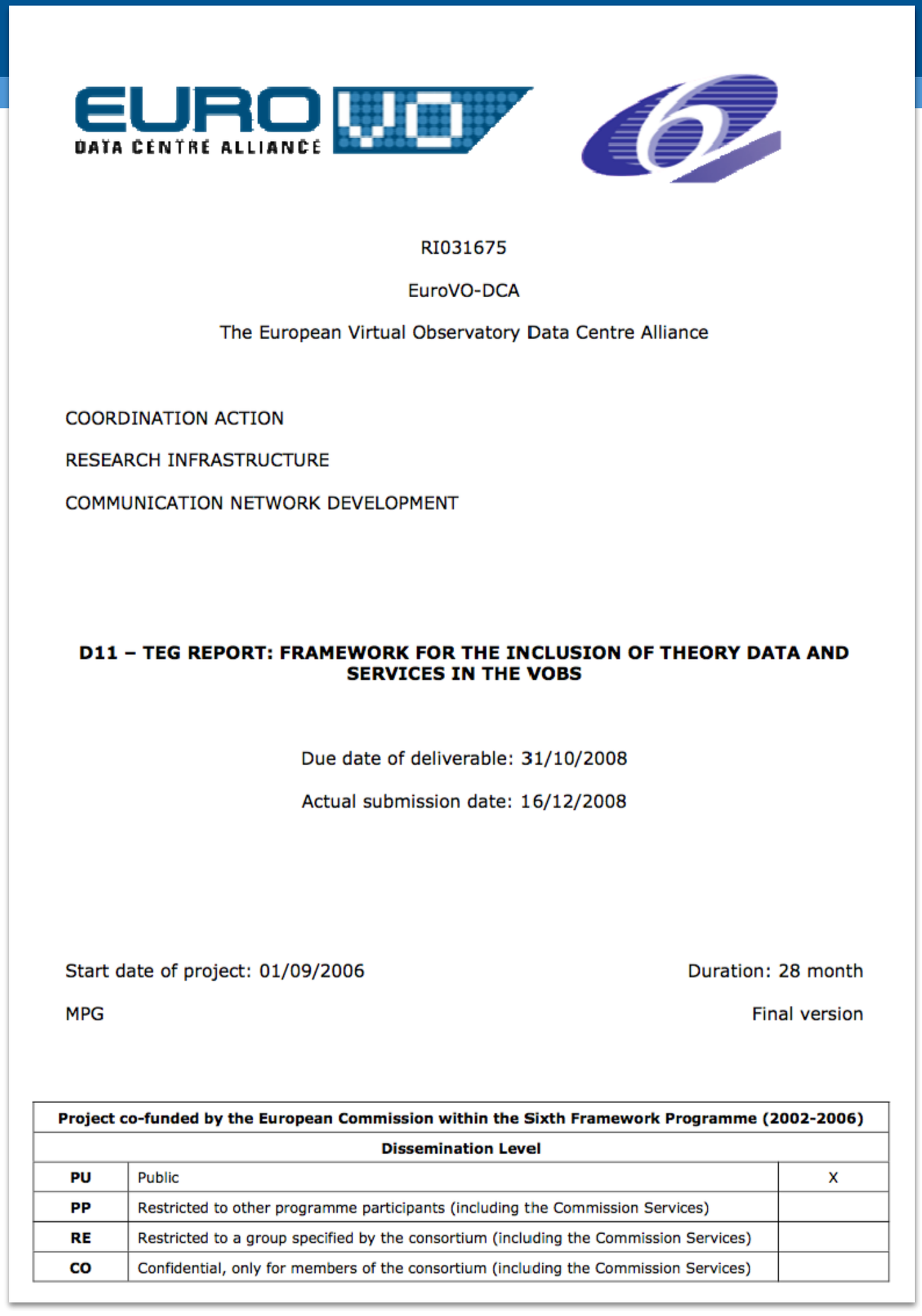
Many evolutions of the goals for Theory standards:

- particle simulations
- cosmological simulations
- all kind of simulations

Whitepaper Euro-VO DCA

Editors: G. Lemson, H. Wozniak, J. Zuther

Date: 2008



RI031675

EuroVO-DCA

The European Virtual Observatory Data Centre Alliance

COORDINATION ACTION

RESEARCH INFRASTRUCTURE

COMMUNICATION NETWORK DEVELOPMENT

D11 – TEG REPORT: FRAMEWORK FOR THE INCLUSION OF THEORY DATA AND SERVICES IN THE VOBS

Due date of deliverable: 31/10/2008

Actual submission date: 16/12/2008

Start date of project: 01/09/2006 Duration: 28 month

MPG Final version

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

IVOA - VO-Theory

Why to *publish theoretical data* ?

- Access to the data described in your publications, so readers can extend the work **➔ increase the impact of papers**
- readers can verify the results **➔ increase the quality of papers**
- allows to benchmark other works with your results / **reproductibility of results**
- incresingly **mandated by funding agencies**
- **showcases** for future proposals
- …

Why to publish theoretical data *in the Virtual Observatory* ?

- makes results available in a **standard way**
- forces you to think carefully about your results and **improves re-usability for you**
- *maybe not give obvious benefit to you but you may agree that if others do it, you would have an easier job using their data*

- facilitates **comparison models - observations**

Diffusion of simulations in the VO in France

Important role of France for the definition of IVOA standards

Euro-VO
- HORIZON-CRAL
- GALMER (projet HORIZON)
- Modèles PDR

Astronet Projet : STARFORMAT (P. Hennebelle)
- STARFORMAT (simulations RAMSES et autres) -> **Galactica**
- PDRDB -> **ISMDB**

- DEUVO (J.M. Alimi et collaborateurs)

IVOA - **Standards**

① Simulation Data Model (SimDM)

IVOA Recommendation

International Virtual Observatory Alliance

IVOA Documents

Simulation Data Model
Version 1.0

IVOA Recommendation 03 May 2012

Interest/Working Group:
http://www.ivoa.net/twiki/bin/view/IVOA/IvoaTheory

Author(s):
Gerard Lemson, Laurent Bourges, Miguel Cervino, Claudio Gheller, Norman Gray, Franck LePetit, Mireille Louys, Benjamin Ooghe, Rick Wagner, Herve Wozniak

