

CENTRE DE DONNÉES IPGP

Compte rendu du Comité de Pilotage - Salle des conseils (108)

5 février 2013

Participants

Membres du comité : Nikolai SHAPIRO, Constanza PARDO, Arnaud CHULLIAT, Eléonore STUTZMANN, Pierre AGRINIER, François BEAUDUCEL¹, Arnaud LEMARCHAND (excusé), Jean-Pierre VILOTTE, Wayne CRAWFORD, Philippe LOGNONNÉ, Jean-Christophe KOMOROWSKI, Céline DESSERT

Membres du centre (invités) : Sylvie BARBIER, Vincent DOUET, Cyril GUINET, Ludmila KLENOV

Autres invités : Virginie MAURY (Observatoire Magnétique), Romuald DANIEL (Géosciences Marines), Olivier DEWEE (DT INSU)

Programme de réunion

9h30 Présentation générale du centre de données de l'IPGP et de ses missions (N. SHAPIRO)

10h00 Présentation du centre de données de l'IPGP et des projets en cours (C. PARDO)

11h00 Présentation de la gestion des données des OVS / VOLOBSIS (C. PARDO, A. LEMARCHAND, C. GUINET, F. BEAUDUCEL)

14h00 Présentation du projet INSIGHT (sismomètre sur Mars) et de données futures associées (P. LOGNONNÉ)

14h15 Présentation de la gestion des données de Obs. Magnétiques (V. MAURY, A. CHULLIAT)

14h30 Présentation des données géochimiques OVS et OBSERA (C. DESSERT)

14h45 Présentation des données géologiques / SIG (J.C. KOMOROWSKI)

15h00 Présentation des données de campagnes OBS (R. DANIEL, O. DEWEE)

15h15 Discussion sur la salle des observatoires ⇒ non abordé

15h45 Discussion sur les priorités pour l'année à venir ⇒ non abordé

Lexique et liens

CdD Centre de Données IPGP

Geoscope Observatoire Geoscope

OVS Observatoires Volcanologiques et Sismologiques IPGP

ObsErA Observatoire de l'Érosion aux Antilles

1. prise de notes et rédaction du présent compte-rendu

Volobsis	Portail de distribution de données des OVS
INSU	Institut National des Sciences de l'Univers - CNRS
SNO	Service National d'Observation INSU
SIG	Systèmes d'Information Géographique
OBS	Ocean Bottom Seismometers

La séance a débuté à 9h30.

1 Introduction [N. SHAPIRO]

- Présentation du CdD... Il s'agit exclusivement de données numériques.
- UMS IPGP créée en 2011 : l'équipe du CdD en fait partie, comme tous les ITA des observatoires et services de l'IPGP.
- l'INSU a récemment reconnu l'effort de l'IPGP en matière de gestion et distribution de ses données.
- organigramme du CdD : Geoscope et observatoires magnétiques ont une politique de gestion/distribution de données depuis leur création ; les observatoires magnétiques fonctionnent bien et il n'y a pas de raison de modifier/réformer dans l'immédiat.
- Geoscope & OVS = en cours de fusion essentiellement sur base des données sismologiques (en raison du format commun)
- présentation des objectifs de la salle observatoire → sera abordé en fin de journée
- site web du CdD : présentation et liens vers les portails existants
- présentation de l'équipe du CdD
- présentation du comité de pilotage : interne à l'IPGP, son rôle est assez informel, ses membres sont nommés par le directeur du CdD, il n'y a pas de membres externes, son rôle est essentiellement d'assister l'équipe du CdD dans son travail et faire des propositions sur les projets en cours et futurs, la stratégie à long terme (5 ans)
- [E. STUTZMANN] = il y a aussi un besoin de définir une stratégie à court terme (1 an)
- [N. SHAPIRO] = ce n'est pas nécessaire car chaque observatoire a déjà son propre conseil et sa propre gestion...
- [P. LOGNONNÉ] = quels sont les sources de crédits du CdD ?
- [N. SHAPIRO] = il y a un budget de base IPGP de 42 k€+ des financements du Labex
- [J.C. KOMOROWSKI] = propose de changer l'intitulé "SIG" en "données géologiques" pour être cohérent avec les autres types de données ⇒ **Le comité a approuvé**
- [A. CHULLIAT] = comment s'articule le CdD par rapport aux SNO de l'INSU, au niveau financements notamment ?
- [N. SHAPIRO] = le CdD ne reçoit pas d'argent de l'INSU...

Tour de table.

