



Centre de Données de la Physique des Plasmas



L'Observatoire Virtuel et la Physique des Plasmas Spatiaux

Christopher C. Harvey

Centre de Données de la Physique des Plasmas

<http://cdpp.cesr.fr>

cdpp@cesr.fr, harvey@cesr.fr



Résumé

- Le Centre de Données de la Physique des Plasmas est issu d'une collaboration CNRS/INSU/CNES formalisée par un protocole signé en janvier 1998.
 - Il a ouvert son service en octobre 1999
 - Vincent Génot en parlera demain matin.
- Puisque :
 - Aucun centre ne peut archiver les données de tout le mode, et que
 - cette solution n'est même pas souhaitable,
- Il est rapidement devenu apparent que l'interopérabilité entre centres de données de physique spatiale est essentielle.
- Cette présentation est :
 - un historique brève de ce qu'a été fait de 1998 jusqu'à aujourd'hui, et
 - une présentation des perspectives actuelles.



A Long History

1998. In September an ISTP science workshop meeting in RAL recommended that larger data centres should “do something” to make data more accessible and more easily usable. Efforts initiated by CDPD, NSSDC, SwRI and RAL.

2000. A proposals was submitted to the NASA AISRP (Advanced Information Systems Research Program). Proposal rejected, because “Implementation not at the leading edge of Information Systems Technology”.

2002 : ESA Science Director announced his intention to propose **CAA** (Cluster Active Archive) to support science analysis of data from the Cluster mission. The **SPASE** (Space Physics Archive Search and Extract) collaboration was formalise by a charter agreed in October.

2003. **CAA** approved by ESA/SPC at its meeting in February.

SPASE consortium submitted a proposal to NASA’s “Living With a Star” (LWS) “Targeted Research and Technology” (TR&T) program.

2004. **SPASE** TR&T proposal accepted with funding sufficient to support significant activity in USA until mid 2007. (No European support from NASA)

After a slow start, ESA/CAA started :

- supporting the effort of Cluster European PIs to produce archivable datasets, and
- building the archiving system.



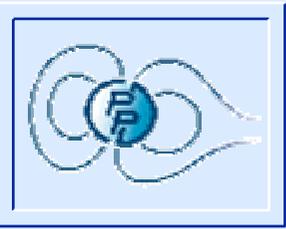
Interoperability

- The science user requires information, data, and services.
- His local data system should appear as part of a single coherent global system, which in reality it consists of :
 - many geographically dispersed centres,
 - which store data from different missions or different experiments,
 - possibly in different formats, and
 - offer different value-added services.
- The objective of “Interoperability” is to allow the user, wherever he is situated, to find via a graphical interface with which he is familiar:
 - the **information** he wants,
 - the **data** he needs, and
 - the **tools** to exploit it,
 - all in a form which he can **readily use**.
- Thus interoperability creates a “Global Virtual Observatory”



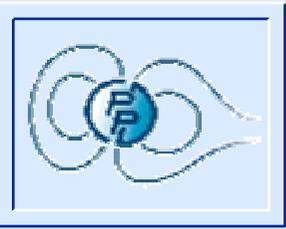
SPASE Objectives (1/2)

- Space Physics Archive Search and Extract (**SPASE**) is an international consortium of space physics archiving organisations.
- It is developing a **data model** and **recommendations for use** of that data model, which will :
- Permit data centres to implement tools to allow their science users :
 - to **find** space physics data which interests them, wherever it is,
 - **intercompare** the data found, and
 - **retrieve** selected datasets or portions of datasets.
- Search across **multiple** data centres through a **single** query
- With a **homogeneous** presentation of the results, using **common** terminology, and taking account of each centre's ability to **discriminate** when searching.
- SPASE **will** demonstrate the feasibility of such a system.
- SPASE **will not** run the individual archives, which remain the responsibility of their parent organisations.



SPASE Objectives (2/2)

- 1) **Demonstrate** a system that meets the global Space Physics User Functional Requirements.
- 2) Facilitate circulation of Space Physics **scientific & technical information**.
- 3) Facilitate the (two-way) **interface** with international organisations responsible for technical standards and developments.
- 4) **Avoid duplication** of effort between space physics archiving centres.
- 5) Ensure **compatibility of architecture** used for the global distributed system with as many of the individual data centres as possible.
- 6) **Minimise costs** by development of a system that :
 - uses widely approved technical standards,
 - Interfaces with existing (legacy) data centres,
 - is easy to maintain (both globally, and in each archiving centre),
 - can evolve so as to profit from likely future technological development,
 - can interface with adjacent disciplines and, one day, with a global system.



