

PREDON

Scientific Data Preservation Project

Call: Mastodons

MASTODONS

- Multi-disciplinary Department of CNRS launched a call in April 2012:
 - Data exceeds storage
 - More science in Data
- Possible Directions:
 - Stockage et gestion de données (par exemple, dans le Cloud), sécurité, confidentialité.
 - Calcul intensif sur des grands volumes de données, parallélisme dirigé par les données.
 - Visualisation de grandes masses de données.
 - Extraction de connaissances, datamining et apprentissage.
 - Qualité des données, confidentialité et sécurité des données.
 - Problèmes de propriété, de droit d'usage, droit à l'oubli.
 - **Préservation/archivage des données pour les générations futures.**

PREDON

- **Projet: PREDON** C. Diaconu (CPPM), G.Lammana (LAPP). S. Kraml (LPSC)
- **Extend the scientific, technical and collaboration assets of (dp)HEP and astroparticles**
 - Define a robust environment for scientific data preservation:
 - Policies
 - technical solutions
 - long term support

Project

- Short term (2012):
 - Demonstrate the interests of several national labs in complex scientific data preservation
 - CPPM/IN2P3 Univ. Aix-Marseille
 - LAPP, LPSC and APC
 - OAMP/LAM (INSU)
 - LIRMM (INS2I) Univ. Montpellier
 - Univ. Paris 5
 - Univ. Paris 13
 - CC-IN2P3 centre de calcul IN2P3
 - CINES
- Medium Term (2013/2014)
 - Communication: exchanges and workshop
 - White book specifications
 - Demonstrator of multi-discipline access and preservation unit (focus on scientific complex data)
- Long term (2015/2016)
 - Install “National Data Observatory”
 - Coalition of large centres and new multi-disciplinary data access projects
 - Survey/support production and preservation of scientific data

Reunion Mastodons 5 Decembre

- Cette première année de projets étant réduite à quelques mois seulement, il ne paraît pas raisonnable de mettre en place un processus d'évaluation individualisé des projets. L'objectif de ce premier colloque Mastodons sera donc une occasion d'avoir une vue d'ensemble des projets et d'initier un forum de dialogue entre les partenaires de ces projets et avec la MI du CNRS.
- Dans ce but, je sollicite de chaque porteur de projet une présentation selon le format suivant:
 - Challenges scientifiques du projet à 4 ou 5 ans et affinage éventuel de ces challenges depuis la soumission des projets,
 - Organisation du projet, modalités de travail collaboratif, agenda des réunions techniques de l'année 2012,
 - Premiers résultats scientifiques obtenus ou le cas échéant identification des premières pistes de recherche explorées et positionnement par rapport à l'état de l'art,
 - Objectifs scientifiques pour 2013, en donner une vue plus détaillée que celle du point 1.
 - 1 slide sur l'utilisation du budget alloué au projet.

Pistes

- Robust systems/technologies for data preservation:
 - DPHEP-connection, CC-IN2P3, CeSAM
 - “Preservation des bits”
 - Infrastructures/technologies: clouds, access....
- Methodes dans la preservations de données:
 - CC-IN2P3/CINES/CeSAM
 - Workflows, specs, organisation, indexation, metadonnees
- Scientific data formats for data preservation
 - HEP case
 - Astro case