

RAPPORT D'ACTIVITÉ

CALCUL NUMÉRIQUE INTENSIF | ARCHIVAGE PÉRENNE | HÉBERGEMENT

2015-2017

SOMMAIRE



› Éditorial	6
› Présentation du Cines	7
› Les missions nationales	7
› Organisation	8
› Instances	8
› Orientations stratégiques	9



› Infrastructures de niveau Tier3+	10
› Nouvelle salle machine de 600 m ² dédiée au calcul	11
› Ressources énergétiques	11
› Réseaux et Sécurité	13
› Évolution des infrastructures entre 2008 et 2017	15



› Calcul de haute performance	16
› De JADE à OCCIGEN	17
› Stockage, architecture "data centrique"	19
› Logiciels disponibles	20
› Utilisation des ressources	20
› Collaborations	21
› Veille technologique	22
› Services	22
› Comité des utilisateurs	23



➤ Archivage 24

- Contexte historique 25
- Équipe d'experts 25
- Refonte de la plateforme PAC 26
- Avancées en matière d'expertise sur les formats 27
- Projet d'agrément HADS (données de santé) 28
- Accroissement des archives, partenaires et typologies 28
- Reconnaissance nationale et internationale 34
- Focus sur quelques projets stratégiques 37



➤ Hébergement 40

- Une mission à part entière 41
- Offre de haut niveau 41
- Hébergés nationaux et locaux 41
- Projet de data center 41

➤ Services informatiques transverses 42

- Serveurs 42
- Capacité de stockage 42
- Sauvegardes 42
- Réseaux 43
- Métrologie 43
- Sécurité 43
- Informatique et libertés 43
- Supervision et astreintes 43



› Ressources humaines 44

- › Postes et effectifs 45
- › Promotions 45
- › Hygiène et sécurité 45
- › Mutualisation de la gestion des pensions (Adhésion à PETREL) ... 45



› Gestion financière et administration 47

- › Missions et déplacements des agents 47
- › Finances 47
- › Dématérialisation 48

› Formations 49

- › Formations organisées par le CINES 50
- › Formations continues suivies par le personnel CINES 50
- › Adhésion aux formations « communes » 50



› Archives du CINES 51

La période 2015-2017 coïncide avec des événements marquants pour le CINES. En premier lieu le renouvellement du supercalculateur. Après des années de bons et loyaux services JADE a pris une retraite bien méritée pour faire place en janvier 2015 à OCCIGEN, un calculateur conçu par Bull à base de processeurs Intel et acquis par GENCI suite à un dialogue compétitif. Cette nouvelle machine a offert dans un premier temps à la communauté scientifique 2,1 Pflops de puissance crête avant de bénéficier d'une extension portant cette puissance à 3,5 Pflops en 2017. Parallèlement l'installation de deux prototypes comportant des processeurs Intel KNL (de type Xeon Phi) a permis de réaliser une veille technologique sur les possibles évolutions des calculateurs et des techniques de programmation. Le Centre a poursuivi sa participation aux différentes phases d'implémentation de PRACE et en particulier aux formations effectuées dans le cadre du PATC (Prace Advanced Training Center).

Cette période a vu le développement d'un important projet concernant la conservation des herbiers numérisés, mené en partenariat avec le Museum National d'Histoire Naturelle. Ce projet combine archivage et calcul, en s'appuyant d'un côté sur les outils développés dans le cadre du projet européen EUDAT2020, interfacés avec la plate-forme d'archivage du CINES, et d'un autre côté sur un traitement d'OCR délégué au calculateur. Un travail stratégique a également été initialisé, avec l'appui du Comité d'orientations stratégiques du centre, sur les données environnementales pour définir une démarche générique pour l'archivage des données scientifiques. Par ailleurs le CINES a signé un contrat de partenariat avec le projet interministériel VITAM de développement d'un logiciel d'archivage numérique, socle commun pour les administrations. Ce programme lancé en 2015 est piloté conjointement par le Comité interministériel aux Archives de France (CIAF) et la Direction interministérielle du numérique et du système d'information et de communication de l'État (DINSIC).

Par arrêté du 6 novembre 2015 le CINES est devenu officiellement une ZRR (Zone à Régime Restrictif), c'est-à-dire un site et des accès à distance hautement sécurisés. Bien que nombre des éléments sécuritaires aient déjà été opérationnels, certaines améliorations ont été apportées. Des investissements ont été réalisés, notamment la généralisation de la vidéosurveillance dans le cadre de la sécurisation du bâtiment et la construction d'une zone publique afin de permettre au CINES d'assurer les formations et l'accueil de personnes extérieures. Concernant les accès virtuels, à la suite d'un audit effectué par l'ANSSI l'environnement du supercalculateur OCCIGEN a été homologué pour la sécurité. Enfin un système d'extinction incendie a été déployé en 2017 dans la salle machine hébergeant OCCIGEN et les données associées, et se poursuivra dans les autres salles machines.



Grâce à l'ensemble de ses infrastructures hautement sécurisées, basées sur une architecture en miroir, le CINES offre aujourd'hui aux partenaires hébergés un service de niveau Tier 3+, reconnu par Infranum. Chaque hébergé bénéficie d'une mutualisation des infrastructures et d'une optimisation des consommations énergétiques reposant sur une politique d'urbanisation efficace des salles machines, confinement des couloirs froids notamment, et sur un suivi des mesures de consommation électrique individualisées.

Dix ans à la direction du CINES, un sacré bail ! Mon mandat a pris fin le 30 septembre 2017. Je profite de cette dernière occasion pour adresser mes sincères remerciements aux personnels du CINES, à son Secrétaire Général, en n'oubliant pas ceux qui vivent aujourd'hui une retraite bien méritée. Sans leur collaboration rien n'aurait été possible. Mes remerciements également aux Directeurs qui m'ont précédé : Jean-Claude Ippolito, Jean-Loïc Delhay, Philippe Rouzard, Alain Quéré, Thierry Porcher. Ils ont légué un instrument de premier plan rendant possible la tâche accomplie pendant mon mandat. Je ne voudrais pas oublier les instances du CINES, notamment les membres du Conseil d'Administration qui tout au long de ces dix années ont soutenu le Centre et les différents projets proposés. Enfin merci aux nombreux partenaires qui ont accompagné le Centre, au comité

des utilisateurs pour ses suggestions pertinentes et à l'ensemble des utilisateurs pour leur confiance et leur fidélité.

Bon vent au CINES et mes vœux de succès dans ses diverses missions. ■



PRÉSENTATION DU CINES

Le **C.I.N.E.S.** (Centre Informatique National de l'Enseignement Supérieur) est un Etablissement Public à caractère Administratif national (EPA), basé à Montpellier, et placé sous la tutelle du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI). Le CINES propose des moyens exceptionnels, en particulier pour le calcul de haute performance, à l'ensemble de la communauté scientifique dans le cadre de ses missions statutaires stratégiques nationales.

LES MISSIONS NATIONALES

Le Décret no 2014-303 du 6 mars 2014 modifie les statuts du CINES et officialise les nouvelles missions du Centre :

- **Le calcul numérique intensif**, soit par les actions que le centre conduit directement soit par les conventions qu'il passe avec d'autres organismes publics ou privés, nationaux ou étrangers, en particulier dans le cadre des accords passés entre le centre et la société GENCI,
- **L'archivage pérenne** de données électroniques afin de contribuer à la préservation du patrimoine scientifique numérique national ;
- **L'hébergement** de matériels informatiques à vocation nationale dans la mesure de la disponibilité de locaux et de capacités techniques, électriques et de climatisation, ne compromettant pas l'exécution et l'évolution des deux missions précédentes.



ORGANISATION

Le centre est divisé en quatre départements sous la responsabilité d'un Directeur assisté d'un Secrétaire Général. Le Directeur du centre est nommé par arrêté du ministre chargé de l'enseignement supérieur et de la recherche pour une durée de trois ans renouvelable.

Département Calcul Intensif (DCI) :

Le DCI est en charge des environnements de calcul hébergés au CINES, du support, de l'aide aux utilisateurs pour la parallélisation et l'optimisation des codes, et de la promotion des techniques de calcul parallèle. Il fournit l'expertise informatique aux nombreuses disciplines scientifiques qui utilisent les supercalculateurs du Centre pour la résolution de problèmes qui exigent des puissances de calcul extrêmes associées à de grandes capacités mémoire et des grandes masses de données.

Département Archivage et Diffusion (DAD) :

Le DAD définit et met en place les outils et les procédures qui permettent de garantir la pérennité des données et documents numériques qui sont confiés au CINES pour un archivage à très long terme agréé par le SIAF (service interministériel des archives de France). Il apporte conseil et expertise dans le domaine et participe à des projets nationaux et internationaux.

Département Services Informatiques et Infrastructures (DS2I) :

Le DS2I est en charge des infrastructures informatiques communes sur lesquelles s'appuient les équipements des autres départements. Il veille à la disponibilité des ressources 24 heures sur 24 et à la sécurité des données. Il assure également le centre d'appel de premier niveau.

Département Administration et Ressources Humaines (DARH) :

Le DARH est chargé de la gestion financière et administrative du centre. Il prépare les marchés avec les différents fournisseurs et veille à leur exécution. Il assure par ailleurs la gestion des ressources humaines du CINES.

INSTANCES

Statutairement le CINES est administré par un conseil d'administration et est doté d'un comité d'orientations stratégiques.

Le Conseil d'Administration du CINES comprend :

- 1° Quatre membres de droit :
 - a) Le directeur chargé de la recherche au ministère chargé de l'enseignement supérieur et de la recherche ou son représentant ;
 - b) Le directeur chargé de l'enseignement supérieur au ministère chargé de l'enseignement supérieur ou son représentant ;
 - c) Le responsable du service chargé de l'information scientifique et technique au ministère chargé de l'enseignement supérieur ou son représentant ;
 - d) Le directeur général du Centre national de la recherche scientifique ou son représentant.

- 2° Cinq personnalités nommées par arrêté du ministre chargé de l'enseignement supérieur et de la recherche en raison de leurs compétences dans les domaines définis à l'article 2, dont deux présidents d'université désignés sur proposition de la conférence des présidents d'université ;

- 3° Trois représentants élus des personnels.
Le président du conseil d'administration est nommé, par arrêté du ministre chargé de l'enseignement supérieur et de la recherche, parmi les cinq personnalités nommées.

Le Comité d'Orientations Stratégiques comprend :

- 1° Trois membres de droit :
 - a) le directeur chargé de la recherche au ministère chargé de la recherche ou son représentant ;
 - b) le directeur chargé de l'enseignement supérieur au ministère chargé de l'enseignement supérieur ou son représentant ;
 - c) le Président du conseil d'administration du centre.

- 2° Trois personnalités nommées par arrêté du ministre chargé de l'enseignement supérieur en raison de leurs compétences dans les domaines définis à l'article 2, pour une durée de trois ans renouvelable.

Le directeur du centre assiste avec voix consultative aux séances du comité d'orientations stratégiques. Le comité d'orientations stratégiques élit, parmi ses membres, un président pour une durée de trois ans renouvelable, dans les conditions fixées par le règlement intérieur de l'établissement.

Le Comité Technique et le Comité d'Hygiène, de Sécurité et des Conditions de Travail

Le **Comité Technique** du CINES comprend deux représentants de l'Administration, (le Directeur et le Responsable en charge des ressources humaines) et trois représentants titulaires du Personnel (ainsi que trois suppléants). Les suppléants peuvent assister aux séances du comité sans pouvoir prendre part aux débats. Les élections des représentants du personnel au CT CINES ont eu lieu le 4 décembre 2014.

Il est présidé par le Directeur du CINES.

Le **Comité d'Hygiène, de Sécurité et des Conditions de Travail** du CINES comprend deux représentants de l'Administration, (le Directeur et le Responsable en charge des ressources humaines) et trois représentants titulaires du Personnel (ainsi que trois suppléants qui peuvent venir à la réunion mais qui ne peuvent pas intervenir), l'Agent de Prévention du CINES, le Médecin du travail, l'Inspecteur Hygiène et Sécurité et occasionnellement, toute personne qualifiée de l'établissement désignée par le comité.