Principal SPASE Participants

- CNRS/CNES Plasma Physics Data Centre (CDPP)



- NASA/National Space Science Data Center



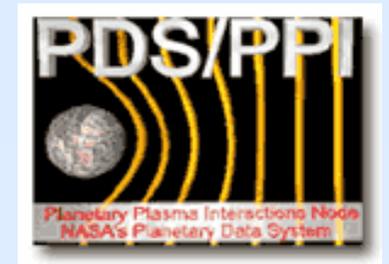
- Rutherford Appleton Laboratory



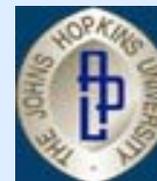
- Southwest Research Institute



- Planetary Data System - UCLA Plasma Physics Interactions Node



- Applied Physics Laboratory





Cluster Active Archive

- Cluster Active Archive est un projet ESA.
- Il y a deux activités distinctes, qui utilisent chacune environ 50% des ressources allouées par le SPC :
 - la **collecte des données et des metadonnées**, avec l'aide des CDDs employés dans les labos PIs et financés par l'ESA ;
 - la **construction du système d'archivage** : infrastructure, logiciels, ..., avec l'aide de trois ingénieurs à temps complet installés à l'ESTEC.
- Le projet est géré à partir de l'ESTEC/RSSD, où le Project Scientist Trevor Sanderson sera bientôt remplacé par Harri Laakso.
- Le Responsable Technique du projet est Chris Perry, mis à disposition 50% par le RAL sous contrat ESA.
- Le CDPP participe aux deux activités citées ci-dessus :
 - Il aide à la préparation des données (service rendu aux PIs français) ;
 - Responsabilité (contrat ESA) pour le Metadata Dictionary utilisé par CAA.
- Le CDPP doit être prêt à se positionner pour l'après CAA (archivage long-terme), peut-être dans le contexte Européen décrit ci-après.
- Le CDPP archive les jeux de données CAA issus des expériences françaises.



Space Plasma Data Description

- Metadata is data which describes the data. Its use is twofold :
 - to find data within archives, and
 - to exploit the data once it has been found and recovered.
- The requirements are not the same for these two purposes, but there is considerable overlap.
- The metadata should contain all the information required :
 - to find the data (can be difficult for in-situ measurements made along a trajectory in space-time),
 - to read the data,
 - to understand the data (syntactic description), and
 - to understand what it represents (semantic description), including how it was obtained : the latter impacts upon the scientific significance of the data.

When searching for data, the precision of the results cannot be better than our ability to describe what we want.



Ce qu'on a appris

- SPASE et par CAA, tous les deux, ont beaucoup investi dans leurs dictionnaires de metadonnées.
- Au début du projet (en 2003), CAA pensait utiliser le SPASE Metadata Dictionary. Cela n'a pas été possible, parce que :
 - CAA ne s'occupe que d'une seule mission, mais a besoin d'une description :
 - » **fine, pour révéler la richesse des mesures, et**
 - » **stable, pour permettre la rédaction de documents contractuels (ICDs).**

La version 2 du CAA Metadata Dictionary vient d'être publiée :
http://www.cesr.fr/~harvey/DataD_V20.pdf

- SPASE est par définition multi-mission, et a rencontré beaucoup de problèmes lié aux concepts parfois élémentaires. Par exemple, le concept de « Jeu de données » est :
 - » **pour les plasmiciens, un séquence d'informations de la même nature, issues du même instrument, traitées par les mêmes logiciels, et archivées dans des fichiers du même format ;**
 - » **pour la physicien solaire, un ensemble de toutes les informations connues, en provenance des instruments dans l'espace ou au sol, et qui rapportent au même événement sur le soleil.**



Aujourd'hui (1/2)

- Le dictionnaire CAA a été conçu pour être extensible à d'autres missions « plasma ».
- Le CDPP va essayer d'utiliser ce dictionnaire pour sa nouvelle version, le SIPAD-NG actuellement en développement.
- Le dictionnaire SPASE est toujours en évolution, maintenant à cause de problèmes rencontrés lors de l'implémentation XML. Il va falloir approfondir l'utilisation de l'ontologie.
- Il me semble (opinion personnelle) que :
 - Chaque data centre doit maintenir son propre dictionnaire pour pouvoir maîtriser :
 - » **la qualité (précision, étendu) de la description,**
 - » **l'évolution (contenu et calendrier)**
 - Le dictionnaire SPASE sera en évolution régulière, et à tout moment fournir une description des données moins détaillée que celle d'un centre disciplinaire, mais plus globale (multi-disciplinaire).



Aujourd'hui (2/2)

- Chaque participant SPASE est en train de préparer, en utilisant le modèle actuel de données, la description XML de quelques jeux de données, pour permettre la réalisation d'une maquette d'interopérabilité. (Au CDPP on utilise deux jeux de la mission WIND)
- En parallèle, SPASE essaie d'étendre son dictionnaire pour pouvoir offrir un service le plus complet possible.
- Une réunion plénière de revue et de planning SPASE est prévue au SwRI (San Antonio, TX) le 10-11 mai.
- CAA est en train de réceptionner les premiers jeux de données Cluster, et l'archive sera en β -test à partir du mois de mai. Les données et metadonnées seront toutes livrées en format standard CEF-2 développé pour et par Cluster.
- CAA doit commencer à penser aux outils d'exploitation des données. Avec un format de livraison standard, on peut développer des outils qui seront ensuite échangés entre les différents laboratoires. Il faut coordonner les développements pour s'assurer que les besoins sont satisfaits tout en évitant la duplication inutile.