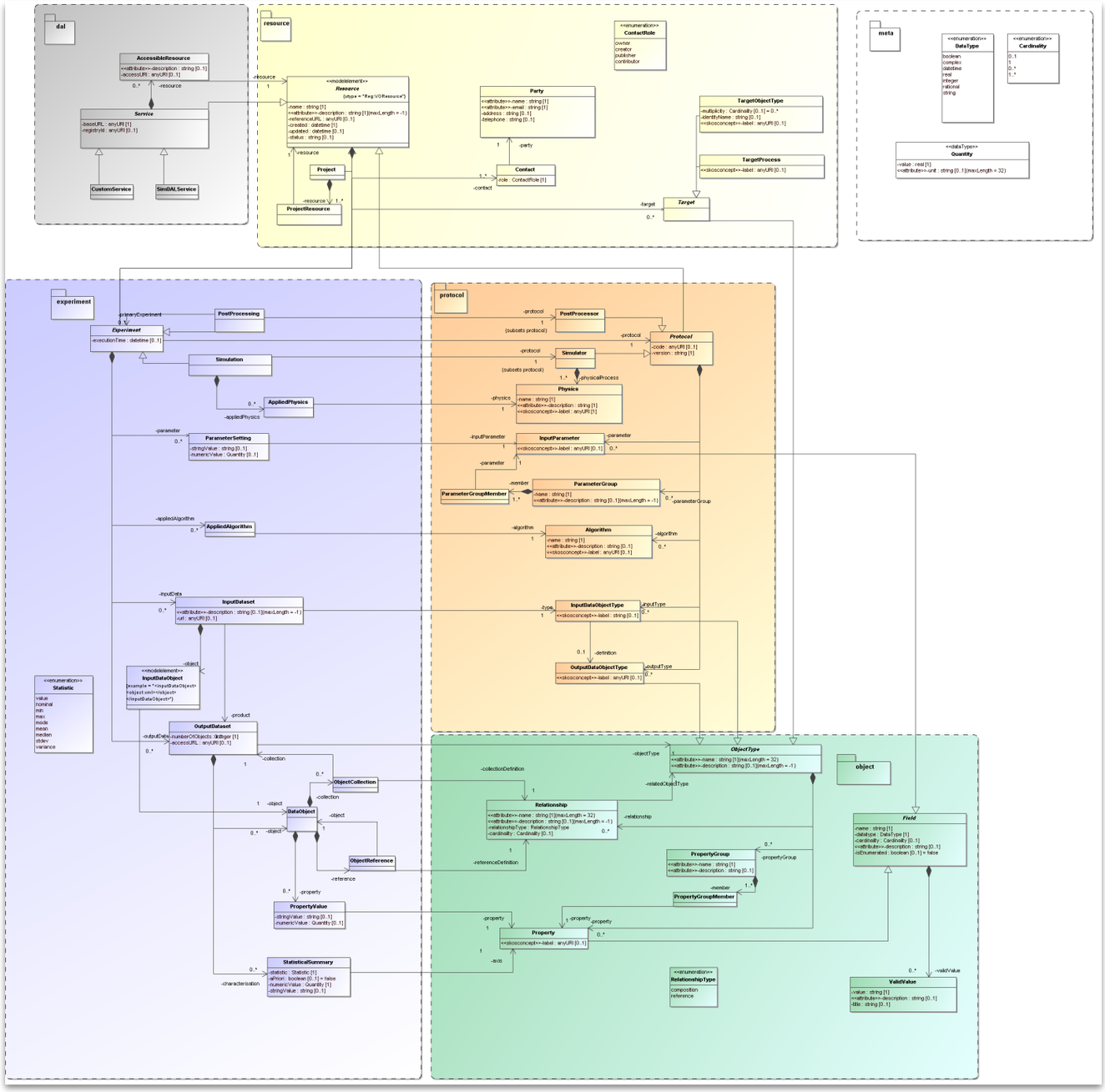
Editor(s):
Gerard Lemson, Herve Wozniak

Abstract

In this document and the accompanying documents we describe a data model (Simulation Data Model) describing numerical computer simulations of astrophysical systems. The primary goal of this standard is to support discovery of simulations by describing those aspects of them that scientists might wish to query on, i.e. it is a model for meta-data describing simulations. This document does not propose a protocol for using this model. IVOA protocols are being developed and are supposed to use the model, either in its original form or in a form derived from the model proposed here, but more suited to the particular protocol. The SimDM has been developed in the IVOA Theory Interest Group with assistance of representatives of relevant working groups, in particular DM and Semantics.

Status of this document

This document has been produced by the Theory Interest Group.
It has been reviewed by IVOA Members and other interested parties, and has been endorsed by the IVOA Executive Committee as an IVOA Recommendation. It is a stable document and may be used as reference material or cited as a normative reference from another document. IVOA's role in making the Recommendation is to draw attention to the specification and to promote its widespread deployment. This enhances the functionality and interoperability inside the Astronomical Community.



Implemented on:

- cosmological simulations
- MHD simulations
- interstellar medium micro-physics models (PDRs)
- …

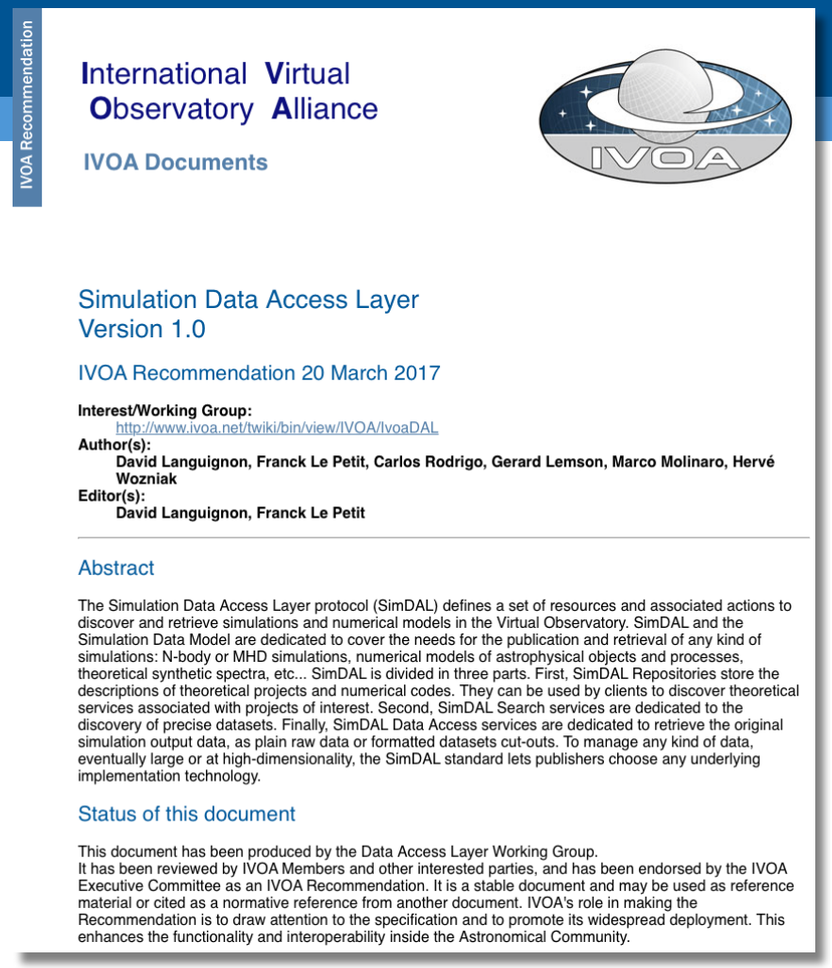
➔ proof of the versatility

IVOA - **Standards**

② Simulation Data Access protocol (SimDAL)

- SimDAL Repository
- SimDAL Search
- SimDAL Cutout

allow to discover & retrieve simulations


IVOA Recommendation

International Virtual Observatory Alliance

IVOA Documents

Simulation Data Access Layer
Version 1.0

IVOA Recommendation 20 March 2017

Interest/Working Group:
http://www.ivoa.net/twiki/bin/view/IVOA/IvoaDAL
Author(s):
David Languignon, Franck Le Petit, Carlos Rodrigo, Gerard Lemson, Marco Molinaro, Hervé Wozniak
Editor(s):
David Languignon, Franck Le Petit

Abstract

The Simulation Data Access Layer protocol (SimDAL) defines a set of resources and associated actions to discover and retrieve simulations and numerical models in the Virtual Observatory. SimDAL and the Simulation Data Model are dedicated to cover the needs for the publication and retrieval of any kind of simulations: N-body or MHD simulations, numerical models of astrophysical objects and processes, theoretical synthetic spectra, etc... SimDAL is divided in three parts. First, SimDAL Repositories store the descriptions of theoretical projects and numerical codes. They can be used by clients to discover theoretical services associated with projects of interest. Second, SimDAL Search services are dedicated to the discovery of precise datasets. Finally, SimDAL Data Access services are dedicated to retrieve the original simulation output data, as plain raw data or formatted datasets cut-outs. To manage any kind of data, eventually large or at high-dimensionality, the SimDAL standard lets publishers choose any underlying implementation technology.