Il est présidé par le Directeur du CINES.

Un audit des inspecteurs hygiène et sécurité a eu lieu les 9 et 10 octobre 2014 concluant à « une évolution positive de la prise en compte des questions de santé et sécurité au travail par le centre ».

ORIENTATIONS STRATÉGIQUES

Conformément aux nouveaux statuts le comité d'orientations stratégiques a été mis en place en 2016. Il a pour mission de conseiller la gouvernance du CINES.

« Art. 16. – Le comité d'orientations stratégiques propose au conseil d'administration et au directeur les orientations stratégiques pour le centre.

Il élabore le plan stratégique pluriannuel en conformité :

- 1- Au titre du calcul numérique intensif, avec les orientations nationales du comité stratégique pour le calcul intensif et les plans stratégiques de la société GENCI ;
- 2 - Au titre de l'archivage pérenne de données électroniques, avec les préconisations du comité Interministériel aux Archives de France. »

Il résulte de ses travaux et des stratégies fixées par la tutelle, un certain nombre d'axes majeurs pour le centre, au moins pour la période 2016 - 2018 :

Pour le calcul de haute performance :

- une orientation vers des machines généralistes, combinées à des ressources de pré-post traitement et à l'accueil de prototypes représentatifs des technologies à venir,
- une synergie avec les ressources, notamment humaines, des autres centres nationaux, en essayant de mutualiser des compétences,
- une optimisation des dépenses énergétiques en améliorant le PUE et le TCO,
- la préparation des codes et algorithmes pour les technologies des machines du futur (exaflopiques manycore, ARM, machines quantiques, ...),
- la gestion et le traitement des grandes masses de données,
- la participation au projet européen PRACE.

Pour l'archivage :

- poursuite de la politique d'agrément (SIAF, DSA) et extension aux données de santé sur support numérique,
- validation d'une démarche générique robuste pour l'archivage des données scientifiques, première étape (2017-2018) sur les données environnementales,
- participation au projet interministériel VITAM, de constitution d'un outil logiciel commun,
- consolidation des collaborations avec les autres acteurs majeurs de l'archivage nationaux (Archives Nationales, Archives de France, BNF, ...) et européens dans le cadre d'EUDAT et de ses successeurs,
- capacité à archiver des grandes masses de données, positionnement du CINES dans le cycle de vie de ces données.

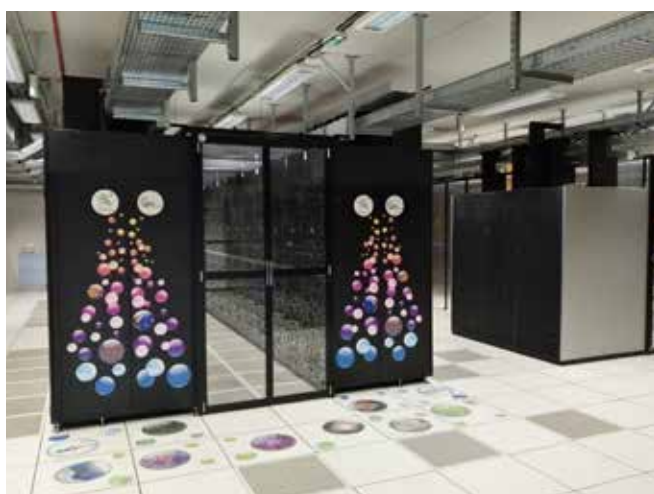
Pour l'hébergement :

- Conformément aux statuts du CINES l'hébergement ne doit pas compromettre l'exécution et l'évolution des deux missions précédentes : calcul et archivage,
- le CINES est opérationnel pour de l'hébergement de plateformes « sensibles » avec des infrastructures de niveau Tier3+. Il en héberge déjà (ABES, DSI INSERM, IRSTEA, ...),
- en tant que centre national il a une mission prioritaire d'hébergement vis-à-vis des plateformes nationales, mais reste ouvert pour de l'hébergement local,
- il peut accueillir des machines de type gestion comme des clusters de calcul, pour cela il dispose de salles machines équipées de moyens de refroidissement par air ou par eau (froide et « chaude »),
- une extension des capacités d'hébergement du centre pourra être étudiée si la demande se révèle supérieure aux capacités actuelles et qu'il n'existe pas d'autres solutions envisageables « rentables »,
- si la demande d'hébergement de données de santé se confirme, un dossier d'agrément devra être soumis aux autorités compétentes.

› Nouvelle salle machine de 600 m ² dédiée au calcul	11
› Ressources énergétiques	11
› Réseaux et Sécurité	13
› Évolution des infrastructures entre 2008 et 2017	15

NOUVELLE SALLE MACHINE DE 600 M² DÉDIÉE AU CALCUL

La construction d'une nouvelle salle machine (SM5) dédiée au calcul a été lancée via un appel d'offres en 2013. Les 600 m² de la SM5 équipée des infrastructures électriques et de climatisation taillées pour accueillir le nouveau supercalculateur, ont été livrés fin 2014. La salle machine bénéficie d'un système de climatisation hybride : 75% eau « chaude » (35°), grâce à l'installation sur le toit d'aéroréfrigérants, pour la partie réservée aux calculateurs, et 25% air (à base d'eau glacée) pour la régulation de la température d'ambiance et l'espace réservé aux baies de disques.



Le découpage de la salle permet de gérer le chevauchement entre les générations n et n+1 des environnements de calcul et réserve un espace pour les tests.

Fin 2017 la salle machine N°5 offre une puissance totale de 1,8 MW. Les infrastructures de climatisation et les baies de disques hébergeant les données étant sécurisées en miroir par une double alimentation vers deux sources ERDF distinctes, elles-mêmes sécurisées par des onduleurs en N+1 et des groupes électrogènes.



RESSOURCES ÉNERGÉTIQUES

Évolution des infrastructures

Après la construction d'un second pôle énergie en 2010 pour accueillir le supercalculateur JADE, le CINES a procédé en 2013 à une amélioration majeure de ses infrastructures énergétiques en doublant ses alimentations électriques auprès d'ERDF, afin de réaliser une architecture en « miroir ». Une deuxième ligne ERDF à 10 MW « dédiée » a été installée et mise en service sur le nouveau pôle énergie. Ainsi toutes les salles machines du CINES disposent aujourd'hui :

- d'un double cheminement électrique de bout en bout pour l'ensemble des matériels informatiques « sensibles » avec deux adductions (ERDF/ENEDIS) séparées et la sécurisation par onduleurs redondés en N+1 et des groupes électrogènes sur chaque chemin.
- d'une climatisation sécurisée et redondante (en N+1)

Les infrastructures du centre de niveau TIER3+ assurent une continuité du service 24h sur 24 et permettent en particulier de réaliser des maintenances à chaud des installations électriques

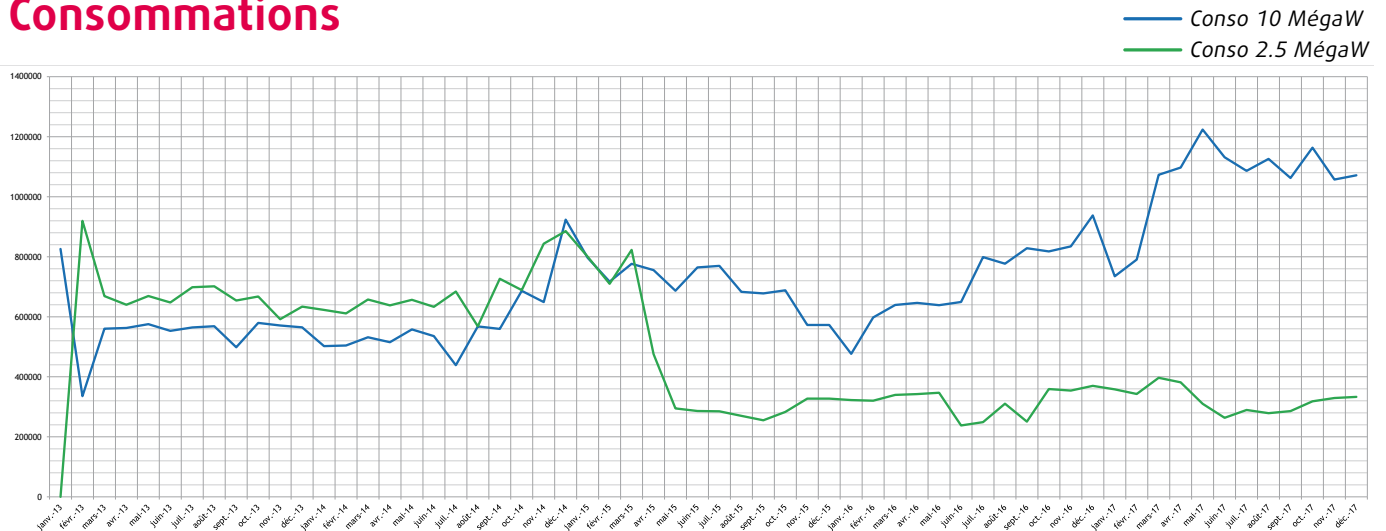
En 2016, suite à une réorganisation des salles machines « historiques » la capacité énergétique de la salle machine n°3 a été augmentée (+210 kW) et les équipements redondés afin d'être conformes au niveau Tier3+ et de dédier cette salle à la mission hébergement, dès la fin d'année 2016. Les machines préalablement hébergées dans cette salle ont été transférées dans la salle machine N°1.

Afin d'urbaniser au mieux la salle machine, le CINES fournit dorénavant les baies sécurisées, par un digicode, et met en place un confinement des allées froides.

Parallèlement l'arrivée de l'extension d'OCCIGEN a nécessité l'augmentation de 1 MW des capacités de la salle machine N° 5. Cela s'est traduit par l'ajout d'onduleurs et de climatiseurs supplémentaires.

En 2015 est entrée en vigueur la déréglementation pour la fourniture d'électricité qui a entraîné une mise en concurrence des fournisseurs et la signature avec ceux-ci d'un contrat cadre et l'attribution de marchés subséquents pour chaque année au fournisseur le moins disant.

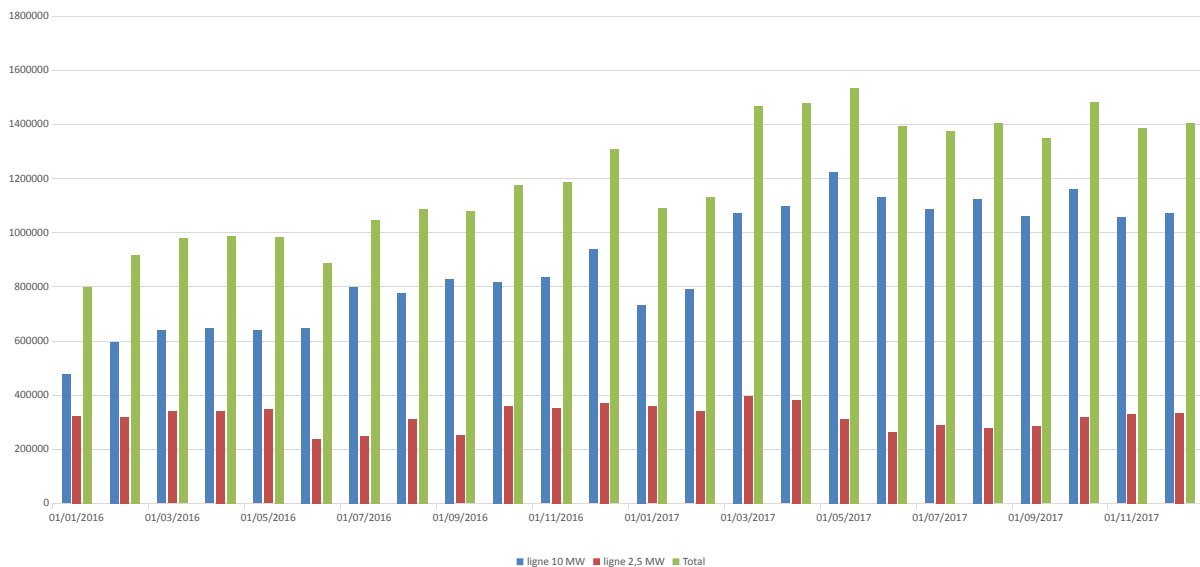
Consommations



L'extension de la configuration d'OCCIGEN à partir de janvier 2017, qui a multiplié la puissance par 1,7 en faisant passer la performance crête de 2,1 Pflops à 3,5

Pflops, a entraîné une augmentation de la consommation électrique, sans impacter celle-ci dans la même proportion.

