

Le format NeXus,

un pas de plus vers l'homogénéité



“

Nexus est un bon candidat pour harmoniser les données”

Stéphane Poirier,
division Informatique de
SOLEIL – groupe « contrôle
commande acquisition »

Qu'est-ce que le format NeXus, quel est son intérêt?

Le format NeXus n'a pas été inventé pour le synchrotron SOLEIL. Il est né aux alentours de 1996 dans des instituts de recherche qui utilisent des sources de neutrons pour analyser la matière. Les laboratoires avaient besoin d'un format permettant d'enregistrer des données plus importantes en volume et en variété. On peut y stocker à la fois les données expérimentales et les données du contexte de l'expérience. Ils se sont appuyés sur un format déjà existant développé dans les laboratoires américains : le format HDF (hierarchical data format), extrêmement puissant mais assez complexe à utiliser. Le format NeXus est une couche d'accès qui simplifie l'utilisation de HDF et spécifie aussi le contenu des fichiers. La plus importante valeur ajoutée de NeXus est qu'il n'impose pas de contraintes en terme de nombre et de volume d'informations que l'on peut stocker dans le fichier, contrairement à la plupart des formats jusqu'alors utilisés.

Pourquoi le format NeXus est-il intéressant pour SOLEIL?

Je suis arrivé à SOLEIL en 2004 et ma mission était de définir un moyen d'enregistrer les données : quel format et avec quels outils. J'ai consacré mes premiers mois à la recherche du format « idéal » et je suis tombé sur NeXus. Les sources de neutrons et de rayonnement synchrotron permettent des recherches qui sont relativement proches, donc la spécification que j'ai trouvée avec le format NeXus était applicable en l'état. Aujourd'hui, de plus en plus, sur les lignes de lumière, les mesures enregistrées représentent des volumes d'information très importants et qui vont toujours croître. Ainsi, les formats textes classiques que les gens utilisaient auparavant se sont révélés inadaptés. D'autre part, l'une des contraintes qui était imposée pour le site était d'utiliser la même chose pour toutes les lignes. Il fallait donc un format générique pour répondre aux besoins divers. Il y a par ailleurs un effort au niveau européen pour essayer d'harmoniser les formats de données et NeXus est un bon candidat. Mais ce n'est pas le seul. Maintenant, à SOLEIL, en

collaboration avec un institut australien (ANSTO), nous essayons d'aller au-delà de NeXus et de nous affranchir complètement du format de données. Avoir standardisé le format des fichiers de données produites par ses lignes de lumière permet à SOLEIL de disposer d'une « bibliothèque » de données expérimentales, bibliothèque dont on peut explorer informatiquement le contenu, par exemple par le biais de l'application Web TWIST. SOLEIL possède ainsi la mémoire scientifique des expériences qui ont eu lieu sur ses lignes. Gérer les données expérimentales à large échelle est novateur dans le monde des synchrotrons, et nous positionne très favorablement dans les différents projets européens (PANDATA, PNI...) liés à cette problématique. Aujourd'hui, nous profitons de cette position pour essayer de convaincre les informaticiens des instituts partenaires (ESRF, DESY, etc) de nous rejoindre dans l'approche « CommonDataModel », projet ambitieux qui a pour but d'unifier l'accès aux données scientifiques depuis les logiciels d'analyse de données. Affaire à suivre, donc.

Avez-vous rencontré des difficultés pour faire accepter NeXus à SOLEIL?

Il a fallu convaincre, ce qui est toujours un peu lent. La plus grosse barrière est que, pour exploiter un fichier HDF, il faut un logiciel qui sache le lire. Nous commençons toute juste à développer les premiers outils d'analyse qui utilisent directement les données au format NeXus. Dans l'intervalle, pour que les gens puissent travailler, nous avons développé une « moulinette » qui permet de relire les fichiers NeXus qui sont produits et de récupérer les données dans des fichiers textes. Il y a un outil, FOXTROT, qui est en production sur la ligne SWING. Il permet d'analyser les images dont les données sont écrites au format NeXus. Il est déjà en production sur cette ligne et on est en train de commencer à le déployer sur d'autres lignes.

→ Pour en savoir + www.synchrotron-soleil.fr/Soleil

ToutesActualites/2009/TWIST

→ Contact : stephane.poirier@synchrotron-soleil.fr

EN BREF

SOLEIL IMPLIQUÉ DANS LE RÉSEAU PARIS-SACLAY-INNOVATION

SOLEIL fait partie des 18 acteurs de la valorisation de la recherche, de la création et du développement d'entreprises de technologies innovantes regroupés dans le réseau Paris-Saclay-Innovation pour permettre au monde économique d'accéder aux moyens expérimentaux de recherche et aux compétences scientifiques et techniques des laboratoires du plateau de Saclay et de la vallée de l'Yvette. Le réseau anime un site web bilingue de présentation de son offre (www.paris-saclay-innovation.fr) et était présent au Salon Technov d'Orly le 10 février 2011.

L'ICM CHOISIT LA MÊME SOLUTION DE STOCKAGE INNOVANTE DE DONNÉES QUE SOLEIL

L'Institut du Cerveau et de la Moelle épinière (ICM) vient de se doter de la solution de stockage innovante de données scientifiques Active Circle, déjà mise en œuvre à SOLEIL. Une visite par l'ICM des installations de SOLEIL et une discussion avec les experts de la Division Informatique ont contribué au choix final de la solution Active Circle, proposée par une entreprise française dont le siège est situé à Jouy-en-Josas, à proximité immédiate du plateau de Saclay.