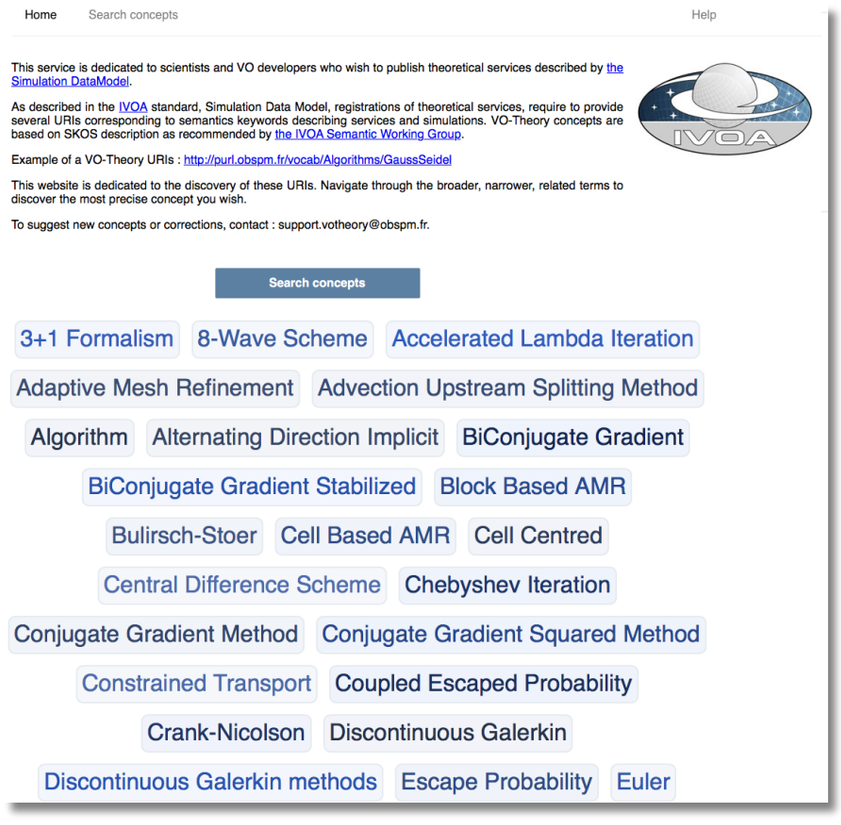
Status of this document

This document has been produced by the Data Access Layer Working Group.
It has been reviewed by IVOA Members and other interested parties, and has been endorsed by the IVOA Executive Committee as an IVOA Recommendation. It is a stable document and may be used as reference material or cited as a normative reference from another document. IVOA's role in making the Recommendation is to draw attention to the specification and to promote its widespread deployment. This enhances the functionality and interoperability inside the Astronomical Community.

③ Semantics / SKOS vocabularies

http://votheory.obspm.fr/

- Algorithms
- AstronomicalObjects
- DataObjectTypes ~ 700 concepts
- PhysicalProcesses
- PhysicalQuantities

