

Réunion annuelle du CU du CDPP, IRAP+distanciel, 26 avril 2024

Participants (CU) : B. Grison BG (IAP, Rép. Tchèque, dist.), N. Aunai NAU (LPP), S. Hess SH (ONERA), L. Griton LG (LESIA, *dist.*), L. Hadid LH (LPP), R. Maggiolo RMA (IASB, Belgique, *dist.*), R. Modolo RMO (LATMOS, *dist.*), P. Henri PH (LPC2E, *dist.*)

Absent (CU) : E. Astafyeva EA (IPGP)

Participants (CDPP): N. André NAN (IRAP, président), B. Renard BR (AKKA) V. Génot VG (IRAP), M. Bouchemit MB (IRAP), C. Jacquy CJ (IRAP), A. Rouillard AR (IRAP), I. Plotnikov IP (IRAP), D. Boucon DB (CNES), D. Leung DL (CNES), D. Heulet DH (CNES), B. Cecconi BC (LESIA), Alexis Jeandet (LPP)

En gras, les recommandations du CU et autres points susceptibles de figurer dans la présentation au CD.

Introduction (NAN)

NAN présente l'OdJ qui est remanié afin que PH (absent l'après-midi) soit présent lors des discussions sur les BdD de simulations numériques dans AMDA.

09:30-10:00 Présentation de sciQLOP (NAU, AJ)

À partir d'un historique sur les limitations de l'analyse des données (pour accéder aux données, puis pour faire un plot, puis faire des études statistiques), NAU dresse le constat suivant : de nombreuses données ne sont pas utilisées pour de mauvaises raisons. AMDA a en partie résolu le problème (facilités pour les plots multi-sources, data mining "conditionnel"), mais il reste encore quelques limitations : fonctions restreintes pour transformer des données ("Derived Parameters"), le data mining à partir de conditions explicites (elle-même limitées par les Derived Parameters) ne permet pas la détection des signatures complexes qui nécessitent des techniques d'apprentissage), lenteur de la navigation dans les données... SciQLOP vise à continuer à développer la vision du CDPP et d'AMDA via l'utilisation de technologies modernes.

Autour de SciQLOP :

Speasy (Equipe CDPP principalement, AJ pour le développement) : Package python open source qui gère l'accès aux données. Un système de cache hiérarchique permet un accès en temps optimal. C'est un projet actif avec de nombreuses versions publiées. Il devient également un composant interne du serveur AMDA.

CDFpp (développement AJ) : Nouvelle bibliothèque C++ son interface python et pour la manipulation de fichiers CDF. Plus rapide, thread safe et facile à installer, elle remplace avantageusement la CDFLib de la NASA.

SciQLOP permet la visualisation des données, une navigation rapide et la création efficace de catalogues d'événements. Les Jupyter Notebook intégrés dans SciQLOP permettent à l'utilisateur de visualiser et partager ses propres fonctions.

TSCat : package python permettant l'exploitation de catalogues d'événements.

Roadmap (futurs développements) :

Les projets de développements concernent l'amélioration de l'installation pour toutes les plateformes, l'amélioration de l'ergonomie, l'ajout d'informations contextuelles (labellisation automatique via IA, lien doi vers un papier étudiant l'événement regardé), le partage des catalogues, la collaboration en simultanée de plusieurs utilisateurs, suggestion par une IA d'événements quand on visualise les données.

L'écosystème SciQLOP regroupant tous ces outils est développé et diffusé sur une organisation GitHub et des dépôts respectifs.

Le workshLOP est prévu le 12-13/06 à Jussieu. Le CU trouve que c'est une bonne idée et SH s'enquiert de la possibilité d'en faire un à Toulouse. Le CU soutient l'ensemble des développements faits pour SciQLOP et la mise en place d'une organisation Github et des dépôts associés. Ceci permet une visibilité optimale des développements et des outils, ainsi qu'une interaction optimale avec la communauté d'utilisateurs et de potentiels contributeurs

10:00-10:30 Premiers tests (il faut l'installer) et retours sur l'outil sciQLOP développé par nos collègues au LPP:
<https://github.com/SciQLop/SciQLop>

BC rappelle que pour les annotations et autres informations contextuelles il est important d'utiliser un standard reconnu (W3C par exemple). Il met en garde contre la tendance qu'ont les conteneurs Docker permettant le déploiement des notebooks à ne plus marcher si personne ne s'en sert pendant 6 mois donc il faut une maintenance (à prévoir). AJ a une expérience différente expliquant que les notebooks de la documentation Speasy s'exécutent depuis plusieurs années avec Binder sans maintenance.