Consommation électrique mensuelle 2016-2017 en kWh



Optimisation énergétique

Au-delà des aspects performances, une préoccupation majeure du Centre consiste à minimiser son impact environnemental en favorisant au maximum les équipements les moins énergivores pour une puissance donnée et en optimisant l'utilisation des ressources énergétiques importantes, comme le montrent les graphiques précédents, mais nécessaires.

Plusieurs axes sont exploités pour atteindre cet objectif :

- lors de la procédure de sélection des machines le CINES donne un poids important aux critères de consommation (PUE, TCO,...). Ainsi lors du choix du calculateur OCCIGEN, en accord avec la maîtrise d'ouvrage GENCI,

la technologie de refroidissement par eau tiède qui permet d'obtenir un PUE inférieur à 1,1 a été privilégiée afin d'obtenir le meilleur rapport performance sur énergie ;

- une fois les plates-formes choisies l'urbanisation des salles machines vise à réduire la consommation d'énergie nécessaire à la climatisation par un confinement des allées froides des baies informatiques en maîtrisant ainsi les flux d'air pour une meilleure gestion des points chauds ;
- le suivi des consommations électriques des divers équipements, informatiques et d'infrastructure, est ensuite réalisé au moyen de mesures individualisées à l'aide de capteurs et de compteurs électriques certifiés

MID complétés par un logiciel d'analyse. L'objectif est alors de disposer d'une capacité d'optimisation des usages ;

- la production de froid nécessaire pour la climatisation d'ambiance des salles machines étant génératrice de chaleur dissipée celle-ci est en partie récupérée pour assurer le chauffage des bureaux ;
- enfin le CINES est partenaire du projet innovant ICARE, piloté par le LIRMM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier) dont l'objet est un prototype d'alimentation d'un cluster de calcul par de l'énergie renouvelable (panneaux photovoltaïque).



Confinement des allées froides

RÉSEAUX ET SÉCURITÉ

Accessibilité

Le point régional RENATER, le cœur du réseau régional R3LR et le cœur du réseau métropolitain HDMON sont tous trois hébergés au CINES. Les partenaires hébergés ont donc toute facilité à se raccorder à ces trois réseaux par des liaisons locales mises à disposition par le CINES. D'autres opérateurs peuvent également installer des arrivées de fibre optique dans les locaux du CINES après étude préalable, si cela est nécessaire pour certains partenaires hébergés.

Sûreté du site

La mise en sûreté du site a été réalisée en étroite collaboration entre le Fonctionnaire Sécurité Défense et le RSSI du CINES et les services du Haut Fonctionnaire de Défense et de Sécurité (HFDS) du Ministère, avec les conseils éclairés du correspondant local de la Direction Générale de la Sécurité Intérieure (DGSi). Elle vise à protéger les équipements et les personnels du CINES de toute atteinte à leur intégrité et à préserver le patrimoine scientifique de l'Etat.

Passage en ZRR

Par arrêté du 6 novembre 2015 le CINES est officiellement classé en ZRR (Zone à Régime Restrictif). Bien que la plupart des mesures attachées à une ZRR aient déjà été appliquées, ce classement a entraîné certaines modifications à la fois dans les dispositifs de contrôle des accès physiques et dans la gestion des accès virtuels informatiques.

Aujourd'hui l'accès à la ZRR est contrôlé à tous les niveaux. Toutes les arrivées et les déplacements à l'intérieur de la ZRR sont enregistrés. Toute visite de la partie ZRR du CINES nécessite une déclaration préalable des personnes concernées, avec transmission des copies des pièces d'identités.

Construction d'une zone publique

Une des conséquences du classement en ZRR a été la nécessité de construire une zone publique, étanche par rapport à la partie ZRR, qui permette au CINES de perpétuer en toute sécurité l'accueil du public extérieur pour les formations, séminaires, réunions et autres manifestations.

Les plans de définition de la ZRR et de la zone publique, présentés en Comité Technique et en Conseil d'Administration du CINES, ont été validés par le HFSD et la DGSi.



Plan de la zone publique

Vidéo surveillance et alarmes

Le dispositif de sûreté du site a également été renforcé par un système de vidéo surveillance périmétrique avec

alarmes associées pour signaler toute entrée illicite dans le champ des caméras. Le système de badgeage a été rénové et permet de tracer nominativement les accès aux salles machines et aux locaux techniques.



Alarme et extinction incendie

Jusqu'à présent le CINES disposait de détecteurs précoces sans système d'extinction automatique dans ses salles machines. Suite à un appel d'offres le CINES déploie en 2017-2018 un système de détection renforcé associé à une extinction automatique à base de gaz inerte. La première salle équipée est la SM5 qui abrite OCCIGEN et les données du calcul.

ÉVOLUTION DES INFRASTRUCTURES ENTRE 2008 ET 2017

Janvier 2008	Fin 2017
4 salles machines	5 salles machines (ajout de la SM5 en 2014) sécurisées en « miroir »
1 pôle énergie « historique »	2 ^{ème} pôle énergie construit en 2010
1 ligne ERDF (1,2 MW)	2 lignes ERDF : nouvelle à 10 MW ajoutée en 2014 passage à 2,5 MW de l'ancienne en 2010
2 transformateurs	6 transformateurs
2 TGBT	7 TGBT
4 onduleurs	8 onduleurs
3 groupes froids (3x415 kW)	6 Groupes froid (ajout de 3x800 kW) 4 aéroréfrigérants adiabatiques
	3 groupes électrogènes
	Extinction incendie automatique (gaz inerte)
	Zone publique (construite en 2017)
	Clôture du site
	Système de vidéo surveillance



Groupes froids



Transformateur



Groupe électrogène

› De JADE à OCCIGEN	17
› Stockage, architecture "data centrique"	19
› Logiciels disponibles	20
› Utilisation des ressources	20
› Collaborations	21
› Veille technologique	22
› Services	22
› Comité des utilisateurs	23

Le Département Calcul Intensif (DCI) regroupe au sein d'une même entité l'ensemble des forces concourant au développement et à la promotion de la mission historique du CINES : le calcul scientifique. Il est en charge des environnements de calcul hébergés au CINES, de l'aide aux utilisateurs pour la parallélisation et l'optimisation des codes, et de la formation aux techniques de calcul parallèle. Il fournit l'expertise informatique aux nombreuses disciplines scientifiques qui utilisent les supercalculateurs du Centre pour la résolution de problèmes qui exigent des puissances de calcul extrêmes associées à de grandes capacités mémoire.

À la fin de l'année 2017, le DCI regroupe seize ingénieurs – dix titulaires et six contractuels – dont l'activité se décline principalement selon trois axes : administration des systèmes et architectures, support aux utilisateurs, projets et collaborations internationales.

DE JADE À OCCIGEN

Après 6 ans de loyaux services, le supercalculateur JADE installé en septembre 2008, 141^{ème} au TOP500 de l'époque, et étendu en février 2010, est arrêté fin 2014 en ayant fourni 150 millions d'heures CPU en 2013.

Suite à un appel d'offres lancé en octobre 2013 par GENCI, sous forme de dialogue compétitif, un nouveau supercalculateur nommé OCCIGEN, fourni par ATOS/Bull, est livré en septembre 2014 en remplacement de JADE. Installé dans la nouvelle salle machine SM5 disponible depuis janvier 2014 et équipée des infrastructures permettant un refroidissement par eau « chaude », il affiche une puissance crête de 2,1 Pflops. Mis à la disposition de l'ensemble de la communauté scientifique le 12 janvier 2015, après une période réservée au passage de grands défis, il est classé 26^e au TOP500 de juin 2015.

La configuration d'OCCIGEN a été étendue en 2016 pour atteindre 3,5 Pflops qui ont été mis à disposition des utilisateurs en janvier 2017. L'ensemble est refroidi par eau « tiède » (35°C), pour un meilleur PUE, grâce à la technologie DLC (Direct Liquid Cooling), et est connecté à un espace de stockage rapide (Lustre) de 5,1 Po avec un débit cumulé de 106 Go/s.

L'intégration de l'extension d'OCCIGEN a nécessité l'augmentation de puissance des infrastructures de la salle machine par le CINES. La mise en production s'est également accompagnée d'un important changement de système (SCS5) sur l'ensemble de la configuration.

La période de deux mois, réservée à la vérification de service régulier, a donné lieu au passage de grands défis dont la restitution a fait l'objet d'un colloque le 7 décembre 2017 et d'un numéro spécial de la gazette du CINES¹.



Supercalculateur OCCIGEN : 3,5 Pflop/s



	OCCIGEN 1 ^{ère} phase	OCCIGEN 2 ^{ème} phase	Total OCCIGEN
Puissance crête	2,1 Pflops	1,46 Pflops	3,5 Pflops
Racks	34 dont 24 de calcul	14 de calcul	38 de calcul
Nœuds bi-processeurs	2 106 de calcul	1 260 de calcul	3 366 de calcul
Processeurs	Intel Haswell EP 2,6 GHz	Broadwell 2,6 GHz	
Cœurs/proc	12	14	
Nombre de cœurs	50 544	35 280	85 824
Mémoire/cœur	50% à 2,6 Go, 50% à 5,3 Go	2,6 Go	
Réseau interne	IB Mellanox 4xFDR	IB Mellanox 4xFDR	

1 : https://www.cines.fr/wp-content/uploads/2017/11/CINES_GAZETTE_GD2017.pdf

Évolution de la puissance de Calcul

Q1 2017 : 3,5 Pflop/s



OCCIGEN

2106 nœuds Haswell (2x12 cœurs)
1260 nœuds Broadwell (2x14 cœurs)
85 824 cœurs, 64 ou 128 Go/nœud,
IB 4xFDR, 5Po scratch Lustre, 105 Go/s

Colloque 21 sept. 2016

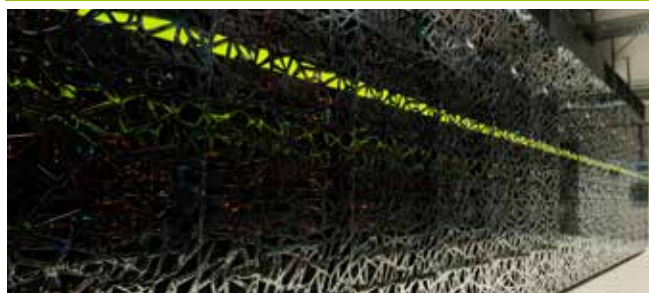
OCCIGEN : ATOS/BULL DLC B720

2 106 nœuds bi-processeurs (2x12 cœurs)
Intel Haswell, IB 4xFDR,
50 544 cœurs, 64 ou 128 Go/nœud,
5 Po scratch Lustre, 105 Go/s

Elec. 935 kW, Refroidissement par Eau tiède
sur procs (35°) + air

PUE < 1,1

Janvier 2015 : 2,1 Pflop/s



Colloque 1 oct. 2010

JADE

2 880 nœuds bi-processeurs
1 536 nœuds Harpertown, IB DDR,
1 344 nœuds Nehalem-EP, IB QDR
23 040 cœurs, 4 Go/cœur, 700 To Lustre
Elec. 1000 kW, Portes eau froide (12°C)

Juillet 2010 : 267 Tflop/s



Inauguration + Colloque 5 nov. 2008

JADE : SGI Altix ICE 8200 EX

1536 nœuds bi-processeurs (2x4 cœurs)
INTEL Xeon – Harpertown

Elec. 600 kW, Portes eau froide (12°C)

Septembre 2008 : 147 Tflop/s



Décembre 2007 : 2,4 Tflop/s



IBM/SP4

9 nœuds 32 Power 4 (64 Go)
5 nœuds 16 Power 5 (32 Go)
4 To de disques (GoFS)

STOCKAGE, ARCHITECTURE "DATA CENTRIQUE"

Une architecture « data centrique » reposant sur un nouveau cluster Lustre.