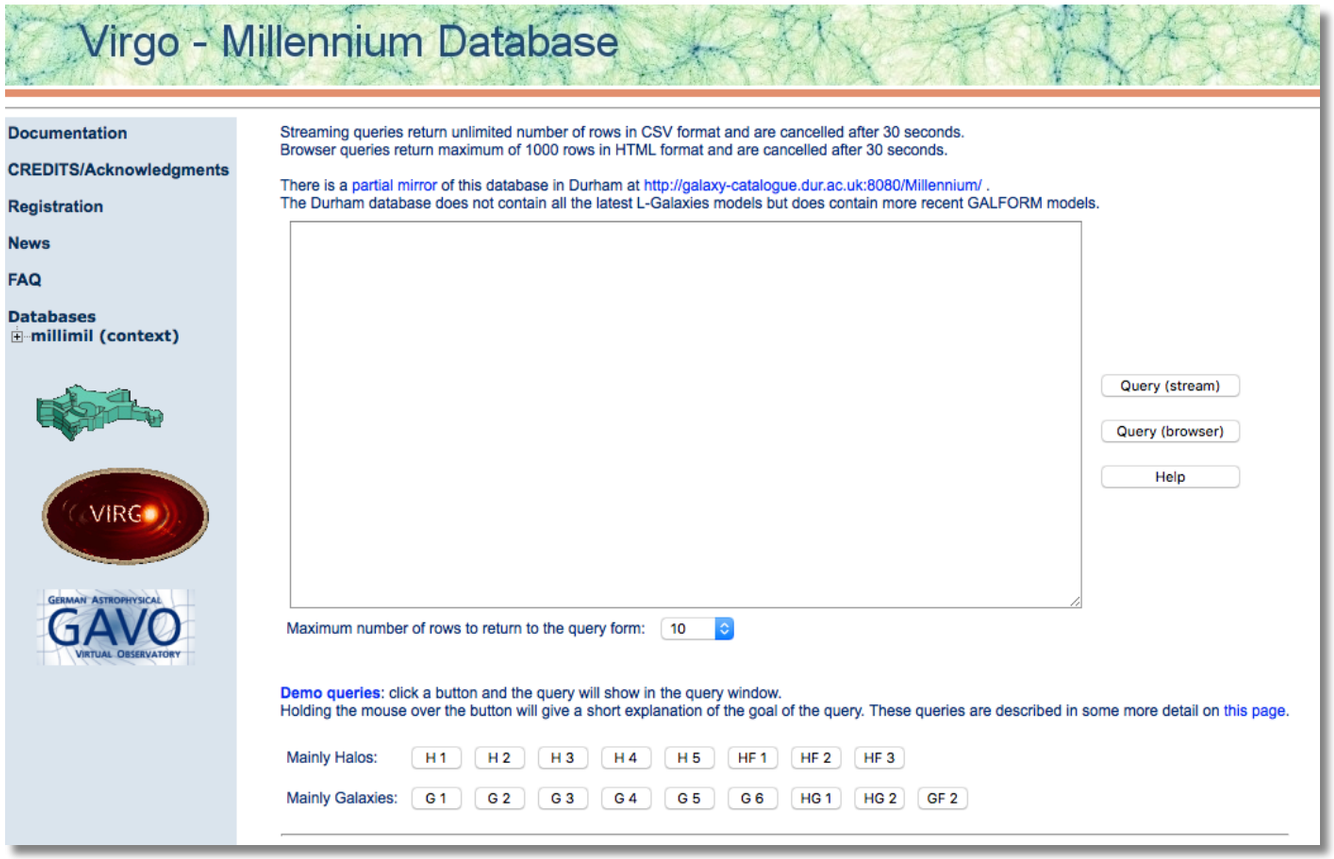


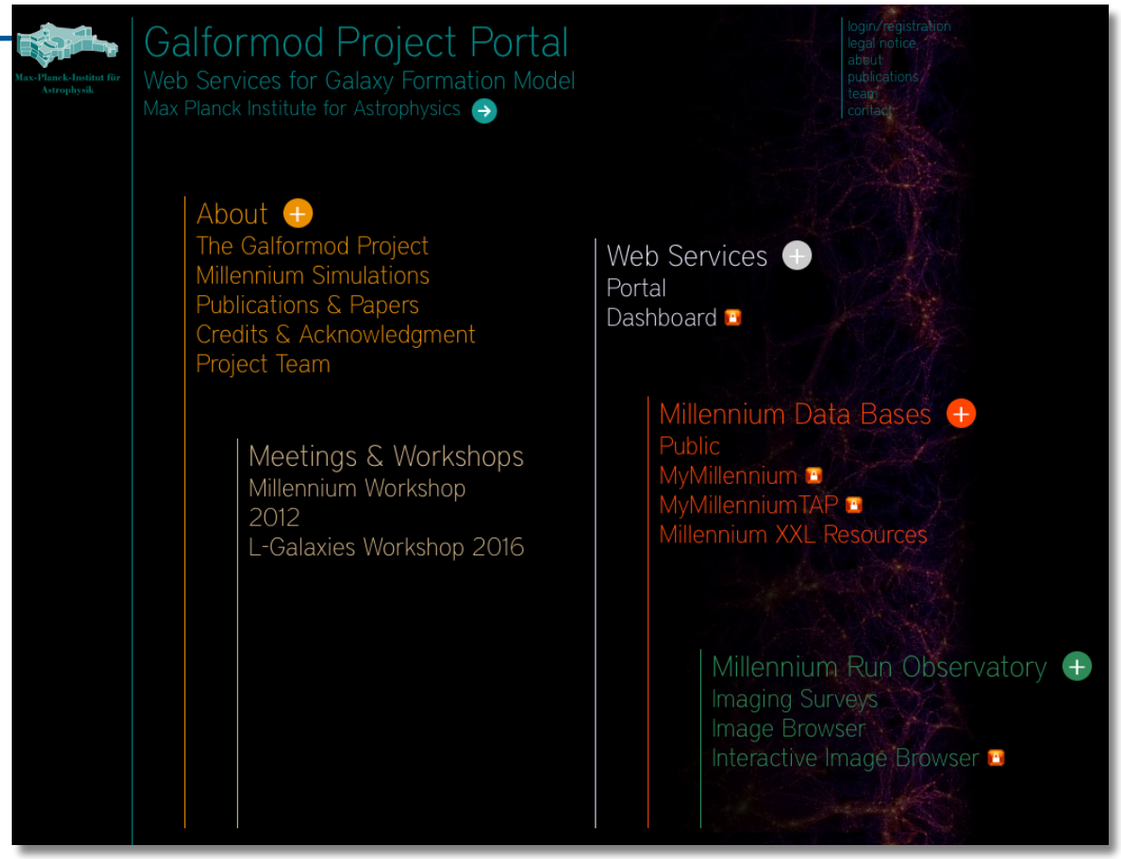
⊗ But no common raw data format

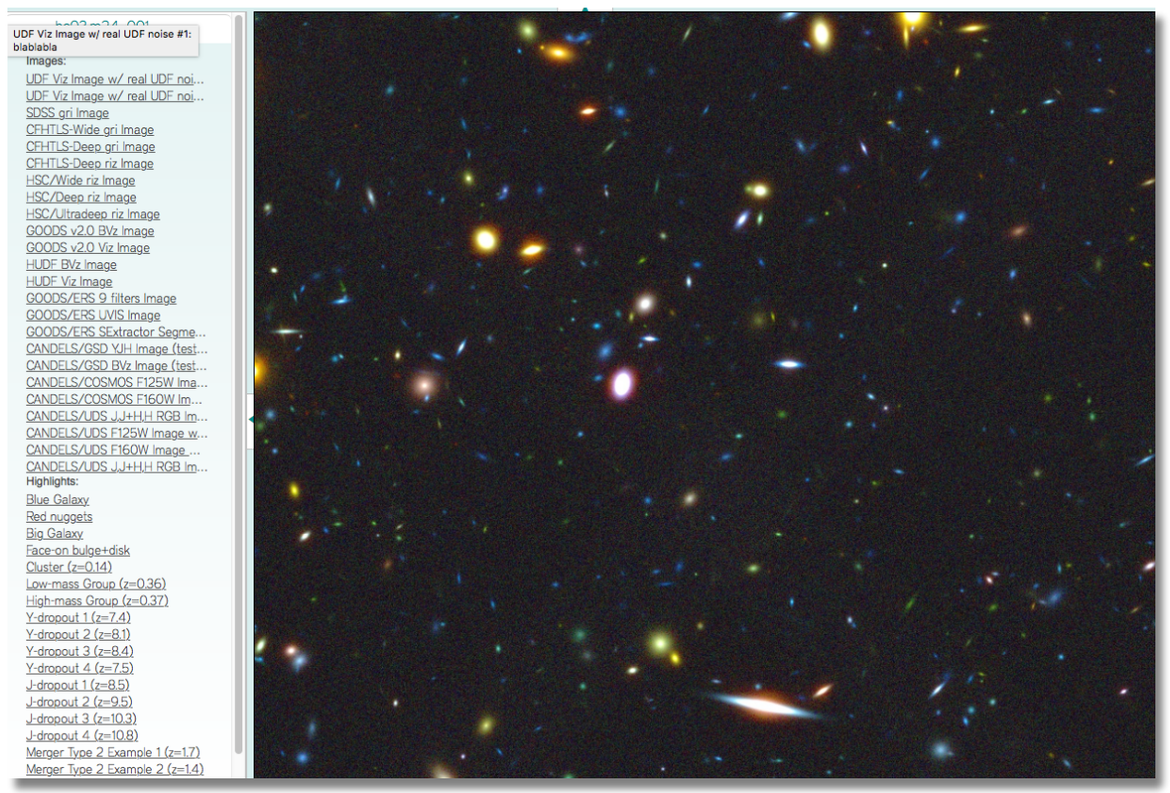
No standard raw data format because of the **heterogeneity** (no FITS for example)
➔ a limitation to use the VO
Recommendation to use VO-Table / FITS whenever possible

Example: Millenium

Millenium simulation





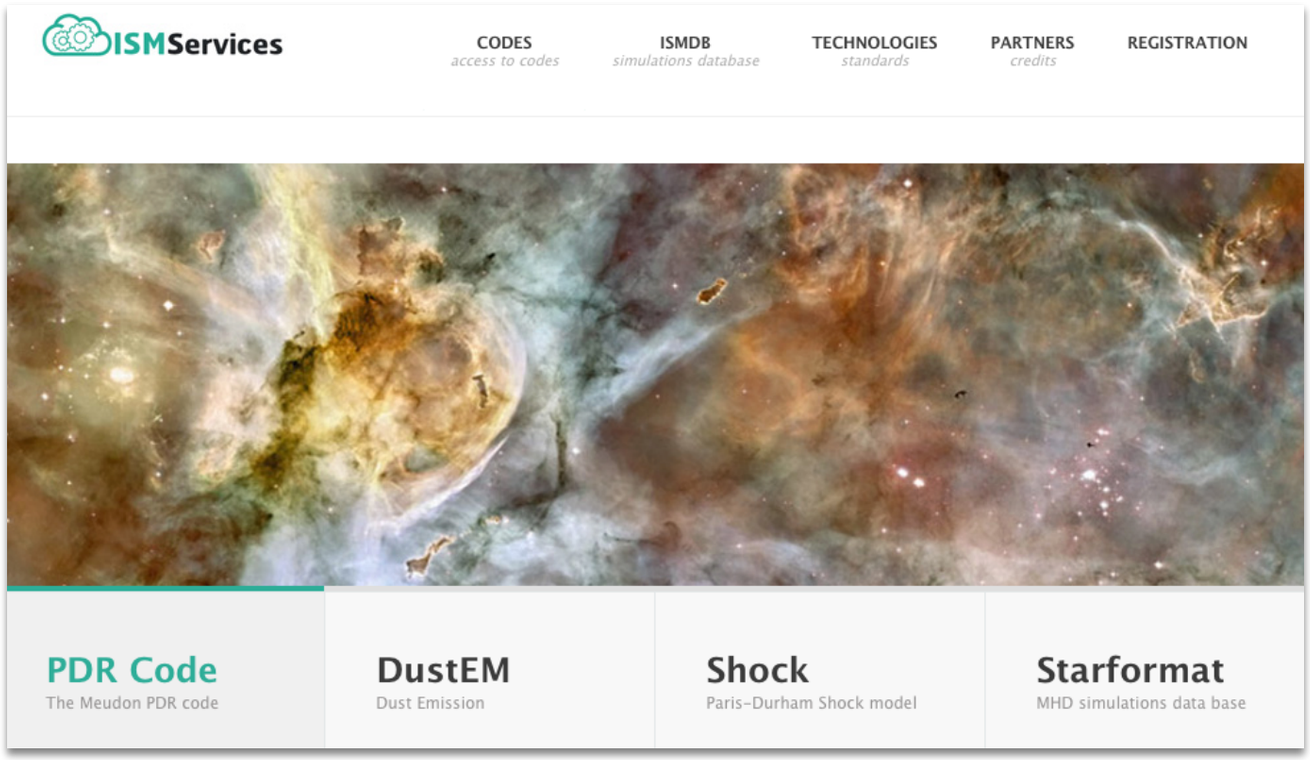


- Access to Millenium simulations
- **Many services developed above the data**
  - access to simulations
  - halo catalogs / Lightcones
  - synthetic images
  - …
- **Prototype for SimDM**
- **VO compatibility for some data (TAP / Images)**