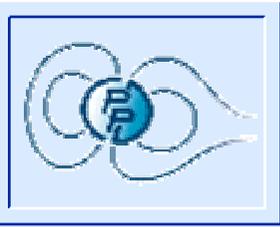
Vers un Observatoire Virtuel « Systeme Solaire »

- Tous les participants aux activités en cours rendent bien compte de la nécessité d'interagir les uns avec les autres :
 - CAA, CDPP, EGSO, et d'autres (EuroPlaNet ?) ou qui pourraient être créés,
 - en utilisant SPASE comme « colle » (outil de cohésion).
- Un atelier sur ce thème a été organisé par le programme NASA « Living With a Star », à Greenbelt MD en octobre 2004.
- Un compte-rendu a été produit, que je vous conseille fortement :
“A Framework for Space and Solar Physics Virtual Observatories”
http://lwsde.gsfc.nasa.gov/VO_Framework_1_Jan_05.pdf
- Sur ~100 participants présents, il n'y avait que 5 européens.
- Les participants européens ont noté que si l'Europe n'est pas plus d'active, la totalité de l'organisation des observatoires virtuels pour le système solaire se fera aux USA.



Informations Issues de l'Atelier LWS-DE

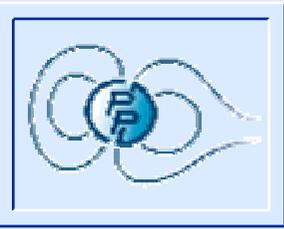
- Un « Observatoire Virtuel » est un ensemble de « dépôts » de données (“ repositories ” en anglais) qui paraissent pour l'utilisateur comme une seule base de données.
- Il fournit un accès uniforme à l'ensemble de données et services.
- En 2005 la NASA va lancer un Appel d'Offres pour des « Observatoires Virtuels » pour le Système Solaire.
- On attend des propositions pour (liste non-exhaustive) :
 - Virtual Solar Observatory
 - Virtual Heliospheric Observatory
 - Virtual Magnetospheric Observatory
 - Virtual ITM Observatory (Ionosphere, Thermosphere & Mesosphere)
- Les différents groupes américains se préparent activement.
- Ces observatoires du Système Solar doivent avoir un certain (TBD) niveau d'« interopérabilité » entre eux.
- Les américains ont l'intention de marquer le 50^{ème} anniversaire de l'IGY, en 2007, par un grand événement « Outreach ». [Colloque d'Ouverture de l'ASOV, 2005/04/04](#)



Activités récentes

L'atelier LWS-DE a stimulé quelques initiatives en Europe :

- Une campagne pour persuader la direction scientifique de l'ESA qu'il faut soutenir, ou au moins encourager, plus activement l'exploitation des données plasma spatiales européennes. On a été aidé par la météo de l'espace (lors d'un atelier à l'ESTEC).
- Une proposition de « Coordination Action » a été soumise à l'UE : CASSIS, Coordination of Access to Solar System Information Systems, <http://www.mssl.ucl.ac.uk/grid/FP6-proposal/>
- Le but de cette proposition est :
 - sensibiliser la communauté « Système Solaire » aux possibilités offertes pour avancer sa recherche,
 - organiser des ateliers pour stimuler la collaboration inter-disciplinaire,
 - utiliser ces ateliers pour étudier la méthodologie de travail des chercheurs, pour mieux définir les besoins scientifiques.
- Le sort réservé pour CASSIS n'est pas connu.
- D'autres activités européennes sont à l'étude.



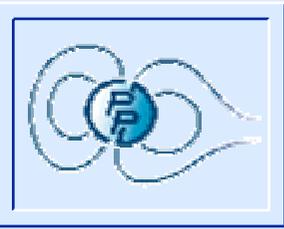
Conclusion

- Différents centres de données qui couvrent un domaine scientifique commun doivent créer des observatoires virtuels pour leurs disciplines (déjà fait par EGSO, à continuer avec le CDPP et le CAA).
- Les observatoires virtuels dans des disciplines voisines, par exemple, tous ceux du Système Solaire, doivent viser un certain niveau (à être défini) d'interopérabilité.
- L'interopérabilité doit être bâtie sur l'existence d'un dictionnaire adéquat.
- Un grand effort est nécessaire pour :
 - Coordonner le développement des outils standards d'analyse, dont certains peuvent être offerts en forme de "Web Services" ;
 - faire accepter un standard pour le format d'échange des données et des metadonnées, ou
 - à la limite, développer des logiciels de conversion de format.
- Sensibiliser la communauté scientifique aux avantages long-terme de l'utilisation de formats et d'outils standards.
- Tout en poursuivant la collaboration trans-Atlantique, essayer d'instaurer un minimum de coordination, et de financement, européenne.



End of the presentation

Thanks



Présentations utilisées pour la préparation de celle-ci

- `\CDPP\transpar\Atelier_CNES_Metadonnees_041125.ppt`
- `\CDPP\transpar\PNST_041019.ppt`
- `\CDPP\transpar\AtelierOVF_041012.ppt`
- `\CDPPcollaboration\SPASE\PV2004-SPASE_041007.ppt`
- `\CAA\Presentations\CAAMetadataDic.ppt`