Le CINES a construit et fait évoluer une architecture « data centrique » où le service de fichiers est positionné au centre de l'environnement calcul afin de servir toutes les entités. Il est à la fois en connexion permanente avec les calculateurs et dispose d'un rythme d'évolution propre. Il est constitué des parties suivantes :

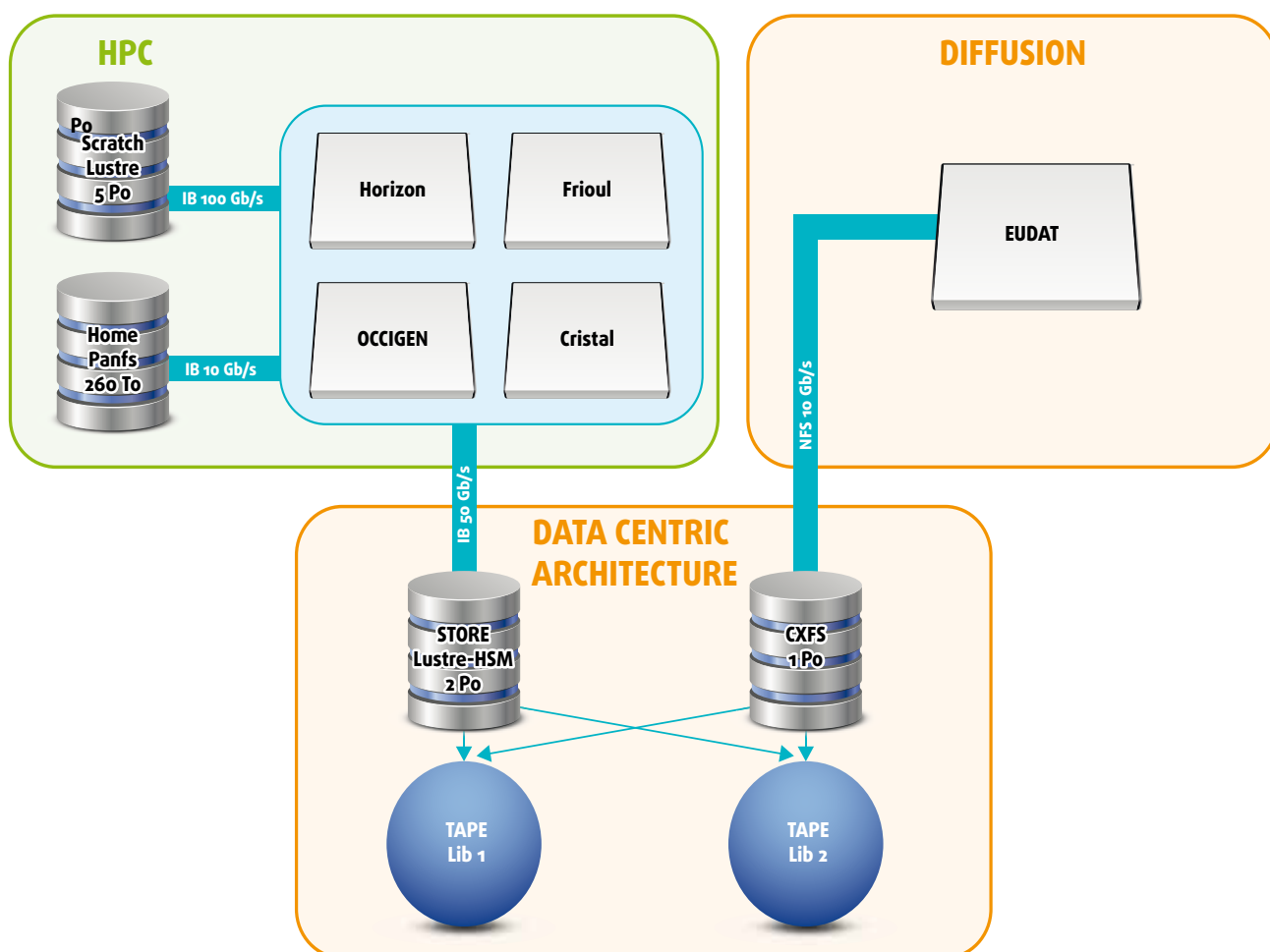
- Le stockage commun sécurisé dit de niveau 2, le niveau 1 correspondant au stockage local rapide du calculateur (/scratch) sous Lustre ;
- L'interface de transfert vers les calculateurs : les passerelles ;

- Le « back-end » HSM qui optimise la gestion de la ressource disque et assure la sécurité des données. Le mouvement de données vers les robotiques est opéré par une série de « datamovers ».

Cet environnement permet d'offrir 2 Po de données sauvegardées sur l'espace (/store2) du serveur de fichiers avec des débits cumulés de 50 Go/s vers le supercalculateur OCCIGEN. L'espace de stockage local de 5 Po directement relié au supercalculateur à haut débit (100 Go/s) n'est pas sauvegardé.

Au sein de l'espace sauvegardé, les fichiers sont automatiquement migrés vers des bandes magnétiques sur deux robotiques (hébergées dans deux salles machines différentes) d'une capacité actuelle de 6 Po chacune (et extensibles à 10 Po), permettant de faire une double copie des données migrées.

L'utilisation globale de l'espace /store a doublé depuis 2015 après la mise en service d'OCCIGEN.



LOGICIELS DISPONIBLES

La liste des logiciels scientifiques par domaine (applications, outils, bibliothèques, compilateurs, graphisme) est disponible sur le site web du CINES :

<https://www.cines.fr/logiciels-scientifiques-occigen/>

UTILISATION DES RESSOURCES

- consommation **DARI** par Comité Thématique, demandes en Heures et nombre de projets,

CT	Nombre d'heures consommées		
	2015	2016	2017
CT1	19 408 617	22 680 431	22 112 996
CT2A	50 261 090	44 603 696	88 813 623
CT2B	51 143 660	61 930 551	91 721 076
CT3	1 460 703	1 228 821	3 491 578
CT4	53 660 901	67 698 947	92 616 731
CT5	45 936 495	47 815 471	78 860 109
CT6	7 120 087	17 163 973	3 393 479
CT7	24 352 938	25 661 163	41 749 709
CT8	14 290 965	17 613 279	20 002 912
CT9	20 340 095	26 052 896	34 432 361
CT10	8 597 175	4 328 675	50 980 269
Total consommées	296 572 726	336 777 903	528 174 843
Demandées	384 404 000	544 197 254	666 776 137
Nb de projets		340	338

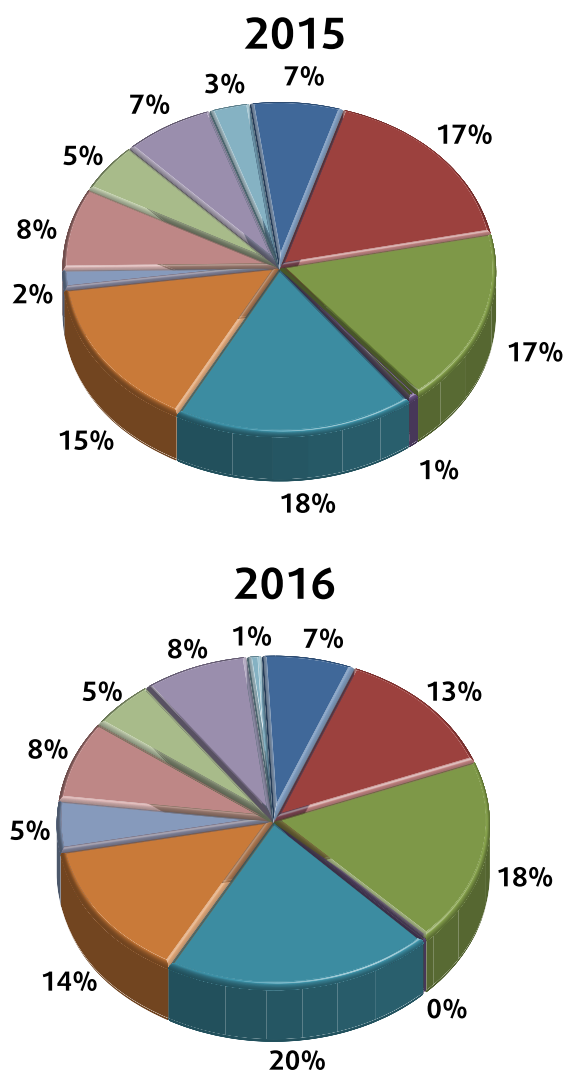
Demandes DARI en 2014 sur JADE :
session 1 = 218 211 400 heures pour
304 projets

Demandes en 2015 sur OCCIGEN :
session 1 = 334 630 000 h
session 2 = 50 774 000 h

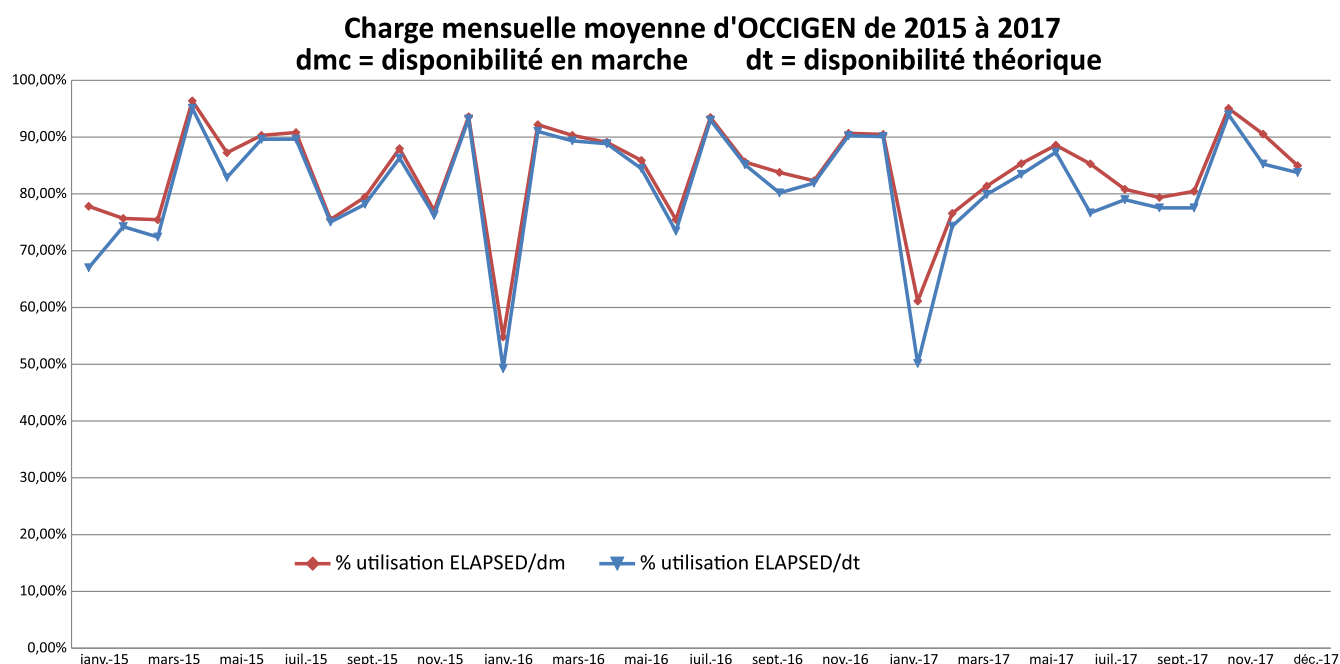
Demandes en 2016 sur OCCIGEN :
session 1 = 457 832 254 h pour 310 projets
session 2 = 86 365 000 h

Demandes en 2017 sur OCCIGEN :
session A1 = 252 597 129 h
session A2 = 414 179 008 h

Consommation par CT



- Évolution de la charge d'OCCIGEN de 2015 à fin 2017 .