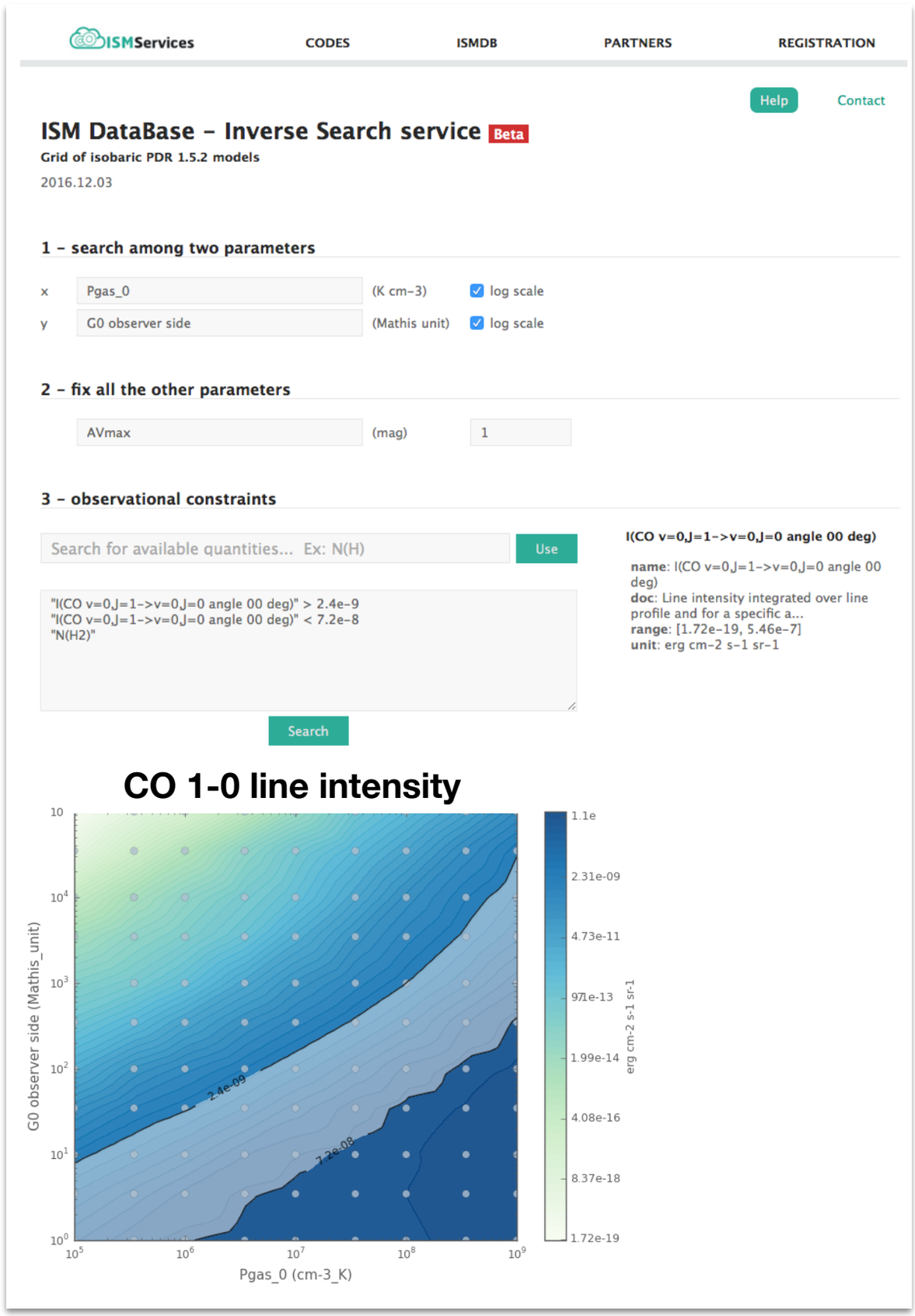
Example: Interstellar medium

ISM Services: models in interstellar clouds

http://ism.obspm.fr



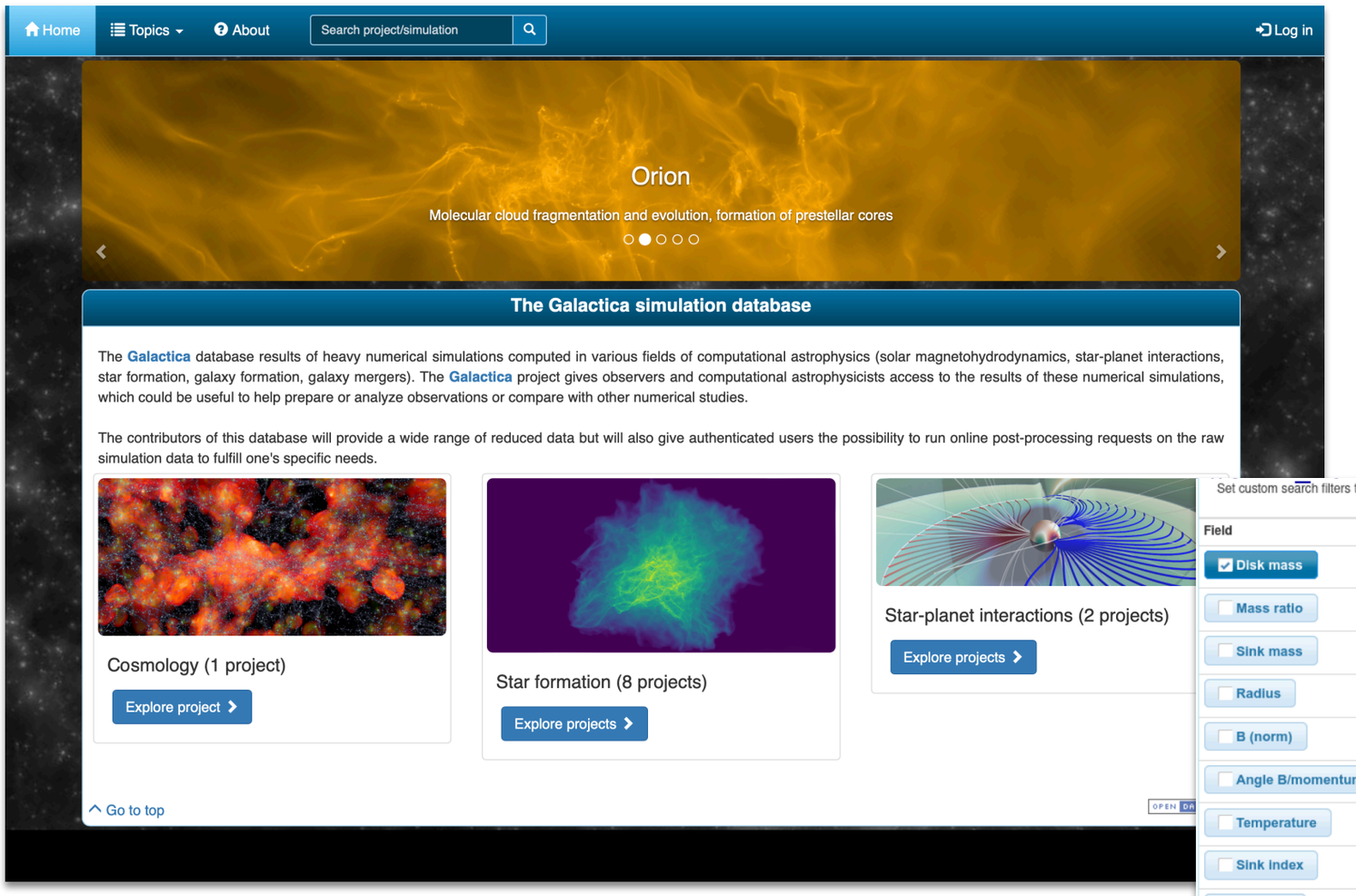
- Access to simulations of the Interstellar Medium
- Prediction of line intensities
- Fully VO-Theory compatible
  - ISMDB: **SimDM** + ***SimDAL*** + **Semantics**
  - Online codes: ***PDL***