LG demande à quel moment des tutoriaux dédiés seront disponibles pour pouvoir utiliser l'outil SciQLOP avec des étudiants. AJ note que l'équipe est prête à aider et à s'impliquer

pour faciliter l'installation et la prise en main pendant que l'ergonomie d'utilisation, d'installation sont améliorées.

LH a utilisé SciQLOP lors d'une étude appliquée pour recherches précises de cas. Avec un support pour se lancer, l'outil est ensuite très efficace. La connaissance d'AMDA aide beaucoup pour utiliser SciQLOP car les interfaces et la philosophie sont similaires.

BG souligne que l'ensemble est fluide avec quelques fonctions non intuitives (la mollette qui zoom en x et en y à la fois : AJ précise que ce point précis est en train d'être changé). Pour la mise à jour de la fenêtre temporelle, AJ précise que c'est juste un glisser/déposer depuis l'endroit où les dates sont tapées. Il faudrait sans doute masquer par défaut le module « IPython ». Ce serait bien d'avoir un exemple de Jupyter Notebook (NAU précise que les Notebooks sont fonctionnels, BG regardait dans le mauvais répertoire). Il est dommage qu'on n'ait pas eu les diapos de la présentation, ou une démonstration via une réunion en distanciel, avant de faire les tests.

RMO a rencontré des problèmes d'installation sur Windows qu'il a fait remonter. SH a rencontré les mêmes problèmes. NAU précise que la compatibilité avec Anaconda est vraiment compliquée en support, et ce n'est pas la priorité du moment. Une clarification du Readme serait donc la bienvenue. NAU prévoit d'avoir une installation classique (type .exe) pour windows dans le courant de l'année.

RMA n'a pas réussi à installer l'outil sur Linux (plug-in manquant). AJ reconnaît qu'un composant généralement installé dans les distributions de linux est nécessaire au fonctionnement de SciQLOP et qu'il ne peut être installé *via* sciqlop. NAU ajoute que cette erreur bien que rarement rencontrée, pourra être mentionnée dans une FAQ à réaliser et mettre en ligne prochainement.

NAU souhaite développer le partage des fonctions d'analyse de données développées par les utilisateurs ce qui valorise le code écrit par les scientifiques (type marketplace). Le CU trouve que l'idée est bonne mais il faudra prévoir un système de classement pour permettre à l'utilisateur de s'y retrouver : info sur le nombre d'utilisateurs d'un package, avoir un label pour les packages les plus prisés de la communauté, regrouper les feedbacks ou les notes des utilisateurs....

L'outil SciQLOP est encore en cours de développement. Le CU a trouvé l'outil très efficace pour une visualisation des données et une navigation rapide. Le CU encourage fortement le développement de l'outil SciQLOP et en premier lieu à mettre à disposition un package d'installation pour les différentes plateformes ainsi qu'un tutorial. Le CU est disponible pour faire des tests sur le partage des fonctions utilisateurs, fonctionnalité qui n'est pas disponible dans AMDA.

10:30-11:00 Discussion autour d'une évolution future possible d'AMDA vers un AMDAdesktop

BG souligne que SciCLOP, vient combler un manque d'AMDA: besoin de naviguer dans les données plus rapidement. SciQLOP est un outil qui vient combler le manque. SH note que le futur d'amda ce sera peut-être en desktop.

NAU précise qu'il faut désormais distinguer le coté "serveur" d'AMDA, dont le rôle est de distribuer des données (notamment via un webservice efficace), de son interface graphique, cliente du serveur, et qui a vocation à donner à l'utilisateur la possibilité d'explorer des données. Il existait jusqu'alors qu'une seule interface graphique, web, d'AMDA, justifiant l'appellation "AMDA" identique pour le service entier. Maintenant SciQLOP vient offrir une alternative (en mode "client desktop") à l'offre d'interface graphique d'AMDA.