COLLABORATIONS

Avec GENCI et les autres centres nationaux et européens :

- Contribution à la mise en place du réseau CCFR (réseau national entre Tier1/0).
- Participation à la mise en place d'une interface commune entre les centres nationaux pour les ouvertures de comptes utilisateur (outil développé, hébergé et maintenu par le TGCC).
- DARI (application développée, hébergée et maintenue par le CINES) :
 - Programmation et mise en œuvre des adaptations liées à la réorganisation des sessions GENCI.
 - Mise à disposition d'outils de rapport et statistiques en ligne.
 - Gestion des différentes sessions.
- Gestion et hébergement du « Peer Review » de PRACE.

Dans le cadre de projets européens :

- HPC Europa :
HPC Europa a offert à des chercheurs européens la possibilité de bénéficier d'un haut niveau de services reposant sur un hébergement dans l'un des pays partenaires et un accès transnational aux équipements de calcul intensif les plus avancés en Europe disponibles dans ce pays. Dans ce cadre le CINES a fourni à des dizaines de chercheurs européens des heures de calcul, le tutorat et le support de spécialistes ainsi que des opportunités de collaboration.
- PRACE :
À ce jour le projet est dans sa cinquième phase d'implémentation (PRACE 5IP) dont le contrat a été signé en

2017. Dans les différentes phases d'implémentation de PRACE le CINES s'est positionné essentiellement sur les « Workpackages » concernant l'amélioration de codes, la définition et la maintenance de « benchmarks » et l'hébergement et la mise en œuvre de prototypes matériels et logiciels. Ainsi dans le cadre du programme PCP (Pre-Commercial Procurement) de PRACE 3IP, le Centre a installé et mis à disposition un équipement d'ATOS/Bull basé sur des processeurs Intel KNL.

Le cluster-prototype PCP est constitué d'un rack équipé de 56 lames Bull Sequana X1210 Intel® Xeon™ "Knights Landing". Chaque lame inclut trois nœuds de calcul Intel® Xeon-Phi™ équipés de carte-mères identiques (Intel® Groveport platform) et d'une carte d'interconnexion InfiniBand EDR, et d'une plaque froide (« cold plate ») commune pour refroidir l'ensemble des composants. Le stockage est assuré par des disques HDD/SDD partagés par réseau Ethernet.

Au total, le cluster met donc à disposition 168 nœuds de calcul (Intel® Xeon-Phi™ Knights Landing 7250 16GB GDDR5 215W).

Le CINES est également impliqué dans les activités suivantes :

- Supporter des PME européennes souhaitant utiliser les infrastructures de calcul PRACE dans le cadre du programme SHAPE.
- Gérer la collaboration avec EUDAT pour valider l'interopérabilité des services de gestion de données.

- Délivrer des formations liées au calcul intensif dans le cadre des PATC (Prace Advanced Training Centers) : Dans le cadre du projet européen PRACE, le CINES contribue à l'initiative de formations calcul PATC qui est coordonnée au niveau français par la Maison de la Simulation.

Une formation « débogage et optimisation » de 4 jours a ainsi été organisée en novembre 2015, 2016 et décembre 2017.

Participation à la cellule de veille technologique :

Sous l'égide de GENCI une cellule de veille technologique regroupant les 3 centres nationaux, l'INRIA, et des représentants de méso-centres a été créée. Elle a donné lieu à l'installation de 3 prototypes permettant de tester les technologies candidates pour la réalisation des supercalculateurs « exaflopiques » de la future décennie.

Dans ce cadre le CINES a accueilli une plate-forme Bull (FRIOUL) à base de processeurs Xeon phi (KNL) et a participé aux tests du deuxième prototype basé sur des processeurs Open Power équipés de GPUs Nvidia installé à l'IDRIS. Le troisième prototype devant être hébergé en 2018 par le TGCC.



La plate-forme FRIOUL

Collaboration avec le LIRMM :

- Collaboration avec le LIRMM dans le cadre du projet ICARE (utilisation des énergies renouvelables pour le « computing », réalisation d'un prototype technologique matériel et logiciel de « supercalculateur » à consommation largement réduite, alimenté par des énergies renouvelables.)

VEILLE TECHNOLOGIQUE

- Installation du prototype FRIOUL à base de Bull Xeon phi, en fin d'été 2016.

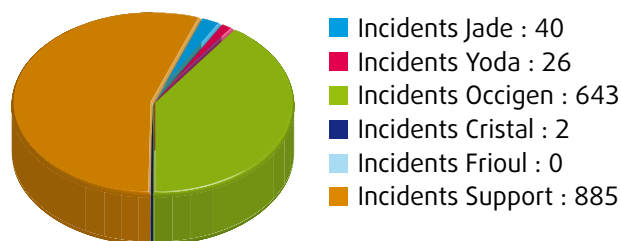
	Frioul
Vendeur	Atos
Type de nœud	Intel Knight Landing (KNL 7250)
Cœurs par nœud	68 cœurs x86
Réseau d'interconnexion entre les nœuds	Infiniband EDR
Nombre de nœuds disponibles	48
Puissance crête	146 Tflops

- Tests et portages de codes sur KNL effectués par des utilisateurs avec le soutien des équipes du CINES.
- Collaboration avec EoCoE (Energy oriented Centre of Excellence) piloté par la Maison de la Simulation intéressée par des tests sur la technologie KNL, séminaire organisé au CINES du 4 au 6 /10/2017.

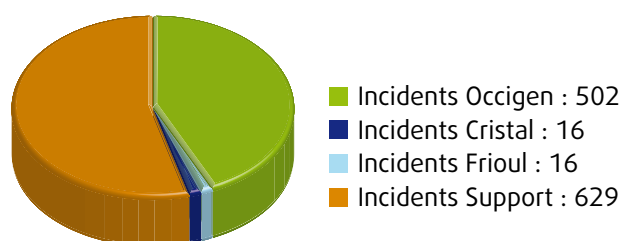
SERVICES

- Support et expertises :

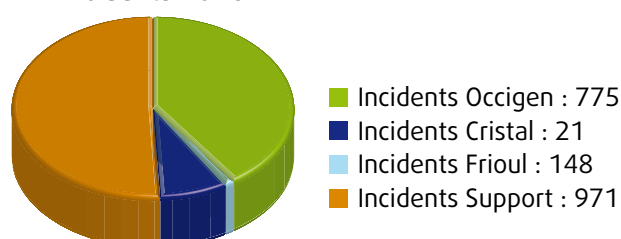
Incidents 2015



Incidents 2016



Incidents 2016



CALCUL DE HAUTE PERFORMANCE

- Participations à des congrès :
Le CINES est présent au congrès « Supercomputing » aux Etats Unis, à l'ISC en Europe, il participe aux manifestations nationales : forums ORAP et Ter@tec, journées France-grilles, etc.
- Participation au programme national SiMSEO et au programme européen SHAPE :
Le programme SiMSEO lancé dans le cadre des Projets d'Investissements d'Avenir du Commissariat Général à
- Formations hors PATC :

l'Investissement, vise à aider les PME et ETI à intégrer la simulation et le calcul intensif dans leur modèle de développement. Dans ce cadre les moyens nationaux de calcul de GENCI, dont le CINES, sont accessibles aux industriels français (grands groupes et PME) pour des travaux de recherche ouverte. Les demandes sont à adresser à GENCI et comporter un engagement de publication des résultats des travaux réalisés sur les ressources attribuées.
Le programme SHAPE est l'équivalent au niveau européen.

	Titre	Début	Fin
CI05	Formation à l'utilisation de gaussian sur OCCIGEN	5 avril 2016 14 h 00	6 avril 2016 12 h 00
CI01	Introduction aux architectures parallèles et au calculateur Occigen	16 juin 2015 14 h 00	17 juin 2015 17 h 00
CI02	Calcul parallèle	9 Mars 2015 14 h	13 Mars 2015 12 h
CI01	Introduction aux architectures parallèles et au calculateur Occigen	30 janvier 2015	

COMITÉ DES UTILISATEURS

- Élection de nouveaux membres :
Les quelques 1500 utilisateurs calcul du CINES sont représentés par le Comité des Chercheurs Calculant au CINES (CCCC ou C4). Le C4 est l'instance d'échanges entre le CINES et ses utilisateurs calcul. Le C4 est élu pour une durée de 3 ans. Les utilisateurs de chaque Comité Thématique (CT) élisent 2 représentants : un titulaire et un suppléant. Les membres du C4 ont été renouvelés en décembre 2016.
Il s'est réuni trois fois en 2017 : le 7 avril, le 6 octobre et 19 décembre.

- Liste des comités :

CT1	Environnement
CT2A	Écoulements non réactifs
CT2B	Écoulements réactifs ou/et multiphasiques
CT3	Biologie et santé
CT4	Astrophysique et géophysique
CT5	Physique théorique et physique des plasmas
CT6	Informatique, algorithmique et mathématiques
CT7	Modélisation moléculaires appliquée à la biologie
CT8	Chimie quantique et modélisation moléculaire
CT9	Physique, chimie et propriétés des matériaux
CT10	Nouvelles applications et applications transversales du calcul

- Liste des membres par comité :

CT1	Sébastien	Theetten
CT2A	Julien	Bodart
CT2A	Florent	Duchaine
CT2B	Florent	Duchaine
CT4	Alain	Minuissi
CT5	Virginie	Grandgirard
CT5	Rachel	Nuter
CT7	Jean-Philip	Piquemal
CT7	Nicolas	Floquet
CT8	Sébastien	Le Roux
CT9	Sébastien	Le Roux
CT9	Evelyne	Lampin

› Contexte historique	25
› Équipe d'experts	25
› Refonte de la plateforme PAC	26
› Avancées en matière d'expertise sur les formats	27
› Projet d'agrément HADS (données de santé)	28
› Accroissement des archives, partenaires et typologies	28
› Collaborations au niveau national et diffusion	32
› Reconnaissance nationale et internationale	34
› Focus sur quelques projets startégiques	37

CONTEXTE HISTORIQUE

Le premier document faisant référence à un projet d'archivage pérenne au CINES, présenté par Francis Daumas à la tutelle et au conseil scientifique du Centre, date de 2002. Suite à un audit effectué en 2007, la lettre de cadrage de la DGRI et de la DGES en date de 2008, redéfinissant les missions essentielles du CINES, confirmait l'archivage comme l'une d'entre elles. L'objectif est de proposer à la communauté de l'Enseignement Supérieur et de la Recherche une solution sûre, efficace et agréée par les Archives de France, pour la préservation du patrimoine scientifique national.

L'archivage pérenne du document numérique a pour objet d'assurer une triple fonction : conserver le document, le rendre accessible et en préserver l'intelligibilité. Ces trois services sont conçus pour toute la durée nécessaire de conservation des informations numériques (soit sur le moyen et long termes).

Pour rappel, en matière d'archivage pérenne, de 2008 à 2014, le CINES a :

- mis en place une **équipe de spécialistes**, composée d'informaticiens et d'archivistes, au sein d'un nouveau département (Département Archivage et Diffusion) ;
- développé et mis en production une **plateforme d'archivage (PAC)** respectant les normes et les standards du domaine, y compris une réplication des données à distance (à Lyon) ;
- obtenu **l'agrément du SIAF** (Service Interministériel des Archives de France) **et du DSA** (Data Seal of Approval, accréditation reconnue par la Commission européenne) ;
- participé, en tant que partenaire spécialiste de l'archivage, au **projet européen EUDAT**, visant à mettre en œuvre une infrastructure de données européenne (Common Data Infrastructure) ;
- été désigné comme **centre archiveur national** pour les thèses de doctorat électroniques, la TGIR HumNum, les collections numérisées du portail PERSEE, les publications de HAL, les fonds de plusieurs bibliothèques universitaires (Cujas, BUPMC, Sainte Geneviève, ...), etc.