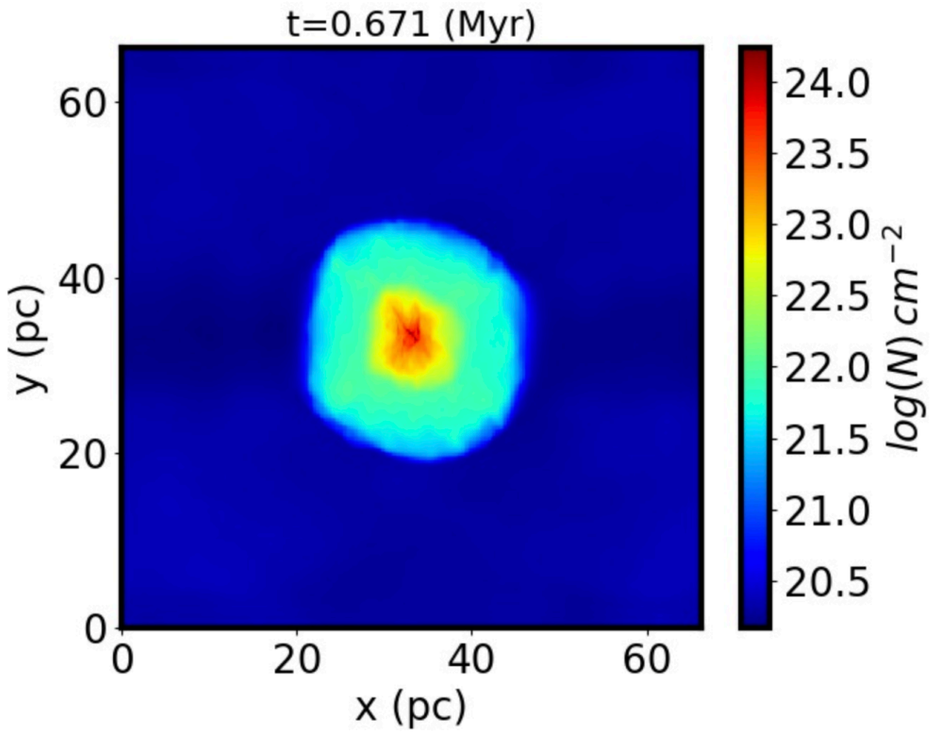


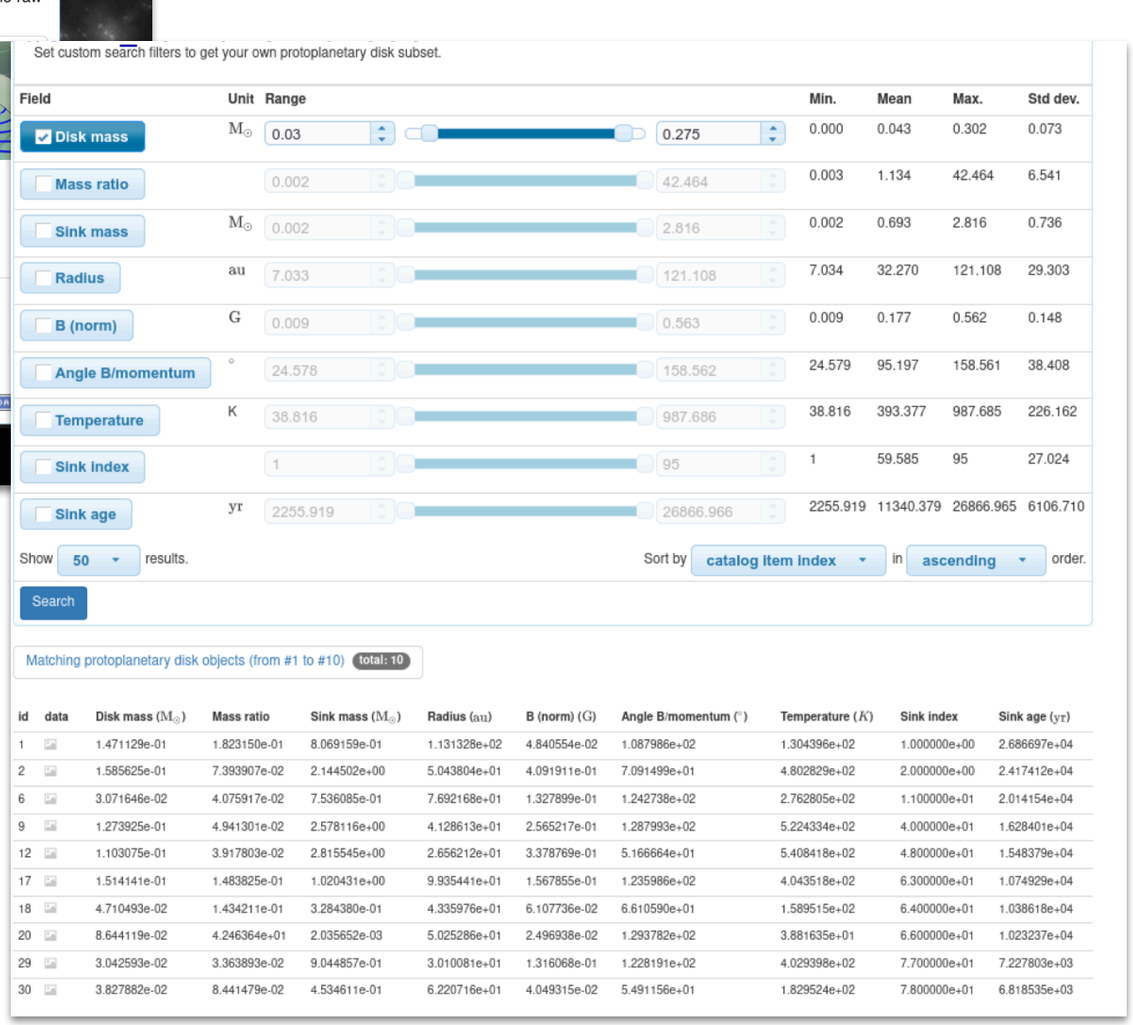
Example: Galactica

3D Simulations :