2 Présentation technique du CdD [C. PARDO]

- missions = gérer les données et fournir leur accès à la communauté → assurer la pérennité des données et l'intégration aux autres structures (RESIF, FOSFORE, EIDA, FDSN, ...)
- historique = création en septembre 2011 sur la base de Geoscope (expertise en SEED, gestion, validation, distribution de données), de ses matériels et de ses personnels techniques
- présentation de l'équipe actuelle, des départs précédents et de la campagne NOEMI 2013, contribution du SI (soutien au stockage), total 2013 = 4.64 ETP
- données = Geoscope / OVS / autres réseaux : Chili (CX), fond de mer (Y8), mission Apollo (XA), noeud EIDA
- volumes de données archivées = essentiellement Geoscope, mais les autres réseaux commencent à remplir la base
- interfaces d'accès : portails, seedlink, NetDC, Arclink, webservice
- présentation de l'architecture du système d'information
- contribution du Labex WP V1 : valorisation, transfert et expertise
- [P. LOGNONNÉ] = l'IPGP apporte une contribution financière au centre François Arago (100 à 150 k€/an) pour la gestion de données
- présentation de RESIF : projet instrumental et SI → l'IPGP participe par le noeud A Geoscope, le noeud A OVS et le développement du portail d'accès. Plan sur 2 ans.
- [E. STUTZMANN] : avec RESIF on se dirige vers un unique serveur Arclink RESIF/Grenoble... l'IPGP ne sera plus du tout visible. Il faut absolument prévoir que l'IPGP distribue ses données en parallèle
- [N. SHAPIRO] : l'IPGP s'est toujours engagé à distribuer ses propres données. La direction soutient fortement le CdD (exemple : demande de poste CdD mis en priorité au niveau de l'IPGP).
- [J.P. VILOTTE] : l'IPGP a des responsabilités d'observatoires opérationnels et doit impérativement distribuer ses données de façon opérationnelle, indépendamment de centres comme RESIF.
- [P. LOGNONNÉ] : Geoscope a été fondateur dans la sismologie globale et l'IPGP en a largement bénéficié.
- [N. SHAPIRO] : la direction de l'IPGP a toujours maintenu et affiché sa volonté de poursuivre la distribution de ses données, il n'y a aucune ambiguïté là dessus.
- [A. CHULLIAT] : où se situe la limite entre la validation et la distribution entre les noeuds A et le CdD ?
- [N. SHAPIRO] : le CdD se limite à une validation "technique" car il n'y a pas de personnels CNAP au CdD. Il ne modifie jamais les données sans autorisation de l'observatoire qui en est la source. La validation scientifique doit être faite par les observatoires via les tâches de service de leur personnel CNAP.
- [R. DANIEL] : les utilisateurs sont-ils informés de changements/correction de données ?
- [N. SHAPIRO] : généralement non. Il n'y a pas d'identification des utilisateurs pour Geoscope, RESIF ou VOLOBSIS
- [P. LOGNONNÉ] : projet de stockage sécurisé de données prévu et financé (70 k€) pour 2014 → il faut se coordonner avec le CdD ⇒ **Le comité souligne l'importance de coordination**

- demandes effectuées au CdD pour une intégration de données : campagnes OBS (R. DANIEL), sécurisation données SWARM (O. SIROL), données acoustiques Yasur (S. VERGNOLLE)
- [N. SHAPIRO] : discussion nécessaire sur une possible mutualisation des services avec les observatoires magnétiques (éléments communs des deux SI, suggestion d'intégrer le format "Station XML" pour les métadonnées, et...). Il faut également une réflexion sur l'intégration des données géologiques (implications sur les formats, etc...)