L'offre du CINES se décline suivant deux niveaux de « service »

- Niveau 1 : il est destiné aux archives à accès fréquent et correspond à :
 - 1 copie sur disque et 2 copies sur bandes
 - 1 de ces 3 copies est faite sur un site distant (à plus de 300 km du CINES)
 - En option une deuxième copie sur disque peut être proposée

- Niveau 2 : il adresse des archives à accès occasionnel et est composé de
 - 2 copies sur bandes au CINES dans 2 salles machines distinctes
 - 1 troisième copie des archives est faite sur un site distant (à plus de 300 km du CINES)

L'archivage de données numériques au CINES est ouvert en priorité à tous les établissements et organismes de l'Enseignement Supérieur et de la Recherche et fait l'objet d'une convention comportant un engagement de confier l'archivage au CINES pour une période supérieure à 5 ans.

ÉQUIPE D'EXPERTS

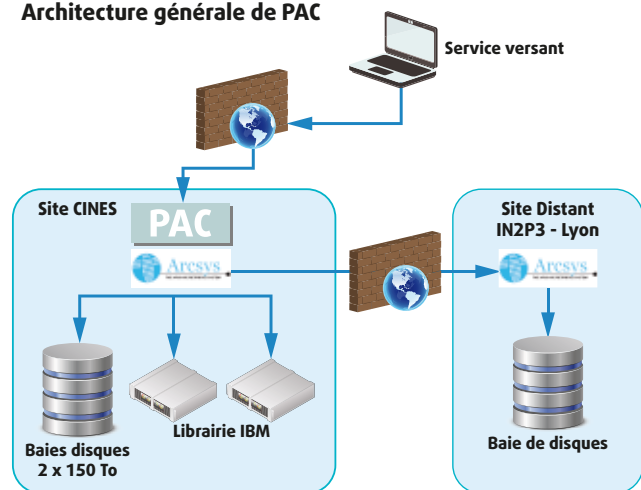
L'équipe d'experts chargés de l'archivage pérenne propose un certain nombre de services, dont :

- L'analyse et la mise en œuvre de projets d'archivage sur la plateforme PAC.
- Des **expertises** :
 - Sur les **formats des fichiers**
 - Sur les **métadonnées** associées aux données à archiver
 - Sur **l'assurance qualité** et les **normes et standards** du domaine
- La constitution de DMP (Data Management Plan ou « plan de gestion de données »).

L'équipe du CINES **accompagne les producteurs de données dans la rédaction et l'implémentation de leur plan de gestion des données**. Les phases préalables à la mise en œuvre d'un projet d'archivage sont notamment l'occasion d'aborder les questions : de droit sur les données, de métadonnées ou encore de durée de conservation.

REFONTE DE LA PLATEFORME PAC

Architecture générale de PAC



Les matériels utilisés pour l'archivage depuis 2008 arrivant en fin de maintenance, il était nécessaire d'envisager leur remplacement. **Ce changement de la totalité de l'infrastructure d'archivage**, sur le site principal du CINES ainsi que sur le site distant du CC-IN2P3 à Villeurbanne, a dû être réalisé entre 2015 et 2017 parallèlement au maintien du flux de production (versement de nouvelles données, accès aux archives, etc.). Les opérations de **migration des données** devaient également permettre de maintenir en permanence le même niveau de service, à savoir au moins 3 copies d'archivage (2 en local et 1 sur site distant).

Une évolution selon 6 axes

La nouvelle architecture PAC-V3, mise en place en 2015 et 2016, présente plusieurs évolutions :

- les machines supportant l'application PAC sont désormais hébergées dans une **architecture de virtualisation** afin notamment d'améliorer le délai de reprise d'activité en cas d'incident majeur ;
- la capacité de stockage sur disques est passée de **40 To à 300 To** (avec la possibilité d'ajouter 300To supplémentaires) ;
- les **copies bandes sont pilotées directement par le SAE** (Système d'Archivage Electronique), et non plus par un logiciel tiers ;
- la copie distante est pilotée directement par le SAE et effectuée désormais de manière **synchrone** ;
- un **renforcement de la sécurité** physique et logique a été effectué ainsi que l'ajout d'un anti-virus dans la chaîne de traitements ;
- **plusieurs niveaux de service** sont désormais proposés (offre de stockage sur supports disques et bandes magnétiques, paramétrage différencié des contrôles qualité effectués sur les fonds d'archives, etc.).

Nouvelles infrastructures matérielles pour le site principal Montpelliérain

Dans le cadre du décret relatif à la Protection du Potentiel Scientifique et Technique¹, et suite à son **passage en Zone à Régime Restrictif (ZRR)** par l'arrêté ministériel du 6 novembre 2015 (diffusion restreinte), le CINES a procédé à un renforcement de ses dispositifs de protection périmétrique ainsi qu'à une réorganisation de ses salles machines. L'installation des nouvelles infrastructures d'archivage respecte toutefois les mêmes principes que précédemment, à savoir, la répartition des différentes copies (disques et bandes) dans des salles machines distinctes. La sécurisation physique des baies de disques a également été renforcée par l'adjonction de serrures avec lecteur RFID et contrôle d'accès.

Nouvelles infrastructures matérielles pour le site secondaire Lyonnais

L'externalisation d'une copie de sauvegarde des données archivées dans la plateforme PAC a été mise en place dès 2012, suite à la signature d'une **convention avec le centre de calcul de l'IN2P3**. Cette réplication distante (de plusieurs centaines de kilomètres) des archives et des éléments techniques de la plateforme d'archivage constitue **une des briques fondamentales du plan de reprise d'activité**.

Le renouvellement de la convention avec le centre de calcul de l'IN2P3 a été l'occasion de redéfinir le service fourni, d'augmenter la capacité de stockage et d'améliorer la réplication sur site distant afin d'avoir une copie d'archivage directement gérée par le SAE (Système d'Archivage Electronique).

Le CC-IN2P3 a ainsi mis à disposition du CINES dès l'automne 2015 un nouveau dispositif matériel dans les mêmes conditions de sécurité physique que la précédente configuration mais avec la **possibilité de le piloter directement au moyen du SAE** (et non plus au moyen d'une application tierce).

¹ Décret N° 2011-1425 du 2 novembre 2011, suivi de l'arrêté du PM du 3 juillet 2012, de la circulaire 3415 de novembre 2012, et de lettres relatives à la définition et conditions d'accès aux Zones à Régime Restrictif – avril 2013

Effacement de données archivées et retrait des anciens supports

Conformément aux résultats du groupe de travail SIAF sur l'élimination des données et à la procédure d'élimination du CINES, les matériels sortis de production dans le cadre du changement d'infrastructure ont subi la procédure suivante : déréférencement complet (tant au niveau du référentiel Arcsys qu'au niveau matériel) et/ou écritures sur l'intégralité de leur surface de stockage, puis extraction de la salle machine et surveillance constante par des agents du CINES jusqu'à ce qu'ils leur appliquent un procédé de destruction mécanique, enfin, envoi à une société spécialisée dans le retraitement des déchets informatiques.

Par ailleurs, un second processus d'élimination de documents, pour raison d'invalidation, suite à un plagiat par exemple, a été mis en place et agréé. Il fait appel à des techniques d'effacement et de destruction de supports (similaire au processus décrit ci-dessus) et donne également lieu à l'édition d'un certificat de destruction.



AVANCÉES EN MATIÈRE D'EXPERTISE SUR LES FORMATS

L'archivage pérenne de documents électroniques, a conduit le CINES à se doter d'un ensemble de moyens humains et techniques pour pouvoir mettre en œuvre une expertise sur les formats de fichiers afin d'intégrer complètement la problématique des formats dans le processus. Les actions menées sont de plusieurs ordres : valorisation, sensibilisation, audit sur des échantillons, utilisation et intégration d'outils informatiques, etc.

Dans ce cadre, le CINES assure :

- Une **veille technologique** sur les formats de fichiers et les contrôleurs de qualité des fichiers générés par la communauté de l'enseignement supérieur et de la recherche ;
- Une sélection des formats d'archivage, mais aussi une identification des formats potentiellement archivables, en voie d'obsolescence et des formats obsolètes ;
- Une **validation technique** (au travers de l'**outil FACILE** et des plateformes d'archivage du CINES) pour s'assurer notamment que les fichiers envoyés au CINES pour archivage soient effectivement dans les formats dans lesquels ils se prévalent ;
- Une **conversion** (ou migration logique) de fichiers archivés dans un format jugé non pérenne en un autre format jugé pérenne, de manière à assurer la lisibilité des informations qu'ils contiennent ;
- La fourniture d'une **assistance** à ses utilisateurs pour la production et/ou la conversion de fichiers compatibles avec les exigences de l'archivage en la matière.



Les principales réalisations de la cellule

« expertise formats » sont :

- L'intégration dans PAC et FACILE des **formats TEI, jpeg2000 et GeoTiff et des formats 3D** (DAE et PLY) en partenariat avec la TGIR HumaNum ;
- L'évolution de l'outil FACILE (facile.cines.fr) : avec une amélioration de la robustesse et des performances, une nouvelle interface, l'envoi de plusieurs fichiers simultanément, des tutoriels pour corriger les fichiers, doc web service, sécurisation de l'interface en https, etc. ;
- L'amélioration de l'identification de fichiers envoyés pour archivage (signatures Droid v88) ;
- Le changement des outils de validation utilisés dans PAC et FACILE (retrait de Mencoder et ajout de MkvValidator...) ;
- La veille sur les projets internationaux de développement d'outils pour la validation des formats de fichiers : projet PREFORMA, projet VeraPDF, programme de refonte de JHOVE...

La suppression du format TXT/ASCII, au profit de TXT/UTF-8.

PROJET D'AGRÉMENT HADS (HÉBERGEMENT ET ARCHIVAGE DE DONNÉES DE SANTÉ)

Conformément au décret du 4 janvier 2006, l'hébergement et/ou l'archivage de données confidentielles de santé requiert l'obtention préalable d'un agrément délivré par le ministre en charge de la Santé, après avis de la CNIL et d'un comité d'agrément créé ad hoc (dont l'ASIP Santé assure le secrétariat). En effet, dans le cas de l'hébergement de données de santé, non seulement l'organisme responsable des traitements des données de santé doit être habilité à manipuler ces données mais également le centre hébergeur afin que la sécurité soit garantie de bout en bout.

Dans la perspective de pouvoir proposer ses services aux communautés de chercheurs dans le domaine de la santé, le CINES a initié en 2016 la constitution de deux dossiers de demande d'agrément : **un pour l'hébergement de serveurs comportant ce type de données** (en lien avec la mission statutaire d'hébergement du CINES), **l'autre pour la pérennisation de données de santé** sur ses infrastructures d'archivage pérenne qui sont agréées depuis 2010 par le SIAF (Service interministériel des Archives de France – cf. supra).

Dans le cadre de ce second contexte, le CINES a contribué à la réflexion nationale, en partenariat avec le SIAF et le Ministère de la Santé, sur l'évolution de la réglementation en vigueur pour une meilleure articulation entre l'agrément HADS (pour les données de santé) et celui du SIAF (pour les archives courantes et intermédiaires publiques). Ce double agrément sera en effet le sésame pour la

conservation pérenne de données de santé dans la plateforme d'archivage du CINES et ouvrir ainsi **l'offre du Centre aux entités détentrices de ce type de données** (INSERM, CHU, etc.).

ACCROISSEMENT DES ARCHIVES, PARTENAIRES ET TYPOLOGIES

Nouveaux partenaires

De nouveaux partenaires ont confié leurs archives numériques au CINES pendant la période 2015-2017 :

- l'Equipex **GEOSUD** (infrastructure de recherche « système terre ») piloté par l'IRSTEA pour la conservation de données satellitaires ;
- le Museum national d'Histoire naturelle (**MNHN**) pour la pérennisation de ses herbiers numérisés ;
- l'Institut national de recherches archéologiques préventives (**Inrap**), pour ses rapports de fouilles ;
- la Maison méditerranéenne des sciences de l'homme (**MMSH**) pour sa phonothèque ;
- les DSI de l'**INSERM**-ANRS, de l'**IRSTEA**, de **Sciences-Po** Paris, du PRES de **l'université de Lorraine** et du SCD **Lille2**.