http://www.galactica-simulations.eu/db/



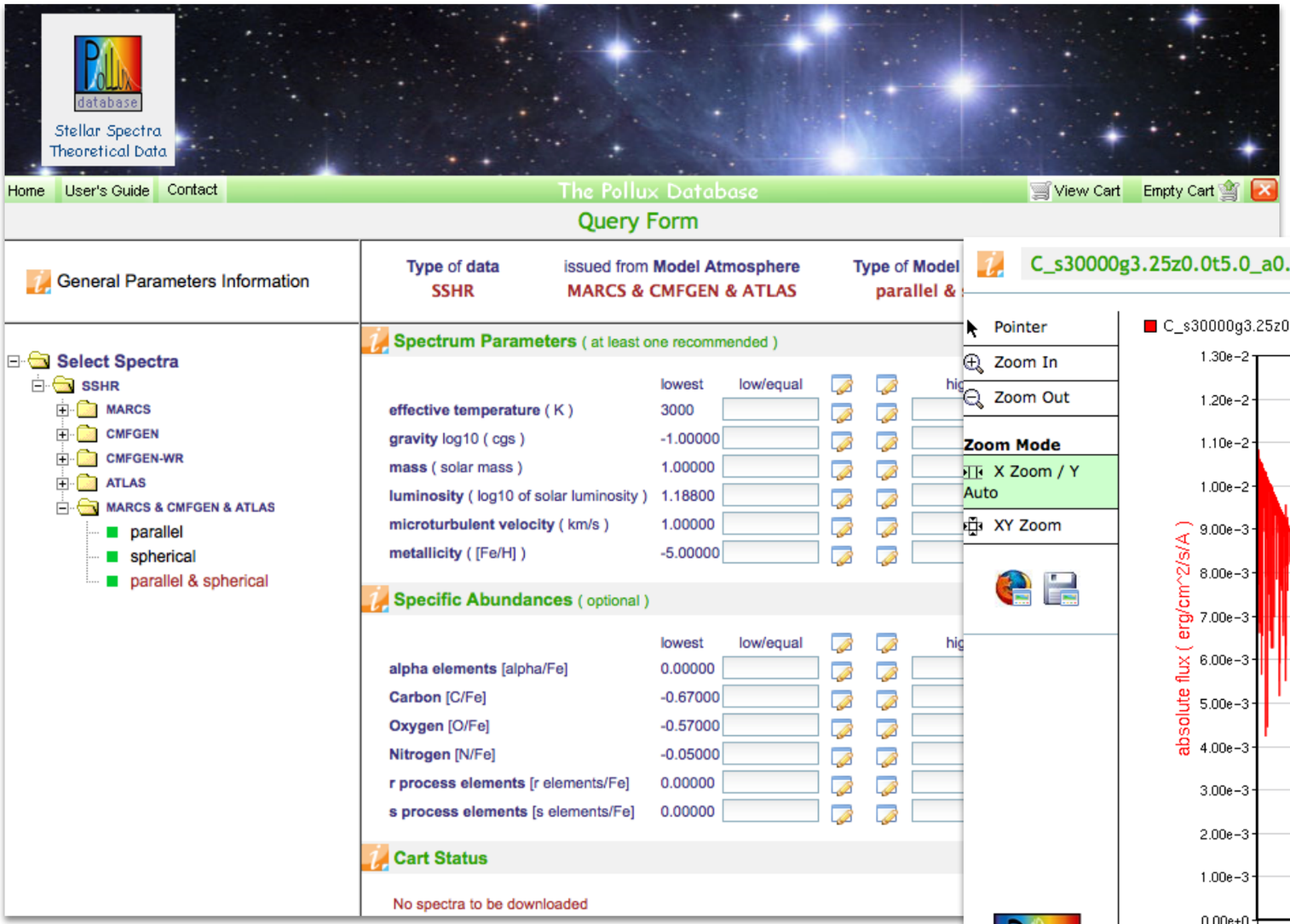


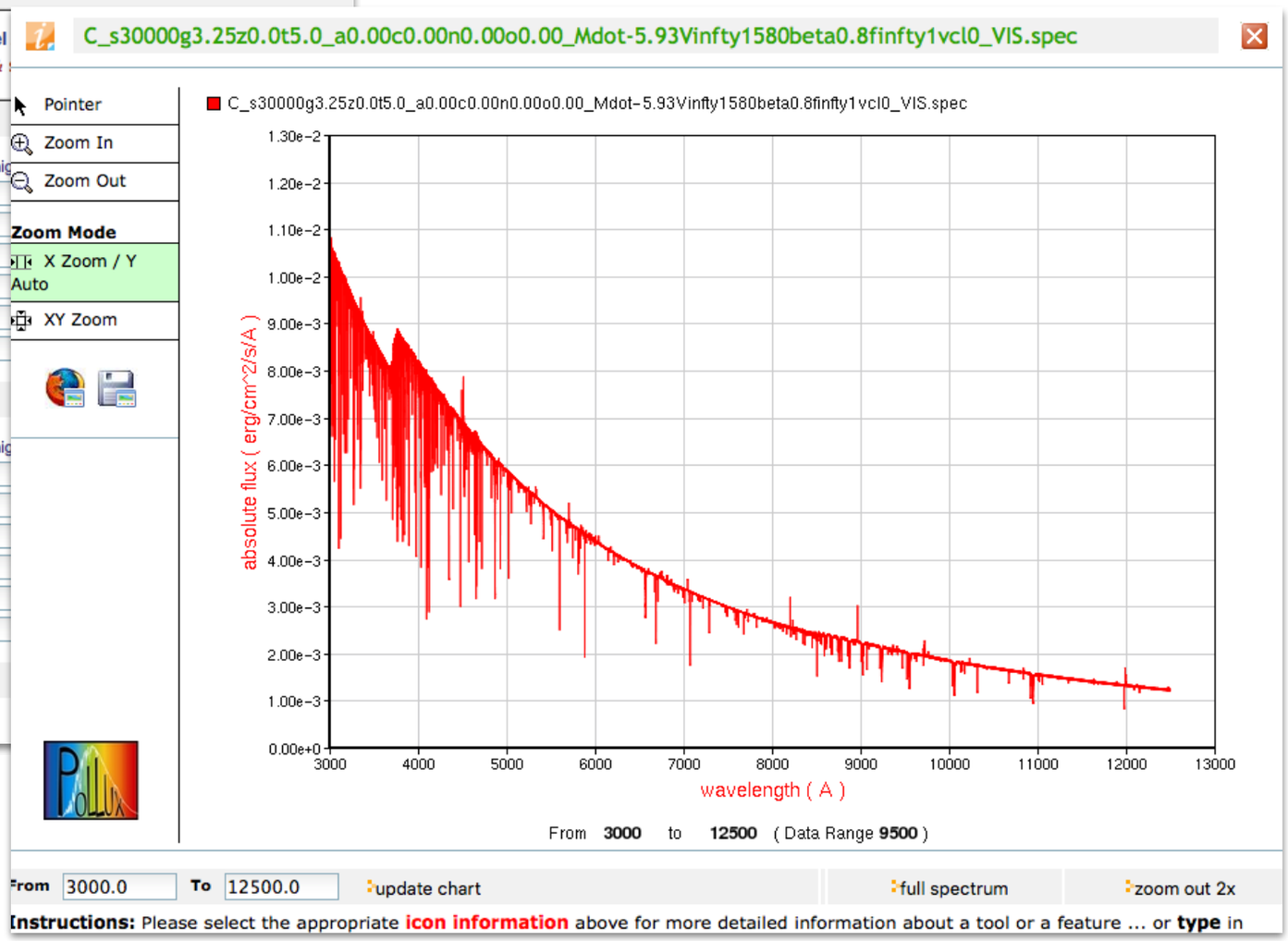


Example: Stellar synthetic spectra

POLLUX database: synthetic stellar spectra

http://pollux.graal.univ-montp2.fr





- Access to synthetic stellar spectra
- Fully VO-Compatible
  - SSA used of theoretical spectra
  - VO-Tools: CASSIS, VO-Spec, …

IVOA - **Evolution**

Diffusion of simulation in **Science Plateforms**

TWiki > IVOA Web > IvoaTCG > ProgramPrepVirtualMay2021 > InterOpMay2021 > InterOpMay2021SPW (2021-05-28, GiulianoTaffoni) Edit Attach

Theory GWS Science Platform Workshop Schedule - IVOA May 2021 Interoperability Meeting

[back to main programme page]

Schedule Summary						
Session	DateTime UTC	UTC-08:00	UTC-05:00	UTC+01:00	UTC+08:00	UTC+11:00
		Victoria BC/Pasadena	*Washington DC*	*Strasbourg*	*Perth/Beijing*	*Canberra*
SP1	May 25 13:30	May 25 06:30	May 25 09:30	May 25 14:30	May 25 21:30	May 25 23:30
SP2	May 26 13:30	May 26 23:30	May 26 09:30	May 26 14:30	May 26 21:30	May 26 23:30
SP3	May 27 13:30	May 27 23:30	May 27 09:30	May 27 14:30	May 27 21:30	May 27 23:30

Theory/GWS Workshop 1

Time: Tuesday May 25 13:30 UTC

Speaker	Title	Time	Material
Simon O'Toole	Introduction	10'+2'	
Gerard Lemson	Cosmological simulations on the SciServer science platform	10'+2'	pdf
Jorge Carretero	Handling Cosmological Simulations at the Port d'Informació Científica (PIC)	10'+2'	pdf
Darren Croton	The Theoretical Astrophysical Observatory	10'+2'	pdf
Shy Genel	Data-sharing at the CCA: Binder and FlatHUB	10'+2'	pdf