BC trouve important de bien réutiliser ce qui est dans AMDA dans SciQLOP (catalogues). Il faudrait que SciQLOP puisse lire toutes les HAPI d'autoplot. AJ mentionne qu'une issue GitHub existe pour que Speasy comprenne le protocole HAPI et une fois réalisée SciQLOP (qui délègue à Speasy l'accès aux données) en bénéficiera automatiquement. Il est tout aussi important de discuter avec les gens de python heliophysics comme le fait AJ, même si la communauté python hélio n'est pas très réactive, ou refermée sur un petit cercle. BC : atelier au mois de Juin, ce serait bien d'y participer car il faut que SciCLOP s'intègre mieux au réseau que ne l'a fait AMDA.

Le CU considère que SciCLOP est la base naturelle pour une extension d'AMDA pour proposer une interface client desktop. Le CU est attaché au maintien de la version Web d'AMDA en l'état (i.e. privilégier les corrections de bugs éventuels à l'ajout de nouvelles fonctionnalités), qui a l'avantage d'une prise en main très rapide ne nécessitant quasiment aucune connaissance spécifique en informatique. Le CU recommande au CDPP de soutenir le développement d'une version aboutie de SciQLOP.

Le CU encourage le CDPP de participer à des colloques comme par exemple "Heliophysics in Europe workshop" (<https://www.cosmos.esa.int/web/esa-heliophysics/heliophysics-in-europe-2023>).

11:30-12:00 Base de données (BDD) de simulations numériques(SN)

PH remercie le CDPP pour tout le travail effectué et souhaite que les groupes de simulations numériques soient guidés vers des solutions compatibles CDPP. Il n'y a pas de bonnes BDD de SN à l'heure actuelle, même si le CCMC est actif sur cette thématique (SH). La

documentation d'interface existante (Impex) ne semble pas suffisante pour toutes les équipes pour que les outils du CDPP attaquent correctement les BDD. SH et RMO notent qu'il y a eu un changement du format trajectoire AMDA avec potentiellement un changement en local à prendre en compte. VG, SH et RMO vont regarder le document fait à l'époque et comment l'actualiser. BC trouve que dans Impex les interfaces sont bien définies mais que cela peut devenir vite difficile si les interfaces sont complexes. Il manque peut être un module de base pour faciliter le début de la prise en main. RMO va essayer de compléter l'ICD (Interface Contrôle Document) avec des exemples concrets pour faciliter l'ajout d'un autre catalogue de simulation. Pour le développement de webservices, les interfaces sont décrites dans le ICD avec des accès au fichier WSDL. La mise en œuvre pratique de ces webservices dépend de l'infrastructure du serveur hébergeant les webservices (ie des BDD externe).

Lors de la discussion sur l'intégration des BDD de SN dans AMDA, il est rappelé que ce n'est pas le CDPP qui stocke les SN et que c'est au simulateur de choisir celles qui sont partagées. LG donne la priorité aux SN publiées. Au final comme pour les observations on peut utiliser les simulations sans contacter le responsable, mais suivant le niveau de description (caveats, description, ...) il y aura souvent des questions.

En résumé, c'est utile de partager les SN. Un travail a été fait (LATMOS, CDPP), il faut le continuer. NAN note une action pour un workshop pour mieux partager les SN.

11:30-12:00 Présentation de REGARDS et de la migration STAF

L'archive CDPP représente 1400 utilisateurs (dont 700 occasionnels), 1246 commandes, DEMETER comme point d'entrée principal, 240 To (MMS en miroir), 1159 jeux de données (stable sur les 2-3 dernières années – migration)

Les nouveautés concernent GEOS (sol en cdf), WIND (gonio L3), GIOTTO, VEX (L2)

Beaucoup de missions sont en attente (SO/RPW, Juice/RPWI, Bepi Colombo MIO et MPO)

La migration a démarré en 2019, avec une double exploitation jusqu'en 2023. La bascule du site officiel est imminente. Cela a été long (beaucoup de cas particuliers). Il y a maintenant de nouveaux modèles de données compatibles SPASE (la migration des méta données a pris du temps).

REGARDS n'est pas encore accessible en web service mais c'est prévu. BC est étonné d'apprendre que des liens actuels ne marcheront plus quand le SIPAD sera fermé. DH note ce point dans le planning à suivre. Il y a des données qui peuvent être mise en ligne mais l'idée est de faire du stockage (exemple des données WIND ne sont disponibles que la). Le CNES rappelle que tout ne peut pas être mis en ligne.