3 Présentation du portail Volobsis [C. GUINET]

- objectifs : accès sans restriction aux données numériques des OVS, accès avec restriction aux données temps-réel
- historique : cahier des charges et appel d'offre en 2006, travail coordonné à l'IPGP mais réalisé par une société externe
- données concernées : essentiellement des données numériques instrumentales mais aussi certaines données manuelles issues des bases de données Webobs (exemple : géochimie, extensomètre)
- [J.P. VILOTTE] : un format "standard" est par définition utilisé par une communauté. Distribuer des données géochimiques au format SEED présente le danger de ne pas correspondre aux attentes des utilisateurs.
- [C. GUINET] : dans Volobsis le SEED est surtout un format de stockage permettant l'homogénéité de la base de données. Des outils sont mis à disposition des utilisateurs pour transformer le SEED en format texte.
- [F. BEAUDUCCEL] : la mise au format SEED des données non sismologiques nous a aussi permis de rationaliser voire compléter / mettre à jour les métadonnées parfois très complexes pour ce type de données.
- [A. CHULLIAT & C. DESSERT] : il faut une discussion sur les formats de données à adopter...
⇒ **Le comité propose que ces discussions commencent dès maintenant**
- données actuellement distribuées : catalogue sismique, GPS continu, flux sismologique continu et bientôt données "lentes" ($< 1Hz$)
- [N. SHAPIRO] : l'IPGP et les OVS ont décidé de distribuer les données en temps-réel, sans délai autre que technique, y compris pour les stations des observatoires volcanologiques. Il y a cependant la possibilité de fermer l'accès en cas de crise.
- [J.C. KOMOROWSKI] : la fermeture soudaine d'accès aux données au moment d'un crise est un message très négatif et qui peut compliquer une gestion de crise... Il faut que notre communauté se prépare à tout faire pour ne jamais avoir à fermer un accès en temps de crise, il y a là une réflexion à avoir. ⇒ **Le comité note qu'il s'agit du problème des OVS et non du CdD**
- présentation du schéma d'archives/traitement/validation des données
- perspectives : format de distribution du catalogue, évolution du portail à 2 ans, intégration EIDA et IRIS, échange métadonnées entre OVS et CdD (via Arclink)
- [N. SHAPIRO] : en complément de la distribution des données au public, il y a aussi des efforts de mise à disposition de données en interne (post-doc, thèse, chercheurs, ...). Des étudiants en bénéficient déjà.
- [P. LOGNONNÉ] : quid des données de type spatiale (imagerie notamment) au sein des OVS?
[NS] : si le besoin est exprimé, le CdD peut y réfléchir mais le CdD ne définit pas les politiques scientifiques des observatoires.

4 Présentation projet spatial InSight [P. LOGNONNÉ]

- la station sismologique sera équivalente à une station Geoscope supplémentaire.
- personnel prévu : 4 ETP = 1 chaire CNES/IPGP (2014), 1 postdoc (2016), 1 IE analyse, 1 IE informatique, 1 qualiticien (0.3 ETP), 1 documentaliste (0.5 ETP)

5 Présentation observatoires magnétiques [V. MAURY]

- flux de données : réseau BCMT (16 observatoires = 11 IPGP + 5 EOST) → serveur Magis → www.bcmt.fr + serveur GIN Intermagnet (réseau international 133 obs.) → www.intermagnet.org (26 stations au total)
- modes de transmission : upload curl, Eathworm, ftp → temps-réel (2-3 secondes de délai)
- type de données = XYZ, scalaire (F), absolu, autres paramètres (température, batterie, ...), échantillonnage = @1Hz (10 obs) ou 1-minute (1 obs), format = IAGA-2002, volume = 6 Mo/jour (@1Hz), 200 octets/s (temps-réel), 100 Ko/j (@1mn)
- plateforme Magis : traitement des données = vérification, ré-échantillonnage, archivage, graphes, outil de "despike", calibration, lignes de base, détection d'orages
- présentation des équipements informatiques
- [N. SHAPIRO] : quelle est la politique IPGP en matière de stockage pérenne de ces données ?
- [A. CHULLIAT] : nous sommes essentiellement mutualisés avec les solutions de stockage/sauvegarde de l'équipe géomagnétisme.
- [N. SHAPIRO] il faudrait connaître les besoins éventuels de la mission Swarm en matière de gestion de données ⇒ **A. CHULLIAT doit contacter G. HULOT à ce sujet**

6 Présentation des données géochimiques [C. DESSERT]

- données géochimiques = OVS (Guadeloupe, Martinique, Réunion) + ObsErA (Guadeloupe) → données multi-observatoires et multi-paramètres
- [N. SHAPIRO] : les données de chimie OVSM sont-elles intégrées à Volobsis ? Le CdD a besoin de référents scientifiques pour chaque discipline qui suive typiquement ce type de problème.
- [C. DESSERT] : il faut définir et stocker les métadonnées liées à toutes les chaînes d'analyse... Il faut une discussion avec le CdD sur ce sujet ⇒ **Le comité a approuvé**
- OVPF = mini-DOAS, maxi-DOAS, etc... ⇒ **ces données devraient être intégrées, à voir.**
- OVSG = sources + fumerolles + boîtes japonaises. Probablement la base de données géochimique la plus complète sur un volcan actif.
- problème de stockage des échantillons (il en resterait encore à P6)... ⇒ **à voir rapidement**
- distribution des données : site web ObsEra <http://morpho.ipgp.fr/Obsera>, Webobs <http://www.ovsg.univ-ag.fr> et <http://webobs.ovmp.martinique.univ-ag.fr/>
- site web ObsErA développé et maintenu par un étudiant en thèse (H. CHAUVET) ⇒ **penser sérieusement à la pérennisation de la maintenance ! Possibilité de transférer vers un système type maintenu par l'IPGP type Webobs ou Volobsis ?**
- [F. BEAUDUCCEL] : faut-il stocker les données instrumentales brutes, résultats des mesures au chromatographe, spectrographe qui produisent des données intermédiaires ?