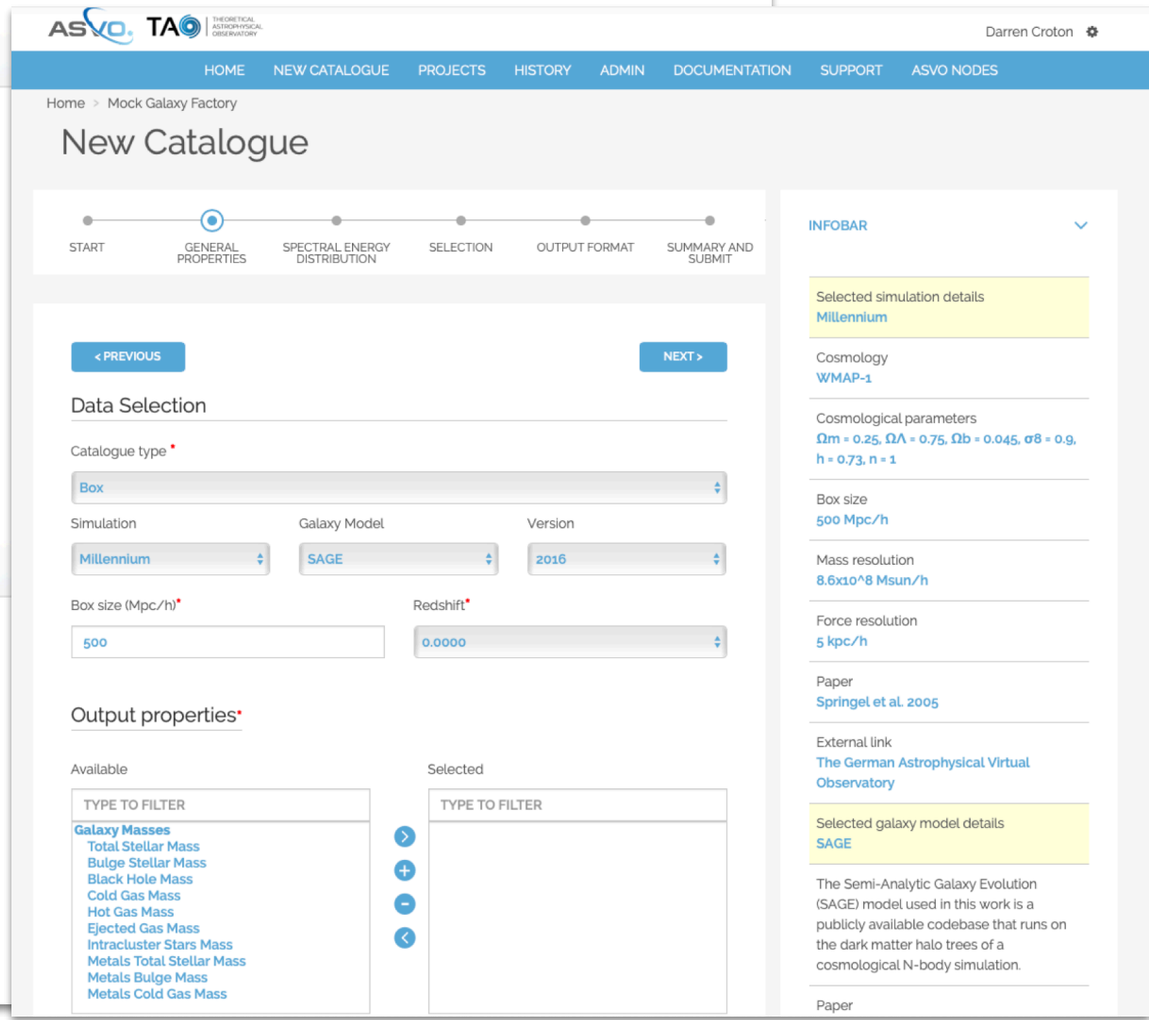
Moderator: Giuliano, Notetaker: Simon, **notes: Etherpad link**

Theory/GWS Workshop 2

Time: Wednesday May 26 13:30 UTC

Speaker	Title	Time	Material
Klaus Dolag	A web portal for hydrodynamical, cosmological simulations	10'+2'	pdf
Dylan Nelson	The IllustrisTNG Simulation Data Platform	10'+2'	pdf
Katrin Heitmann	Simulations for Large-scale surveys	10'+2'	pdf
Matt Turk	Infrastructure for Cosmological Simulations	10'+2'	slides
Andrew Benson	Disseminating Galacticus data through online platforms	10'+2'	pdf

Moderator: Gerard, Notetaker: Simon, **notes: Etherpad link**



IVOA - **Evolution**

Are simulations more difficult to publish than observations ?
Simulations main specificities:
- more heterogeneous
- complex: anything that can be thought can be simulated
- no standard raw data format as FITS

Heterogeneity & complexity are solved with SimDM

Now VO-Theory standards exist, simulations are less difficult to publish

Possible reasons
- VO-Theory standards come lately for some projects
- A question of culture
  - Data centers publishing observational data exist for a long time
  - Publishing of theoretical data is more recent and less organized

- Data centers are focused on observational data: **few help for theorists**
- Requires **time**, **procedures, experience** & **manpower** to publish data
- Needs evolution and maintenance of services

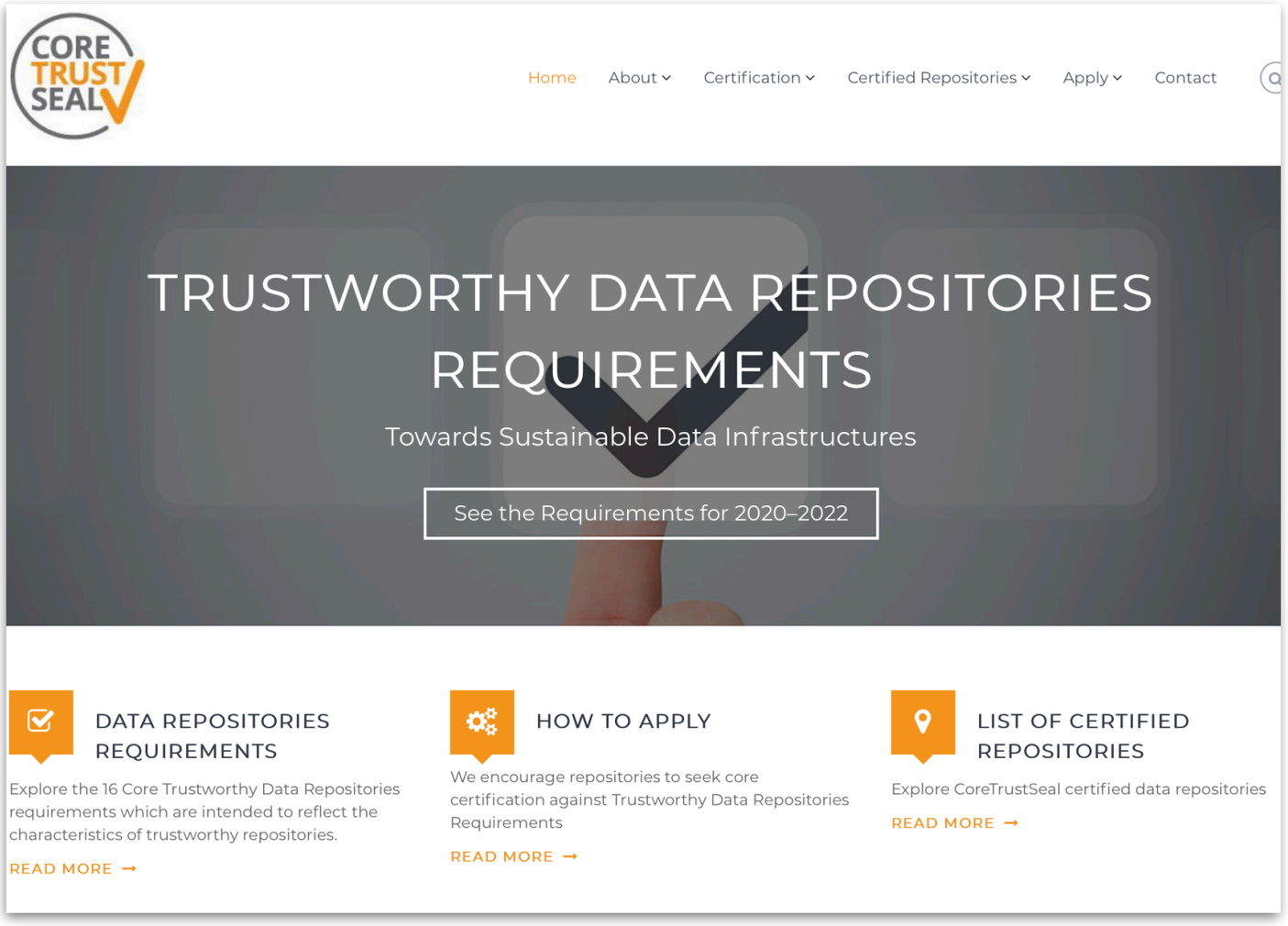
CoreTrustSeal

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



Actuellement : CDS & IFREMER-SISMER

Conclusion

Contexte favorable à la Science Ouverte

- Incitations de l'Europe, du MEN pour une diffusion **FAIR** des données
- Financements : appels à projets

IVOA : une partie du travail est fait

- Standards SimDM, SimDAL, …
- "Plateformes" VO compatibles existent pour diffuser ses simulations : GALACTICA, ISMDB, POLLUX
- Evolution des discussions vers les Science Plateforms

➔ Faut-il s'organiser pour mieux diffuser nos simulations en France ?

ASOV peut contribuer

- Soutien des missions pour collaborations / développement de l'interopérabilité
- Discussions INSU

Inscription mailing lists ASOV

Deux listes

- OV France : liste générale (réunions, AO, annonces)
- OV France dev : liste plus orientée pour les développeurs (échanges d'expertises)

Pour s'inscrire :
Envoyer un mail à sympa@services.cnrs.fr
avec comme objet :

```
subscribe ov-france Prénom Nom
```

Ou

```
subscribe ov-france-dev Prénom Nom
```