Projets d'archives

Parallèlement à la mise en production de nouveaux fonds et collections d'archives provenant de ces établissements, des projets en lien avec les SHS (sciences humaines et sociales) sont en cours avec :

- l'UMS **Archéovision** du CNRS pour des images 3D ;
- **ORTOLANG** (Open Resources and Tools for Language) et l'ATILF (Analyse et Traitement Informatique de la Langue Française) pour des données de linguistique ;
- OpenEdition pour les **revues.org** ;
- l'**EFEU** (École Française d'Extrême Orient) pour des collections scientifiques et techniques issues de sa bibliothèque ;
- Et l'IRHT (institut de recherche et d'histoire des textes) pour ses manuscrits numérisés.

D'autres projets sont programmés avec :

- le **CNRS** (tous les documents issus de la dématérialisation de l'administration du CNRS : dossiers des agents, bulletins de salaires, comptes financiers, marchés publics, factures et leurs pièces-jointes, documents liés aux missions et achats, dossiers médicaux des agents, cahiers de laboratoire, etc.) ;
- l'**INRA** (phénotypage végétal) ;
- la COMUE **Sorbonne** Paris Cité ;
- ou encore la **Bibliothèque universitaire des langues et civilisations** (manuscrits numérisés).

Volumétrie : l'archivage en quelques chiffres

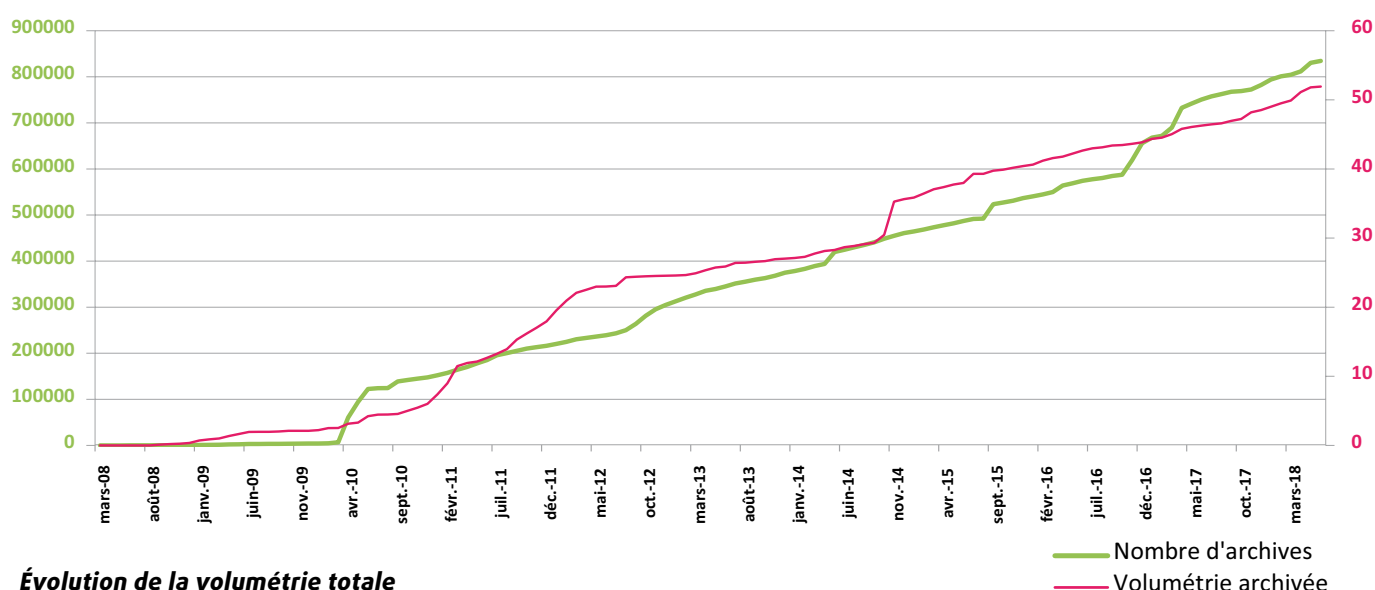
L'archivage au CINES au 31 décembre 2017, ce sont :

- 24 services versants en phase de production ;
- 50 fonds scientifiques, administratifs et patrimoniaux archivés sur les infrastructures d'archivage en production ;
- 55To utiles archivés en production (soit plus de 150To de données réellement stockées sur les infrastructures du CINES) ;
- En 2015 le CINES fêtait la 500 000^e archive confiée à PAC (pour rappel, une archive correspond à une thèse et ses annexes, une archive c'est également

un ouvrage numérisé ou un dossier de personnel, etc.). **En décembre 2017, la plateforme PAC gère 750 000 archives** pour plus de 50 To de données à très forte valeur ajoutée préservées ;

- Et plus de 20 000 000 fichiers archivés.

Les partenaires en production : l'ABES, Persée, Hal, BIU Santé, BUPMC, Cujas, TGIR HumaNum (SLDR, Cocoon, EFEO, Archéovision, ATILF, OpenEdition, MAE-René Ginouvès, IRHT...), BSG, Cour des Comptes, INSERM, Inrap, Université de Lorraine, OATAO, IRSTEA, GéoSUD, IPSL, CORIA-PRECCINSTA, université de Strasbourg, CUNI, MMSH, Sciences Po, Université de Lille2.



Évolution de la volumétrie totale

Nom de l'établissement	Typologie des données archivées au CINES
ABES	Thèses électroniques de doctorat dont le dépôt au CINES est désormais obligatoire pour tous les établissements français (suite à la publication de l'arrêté du 25 mai 2016 fixant le cadre national de la formation et les modalités conduisant à la délivrance du diplôme national de doctorat ; cet arrêté abroge celui du 7 août 2006 qui était relatif aux modalités de dépôt, de signalement, de reproduction, de diffusion et de conservation des thèses ou des travaux présentés en soutenance en vue d'un doctorat).
Persée	Plus vaste ressource d'articles et de documents (revues numérisées principalement) en sciences humaines et sociales francophones.
HAL (Archive ouverte développée et administrée par le CCSD)	Articles scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
BIU Santé	Ouvrages et périodiques numérisés en texte intégral de la Bibliothèque numérique Medic@. Banque d'images et de portraits. Éditions numériques encodées en TEI (Text Encoding Initiative)
BUPMC (Bibliothèque de l'université Pierre et Marie Curie)	Ouvrages anciens scientifiques et médicaux, périodiques, thèses et manuscrits disponibles en ligne au travers de la Jubilotheque.
Bibliothèque Cujas	Ouvrages numérisés du fonds ancien de la bibliothèque.

Nom de l'établissement	Typologie des données archivées au CINES
TGIR HumaNum (dont les services versants sont : • SLDR / UMR6057 – Laboratoire Parole et Langage (LPL) • Cocoon / UMR7107 – Laboratoire de Langues & Civilisations à Tradition Orale (LACITO) / Laboratoire Ligérien de Linguistique (LLL) • Ecole Française d'Extrême Orient (EFEO) • Archéovision / UMS3657 • Centre de données socio-politiques (CDSP) • Ortolang (équipement d'excellence) • OpenEdition • Maison des Sciences de l'Homme René Ginouvès • Maison méditerranéenne des sciences de l'homme (MMSH)	SLDR : Ressources orales de recherche en linguistique. Cocoon : Enregistrements de parole issus de la communauté scientifique en sciences humaines et sociales. EFEO : Photothèque. Archéovision : Données d'archéologie exprimées en format 3D. CDSP : Enquêtes qualitatives et quantitatives effectuées par la communauté scientifique en sciences sociales. Ortolang : réservoir de données (corpus, lexiques, dictionnaires, etc.) en linguistique. OpenEdition : ressources électroniques en sciences humaines et sociales (ouvrages, revues, carnets, événements...). Maison des Sciences de l'Homme René Ginouvès : données en archéologie et ethnologie. MMSH : Enregistrements de paroles issus d'enquêtes effectuées par la communauté scientifique en sciences humaines et sociales et accessibles au travers de la phonothèque de la MMSH.
Bibliothèque Sainte Geneviève & SCD (Service Commun de Documentation) de Paris 3 (université Sorbonne nouvelle)	Livres imprimés au 15^e et au début du 16^e siècles : éditions rares et exemplaires remarquables. Livres du 19^e siècle (1830-1880) conservés à la Bibliothèque Sainte-Geneviève et identifiés comme difficilement accessibles (i.e. uniques ou peu d'exemplaires conservés dans les bibliothèques françaises). Récits de voyages nordiques (i.e. voyages dans les pays du nord et voyageurs nordiques dans le monde) imprimés du 16^e au 19^e siècle. Livres rares anciens imprimés entre la fin du 16^e siècle et la fin du 18^e siècle identifiés comme difficilement accessibles (i.e. uniques ou peu d'exemplaires conservés dans les bibliothèques françaises). Factum (i.e. sources jurisprudentielles imprimées) de la fin du 16^e siècle au début du 19^e siècle. Fonds Dezos de la Roquette : manuscrits et cartes numérisés dans le cadre de l'appel à projets Bibliothèque Scientifique numérique - Axe 5 (présentation de la BSN : cf. supra). Périodiques (toutes périodes et toutes thématiques confondues). Documents iconographiques conservés à la Bibliothèque Sainte-Geneviève (toutes périodes et toutes thématiques confondues). Manuscrits (toutes périodes et toutes thématiques confondues). Fonds Ferdinand Denis : manuscrits et imprimés ayant appartenu à Ferdinand Denis. Sélection numérisée issue du fonds Tessin, fruit d'un don de la bibliothèque de l'Institut suédois (environ 5000 volumes préservés à la BSG dès 1972). Les ouvrages sélectionnés proviennent du fonds d'histoire de l'art constitué par Gunnar W. Lundberg (1903-1986), conseiller culturel à l'Ambassade de Suède à Paris et fondateur de l'Institut Tessin en 1933. Cette sélection comporte des catalogues d'exposition, des ouvrages sur Carl Larsson, Julius Kronberg, Bruno Liljefors, Anders Zorn notamment, des ouvrages illustrés relatifs aux vêtements traditionnels, aux châteaux suédois. Musique ancienne (XI^e – XVIII^e) : sources manuscrites et imprimées associées aux fonds de la bibliothèque Saint Geneviève (Réserve, Nordique, Fonds général).
Cour des Comptes	Archives des activités de contrôle de la Cour des comptes (contrôles administratifs et juridictionnels), des Chambres régionales et territoriales des comptes (CRTC) (contrôles administratifs, juridictionnels et budgétaires), de la Cour de discipline budgétaire et financière (CDBF) et des activités du Conseil des prélèvements obligatoires (CPO).