Le CU recommande, qu'étant donné la vocation essentiellement d'archivage "à froid" des données dans REGARDS, il est important que les données présentes uniquement sur REGARDS soient aussi disponibles "à chaud" via la base AMDA.

Le CU pourra aider à lister des données à mettre en ligne prioritairement.

12:00-12:30 Exploration et retour sur le site web/ la base de données REGARDS du CDPP qui remplace le SIPAD: <https://cdpp-archive.cnes.fr>

La nouvelle base de données marche bien. CJ note que l'interface HM pourrait être un peu plus conviviale (CJ) et il y a des jeux de données un peu caché. Enfin le mail de confirmation de commande ne devrait pas inciter à installer un module de téléchargement puisque les jeux de données seront disponibles via un lien.

Le CU est satisfait des fonctionnalités de la BDD REGARDS et encourage à rétablir le web service existant dans SIPAD le plus rapidement possible.

12:30-12:45 Avis sur l'arrêt programmé de production des données WIND 3DP dans l'archive pérenne du CDPP.

Les données wind 3dp examinées par K. Issautier (action du dernier CU) ne sont pas primordiales pour l'archive (disponibles ailleurs sous format plus lisible, au CDAweb). Les données brutes sont archivées (pas calibrées). La chaîne de production ne peut pas migrer de manière simple (difficultés techniques), **le CU est d'accord pour l'arrêt de la production de ces données car elles sont disponibles sur une autre archive pérenne.** La question du maintien dans l'archive de données non-calibrées se pose.

Il y a aussi une demande de l'ESA pour retrouver les données VEGA pour préparer Comet Interceptor.

14:00-14:30 Retour sur les données disponibles dans AMDA pour Solar Orbiter/Parker Solar Probe

BG et RMA pensent que les données présentes semblent suffisantes pour commencer à travailler. C'est un gros plus d'avoir les données 2D ondes et particules. Le retour sur le travail fait est très positif.

Le CDPP va ajouter de nouvelles données et doit être prêt à ajouter de nouveaux produits (mission toujours active). Ce serait bien d'avoir une idée du nombre des utilisateurs et du nombre de publications.

Le CU soutient l'ajout de nouvelles données Solar Orbiter et Parker Solar Probe dans AMDA.

14:30-15:00 Recommandations pour l'ingestion de futures données manquantes dans AMDA

Une discussion s'engage autour de la complexité de l'arbre et du filtre en temps et en mission qui pourrait être plus convivial. Il est souligné qu'outre l'ajout de données, les outils de transformation de coordination sont à mettre en valeur.

Pour les données elles-mêmes NAU note qu'il n'y a que des données des Prime parameters pour Double Star et qu'il faudrait que plus de données complexes ou plus lourdes, qui n'étaient pas proposées par AMDA car impossible à manipuler via l'interface Web d'AMDA, comme les fonctions de distribution soient disponibles car elles peuvent maintenant être accédées pour analyse via Speasy et via SciQLOP. NAN rappelle qu'il y a des fonctions de distribution disponibles pour MMS. PH note l'absence de données DEMETER dans AMDA qui a des raisons « historiques » rappelées par CJ, le LPCE ayant eu son propre centre de données. Il y aura bientôt la mission Escapade (NASA) vers Mars.

Pour ce qui est des imageurs (par exemple pour la future mission SMILE), 3d-View est un outil plus naturel. Les données de Bepi manquent car les données de la phase de croisière ne sont pas publiques. AMDA sera le distributeur de données Bepi aux PI. La présentation des données des modèles de champ magnétique (Jupiter) devrait être plus homogène (RMO).

La question se pose de l'ajout de données d'occultation radio (RMO), mais celles-ci sont difficiles à visualiser (BC). Pour la diffusion de données radio DAS2 est plus adapté que HAPI (qui est plus léger).

La discussion s'étend et porte sur la possibilité d'avoir une instance locale qui récupère les kernels (distribués par la nasa) tous les soirs et sur la possibilité de déposer ses données au CDPP ce qui serait plus pratique que zenodo.

Le CU suggère dans un premier temps d'ajouter les données Double Star, de réfléchir à quels jeux de données DEMETER à intégrer, et plus généralement de réfléchir aux données REGARDS à interfacier prioritairement avec AMDA.