- [C. DESSERT] : a priori ce n'est pas utile car on conserve les échantillons et on peut relancer les analyses. D'autre part ces mesures sont réalisées avec de forts a priori sur l'élément chimique recherché...
- [P. AGRINIER] : il faut en effet y réfléchir. Les échantillons de gaz par exemple ne sont pas conservables...

7 Présentation des données géologiques [J.C. KOMOROWSKI]

- objectifs : données géospatialisées "5D" = 3D + temps + paramètre. Format/logiciel ESRI (payant)
- type de données : orthophotos, MNT, surfaciques, linéaires, ponctuelles, calculées/interpolées, géographiques, météorologiques, géologiques, histoire éruptive, ...
- exemples : datations C14 à la Guadeloupe, carte de cônes volcaniques au PdF, ... toutes ces données sont indispensables aux recherches et à la gestion de crise
- possibilité d'intégration de données de structures (ex. tomographies), données de terrain diverses, informations sur la tectonique, données de volcanologie marine ...
- données évoluées sur les aléas et les risques : faut-il les distribuer ?
- [N. SHAPIRO] : non ce n'est pas notre rôle ; l'IPGP est chargé de la surveillance mais pas de produire des cartes de risques qui sont des documents de base très officiels. C'est plutôt à des organismes comme le BRGM de fournir ces informations. L'IPGP doit se limiter à une utilisation dans le cadre de la recherche.
- [N. SHAPIRO] : il n'y a pas d'expertise GIS à l'IPGP, mais il faudrait y penser... les 4 ou 5 chercheurs CNAP étiquetés "géologie" devraient pouvoir contribuer à cela... si l'on veut intégrer ces bases de données dans le CdD, il faut aussi commencer par les documenter.
- [F. BEAUDUCCEL] : pour information il existe le logiciel open-source QGIS : il faudrait voir le niveau de compatibilité avec ESRI pour faciliter l'exploitation des bases GIS existantes. Au sujet de l'intégration de données de tomographie, le SIG ne me semble pas le bon outil pour visualiser/exploiter des informations réellement 3D... c'est en revanche parfaitement adapté aux données en couches 2D. Sinon je suggère de commencer par intégrer/distribuer les différents MNT disponibles qui sont extrêmement utiles à tous et pour l'instant distribués de façon trop individuelle et au coup par coup.
- [L. KLENOV] : suggère le format NetCDF pour stocker/distribuer les données 3D.
- [N. SHAPIRO] : les données de tomographie sont trop évoluées pour être distribuées. On pourrait effectivement commencer par distribuer les MNT.

8 Présentation des données marines : logiciel RADO [R. DANIEL & O. DEWEE]

- origine du parc : S. SINGH en 2000. Suite aux récents départs, actuellement le personnel se réduit à une seule personne...
- [W. CRAWFORD] : cette situation est purement temporaire.
- présentation du parc d'instruments, des missions, du stockage = HD ou DVD à Saint-Maur (volume total = 1 To)
- développement d'outils pour la conversion des données en SEED

- [N. SHAPIRO] : il faut faire rapidement une copie des données brutes sur les espaces de stockage à Cuvier.
- [P. LOGNONNÉ] : quelle est la valeur ajoutée du CdD si les PI de projets fournissent déjà des données SEED validées ? Le CdD ne devrait-il pas participer/contribuer aussi à la mise en forme des données ?
- [W. CRAWFORD] : les PI de projets ont besoin d'aide et d'expertise sur la mise en forme de leurs données à des fins de distribution : quels nomination de canaux, quelles informations mettre dans les dataless, ...
- [N. SHAPIRO] : dans ce cas il faut plus de postes... par ailleurs le CdD contribue déjà sur cet aspect pour Geoscope et les OVS...
- [W. CRAWFORD] comment accéder aux données des campagnes précédentes, type Sismantilles ? Le comité de pilotage devrait explicitement demander que les données de campagnes financées par des programmes de recherche terminés soient rendues publiques et accessibles.
- ⇒ Le comité insiste sur le fait que les données des campagnes Sismantilles ont été financées par des fonds publics et devraient être disponibles ouvertement. De plus, les PIs de ces campagnes ont déjà publié leurs principaux résultats et ces campagnes ont été coordonnées et portées par l'IPGP. Pour ces raisons, le comité suggère qu'au moins une copie des données de ces campagnes soit livrée à l'IPGP afin que les membres de notre institut et des observatoires puissent les utiliser dans leurs activités de recherche et d'observation. Le comité suggère que la direction de l'IPGP contacte les responsables de ces campagnes à ce sujet.

La séance est levée à 18h. Les deux derniers points de l'ordre du jour sont reportés à une prochaine réunion.