15:00-15:30 Nouvelles fonctionnalités de AMDA (histogramme)

VG présente ce nouvel outil qui permet de faire des histogrammes 1D, 2D et 3D dans le plot manager

L'utilisateur doit régler les axes à la main ce qui est un peu compliqué et provient d'une limitation technique. Il semble qu'une erreur non reproductible se produit (fond noir chez RMA).

15:45-16:00 3Dview et PropagationTool: problème de stabilisation des outils sur des plateformes différentes

VG montre des réalisations 3dView pour la préparation du survol de JUICE ainsi que des collaborations pour SMILE. L'utilisation de 3dView est en augmentation. La priorité est maintenant la correction des bugs comme le lancement de l'outil sur MAC. Il faut voir avec AR le web-service de distribution des trajectoires et faire un effort sur la documentation ses interfaces.

BC encourage à avoir un doi par version (faisable sur zenodo) ce qui permet de citer la version qu'on utilise.

IP présente la nouvelle version du Propagation Tool qui a un nombre impressionnant de fonctionnalités. **Le CU encourage à mettre en valeur le tutoriel sur la page de lancement.**

Le CU rappelle que ces deux outils connaissent des soucis d'installation sur les plateformes non-windows et que certains administrateurs réseaux n'aiment pas trop les applis développées en Java. Le CU trouve que le nombre de fonctionnalités de ces outils est largement suffisant pour se concentrer sur leur stabilité et la résolution des bugs.

16:00-16:30 Discussion d'une évolution future possible de 3Dview vers un 3Dview-web (comme Aladin et Aladin-lite pour ceux qui connaissent)

Le CDPP réfléchit à une version allégée web de 3DView. BC fait un rappel sur l'outil Aladin du CDS qui a maintenant une version allégée pour avoir rapidement des images (Aladin lite). Un des problèmes de 3DView est qu'il ne permet pas d'avoir la taille réelle des corps célestes (exemple d'une occultation de Jupiter par Ganymède vue depuis Galileo).

Le développement d'une version légère se fait nécessairement dans un autre langage.

Le CU soutient le développement d'une version allégée et comprend la démarche apparemment paradoxale de faire une version web de 3dView et une version desktop de l'interface graphique d'AMDA (SciQLOP). Dans les deux cas, cela va vers une plus grande utilisation (rapidité de navigation et implication de la communauté pour AMDA-desktop ; rapidité également et facilité de lancement pour 3dView). La priorité doit aller à l'optimisation des outils matures existants (AMDA Web, 3DView, PropTool), au développement de SciQLOP (=AMDA Desktop) et à l'implémentation d'un 3D-View Web.

16:30-17:00 Infrastructure CDPP

Les contrats archivage maintenance sont renouvelés. Le contrat industriel de support aux missions (AMDA) est en cours de négociation. Le CNES est en attente d'un recrutement interne. Il faut prévoir l'application de la nouvelle convention, et les départs prochains de MB à la retraite et de NAN de la direction scientifique du CDPP.

Il y a un besoin de mise à jour de l'outillage (remplacement de redmine). La question se pose de l'externalisation de l'hébergement des outils et outillage.

La discussion porte aussi sur la transformation du CDPP en une infrastructure de recherche avec des moyens stables de développement et de maintenance. Cela impliquerait d'absorber des services annexes (HAPI, MASER, STORM). BC souligne que ce regroupement sans devenir une infrastructure de recherche serait compliqué.

Le CU encourage vivement le CDPP à anticiper la succession de NAN à la direction scientifique. Le CU souhaite que le CDPP obtienne des moyens humains pérennes qui sont requis pour de tels projets (voir par exemple le CDS).

17:00+ Fonctionnement du CU

Le nombre de membres du CU présent en fin de réunion ne permet pas une discussion complète sur son fonctionnement. Cependant **le CU du jour qui a fait une grande place aux discussions a été positivement apprécié**. Un bémol venant de la date choisie qui n'a pas permis à la majorité des membres du CU de se déplacer. **Le CU souhaite que la priorité soit donnée à la disponibilité des membres pour prochain CU dont la date devrait être fixée plus en avance afin de rendre la participation "virtuelle" largement minoritaire**

Le CU ne demande pas de renouvellement cette année (il y a eu deux nouveaux entrants l'an dernier). Les questions du remplacement de NAU (double casquette CDPP/Utilisateur) et du président se poseront lors du prochain CU.

Le prochain CU se tiendra autour d'une session de tests avant la fin de l'année.