Nom de l'établissement	Typologie des données archivées au CINES
INSERM	Images fixes et animées issues de la base de données Serimedis de la photothèque de l'Inserm. Cette base de données regroupe les images produites par les différentes entités de l'Inserm, laboratoires et administration.
Agence nationale de recherche sur le sida et les hépatites virales (ANRS)	Cahiers d'observation des essais cliniques et les documents associés issus d'études (essais cliniques).
Inrap	Rapports de fouilles et de diagnostic produits par l'Inrap ou l'association dont il est l'héritier, l'Afan (Association pour les fouilles archéologiques nationales), au terme d'une intervention archéologique.
Université de Lorraine	Thèses de doctorat, thèses d'exercice et mémoires soutenus dans les universités fondatrices Documents patrimoniaux détenus par la bibliothèque
Université de Strasbourg	Données issues des bases arrêtées du logiciel de gestion des ressources humaines « Harpège » édité par l'AMUE (Agence de mutualisation des établissements). Récapitulatifs nominatifs annuels de paie (historiques de paie, déclaration annuelles des données sociales (DADS), déclarations fiscales, etc.).
OATAO (archive ouverte institutionnelle toulousaine)	Publications et documents issus de la recherche scientifique et de l'enseignement produits par les institutions toulousaines.
IRSTEA	Publications numérisées et nativement électroniques de l'établissement.
GéoSUD (équipement d'excellence faisant partie de l'infrastructure de recherche "système terre")	Images satellitaires brutes et orthorectifiées acquises dans le cadre du programme Equipex-GEOSUD.
Institut Pierre-Simon-Laplace (IPSL)	Données satellitaires. Données scientifiques issues de simulation en chimie, advection et météorologie. Données d'observation satellitaires et in-situ (et acquises en temps réel), données modèles (prévisions, simulations, réanalyses) acquises dans le cadre de campagnes de mesures, données instrumentales non-reproductibles, autres données difficilement reproductibles. Bases de données nationales ou internationales. Grands jeux de données satellitaires ou modèles, et autres données reproductibles mais difficiles ou coûteuses à rapatrier. Données résultats d'observations calculées sur le cluster HPC Earth Simulator au Japon.
CORIA-PRECCINSTA (Complexe de Recherche Interprofessionnel en Aérothermochimie)	Données scientifiques issues de simulation en combustion (avec le bruleur PRECCINSTA)
CUNI (université Charles de Prague, République Tchèque)	Données en linguistique et en programmation neurolinguistique issues du répertoire LINDAT/CLARIN.
Sciences Po Paris	Thèses de doctorat, thèses d'exercice et mémoires soutenus à l'Université de Lorraine (depuis 2012) et dans les 4 universités fondatrices.
Université de Lille	Thèses d'exercice et mémoires de l'Université Lille 2 (droit et santé), base de données PEPITE (Panorama d'Es Productions universitaires).
Université de Lyon 1	Thèses d'exercice (médecine, pharmacie, odontologie) de la bibliothèque universitaire.
Bibliothèque universitaire des langues et civilisations (BULAC)	Manuscrits proche et moyen orientaux (arabes, persans et turcs ottomans). Résultats d'autres opérations de numérisations effectuées par la BULAC.
IRHT	Manuscrits du 12 ^e au 18 ^e siècle, incunables (fin 15 ^e , début 16 ^e siècle), cartulaire du XIII ^e siècle. Issus des collections de la bibliothèque et des archives du château de Chantilly, ces documents sont uniques, rares et précieux. Reproductions numériques de documents, principalement des manuscrits, produites dans le cadre de la mission de numérisation systématique des manuscrits médiévaux des bibliothèques publiques de France menée par l'Institut de recherche et d'histoire des textes (IRHT) depuis 1979.

COLLABORATIONS AU NIVEAU NATIONAL ET DIFFUSION

En dehors des nombreuses réunions et discussions qu'entretient le CINES avec les Archives Nationales et les Archives de France, le CINES est impliqué dans plusieurs collaborations avec les acteurs majeurs du domaine.

VITAM

Partant du constat qu'il n'y a pas de dématérialisation possible sans archivage électronique, Vitam est un programme interministériel co-piloté par le Comité interministériel aux Archives de France (CIAF) et la Direction interministérielle du numérique et du système d'information et de communication de l'État (DINSIC). Lancé à l'initiative conjointe des Ministères des Armées, de la Culture et des Affaires étrangères et inauguré en 2015, ce programme a 3 objectifs, soutenus par 5 projets :

- La réalisation de la solution logicielle libre Vitam d'archivage numérique permettant la prise en charge, la conservation et la consultation sécurisée de très gros volumes d'archives numériques définitives, intermédiaires, voire courantes ;
- La mise en place de plates-formes d'archivage utilisant la solution logicielle Vitam, dans chacun des trois ministères porteurs, via les projets ministériels : Saphir (MAEDI), Adamant (MCC/Archives nationales) et Archipel (MinDef) ;
- La diffusion et la réutilisation la plus large de la solution logicielle Vitam, en favorisant et fédérant l'ensemble des actions de soutien financier, de sensibilisation et d'accompagnement en matière d'archivage numérique au-delà des trois ministères porteurs du programme, via le projet Ad-Essor (MCC/Service interministériel des Archives de France).

Le CINES a signé le 20 juin 2016 avec la DINSIC une convention de partenariat VITAM/CINES. À ce titre, le Centre a participé aux ateliers, aux phases de spécifications et de tests, avec pour **objectif à moyen terme de baser sa solution d'archivage électronique sur VITAM.** Les projets Adamant (Archives nationales, Ministère de la Culture), Archipel (Ministère des Armées) et Saphir (Ministère de l'Europe et des Affaires étrangères) sont actuellement en cours d'intégration de VITAM (dont la première version de production a été livrée en avril 2018) afin de disposer d'une solution d'archivage électronique dédiée à leurs besoins métier. Nul doute que PAC évoluera dans le même sens.

Le défi à relever dans le futur proche réside dans l'intégration de solutions permettant de répondre aux besoins d'archivage des masses considérables de données produites.

BSN (Bibliothèque Scientifique Numérique)

Créée en 2009, à l'initiative du ministère de l'enseignement supérieur et de la recherche en fédérant de nombreux acteurs des universités et organismes de recherche, la bibliothèque scientifique numérique (BSN) veille à ce que tout enseignant-chercheur, chercheur et étudiant dispose d'une information scientifique pertinente et d'outils les plus performants possibles.

En accord avec les orientations de la Commission européenne, la BSN privilégie l'accès ouvert aux documents et données scientifiques sous toutes leurs formes. Ce développement d'un accès large et rapide passe par l'innovation, les négociations avec les éditeurs ou le soutien aux archives ouvertes, en tenant compte des différences entre les disciplines.

La coordination des acteurs de l'enseignement supérieur et de la recherche est à la base de BSN.

La BSN comporte 10 groupes de travail :

- BSN 1 : Acquisitions des revues (courant) et des archives
- BSN 2 : Dispositif d'accès et d'hébergement
- BSN 3 : Dispositifs de signalement
- BSN 4 : Open Access
- BSN 5 : Numérisation
- BSN 6 : Archivage pérenne
- BSN 7 : Édition scientifique
- BSN 8 : Fourniture de documents, prêt
- BSN 9 : Formation, compétences et usages
- BSN 10 : Données de la recherche

Le CINES est **pilote du groupe de travail BSN 6** consacré à l'archivage pérenne, activité transverse aux autres préoccupations. Il participe également aux groupes BSN 9 et BSN 10.

Le coût de l'archivage pérenne est tel qu'il nécessite une politique nationale portée au niveau interministériel et facilitant des coopérations internationales, notamment européennes. BSN constitue un des lieux de cette coopération.

En 2018, la **Bibliothèque Scientifique Numérique (BSN) devient le Comité pour la Science Ouverte (CoSO)** présidé par le Directeur général de la recherche et de l'innovation du MESRI.

« Le CoSO sera composé d'experts de toutes professions et disciplines concernés par la Science ouverte, de chercheurs, de professionnels de l'information scientifique... Il aura pour mission de proposer des orientations et d'instruire des sujets sur les questions de science ouverte, d'impulser et d'accompagner les actions associées, dans une structure fluide facilitant l'expression et la remontée des idées, suggestions et contributions aux différents groupes de travail ».

Ce comité est organisé autour de quatre Collèges :

- Un collège Publications
- Un collège Données de la recherche
- Un collège Compétences et formation
- Un collège Europe et international

PIN (Pérennisation des Informations Numériques)

Le groupe de travail PIN a été mis en place en 2000 au sein de l'association Aristote. Le CINES est copilote de ce groupe qui constitue un lieu d'échange de savoirs, de savoir-faire et d'actions dans le domaine de la pérennisation des informations représentées sous forme numérique. Il se préoccupe également de l'accès aux informations qui ont été pérennisées.

Dans ce cadre, le CINES **pilote un cycle de conférences** plénières consacré à la pérennisation des données de la recherche en France et en Europe. Initié en 2017, ce cycle devrait se poursuivre jusqu'en 2020.

Enfin, au travers de l'association Aristote, le Centre **co-pilote la formation continue en archivage pérenne** parmi les plus pointues et complètes du marché.



Instances de normalisation et comités scientifiques

Les experts du Centre sont membres de plusieurs groupes de travail ; on citera notamment les commissions nationales CN171 (Applications pour l'archivage et la gestion du cycle de vie du document) et CG 46/CN4 (Interopérabilité technique) de l'**Afnor**, ou encore le comité de pilotage **SEDA** (Standard d'Echange de Données pour l'Archivage).

Dans ces cadres respectifs, les équipes du CINES ont participé à la rédaction de la norme **Afnor** sur l'implémentation d'un système d'archivage (NF Z42-013, utilisée pour l'agrément SIAF) et du SEDA dont la version majeure 2.0 a été publiée en décembre 2015.

Le CINES est également membre des conseils scientifiques de l'**ABES** et de la **TGIR Huma-Num**, membre

du comité scientifique du **forum des archivistes** 2016, ou encore membre des bureaux internationaux DSA et WDS-DSA (devenu CoreTrustseal début 2018).

Formations/Publications

Le département archivage, c'est chaque année une **trentaine d'interventions et de formations**, une ou deux **journées utilisateurs** et cinq à dix **papiers ou articles-presse**.





Le CINES intervient dans de nombreux cursus de formations initiaux et continues, tant sur le plan national qu'international : Enssib (École nationale supérieure des sciences de l'information et des bibliothèques), Ecole nationale des Chartes, Association « Aristote » - PIN, Direction Générale des patrimoines du Ministère de la Culture, Institut national du patrimoine, Université de Namur, Université de Liège, Archives générales du Royaume belge, EUDAT, etc.

Il organise des journées utilisateurs et des journées « prospects » (22-24 juin 2016, 9 juin 2017) afin de sensibiliser la communauté scientifique à la nécessité de l'archivage pérenne et promouvoir l'offre du Centre.

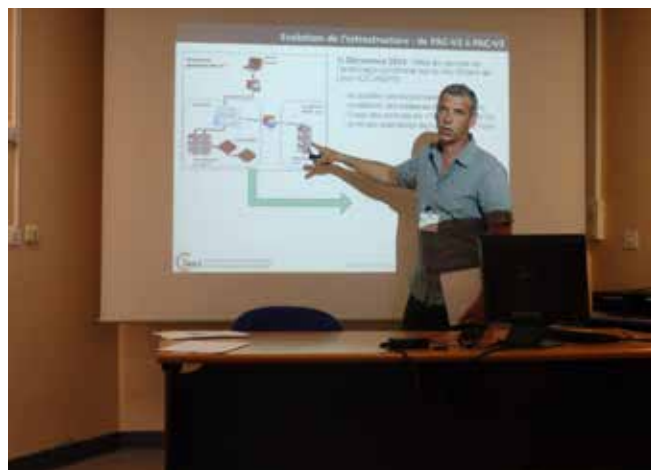
En 2017 l'équipe chargée de l'archivage a publié un **nouveau guide utilisateurs** des services d'archivage du CINES. Ce guide réunit dans un unique document toutes les informations essentielles à connaître pour archiver dans PAC (Plateforme d'Archivage du CINES). Il remplace les spécifications fonctionnelles et techniques de la plateforme jusqu'alors fournies en annexe des conventions d'archivage.

Les principales publications du CINES en 2015, 2016 et 2017 sont :

Pascal Dugenie, Nuno Freire, Daan Broeder, Building new knowledge from distributed scientific corpus ; HERBADROP et EUROPEANA : two concrete case studies for exploring big archival data, *International IEEE workshop on Computational Archival Science* (Boston, USA, 13 décembre 2017)

Cristinel Diaconu, Marion Massol, Préservation et archivage, *Les big data à découvert*, CNRS éditions, Février 2017

Pascal Dugenie, Preservation of natural and cultural heritage Herbadrop & Europeana case studies, *International conference on Digital Infrastructures for Research "DIAR-2017"* (Brussels, Belgium, November, 30th, 2017)



Marion Massol, Pérennisation à long terme : les services mutualisés du CINES, *Journées réseaux de l'enseignement supérieur et de la recherche "JRES2017"* (Nantes, France, du 14 au 17 novembre 2017)

Alexia de Casanove, Sélectionner les données pour l'archivage : réflexions et pratiques au sein de la plate-forme PAC du CINES, *La Gazette des archives*, N°243, 2016

Marion Massol, Nicolas Larrousse, La préservation à long terme des données de la recherche en Sciences Humaines et Sociales : un retour d'expérience, *Journées réseaux de l'enseignement supérieur et de la recherche "JRES2015"* (Montpellier, France, du 8 au 11 décembre 2015).

Présentation du CINES et de son offre d'archivage à moyen et long termes actes des journées de l'association des DSI de l'ESR.

RECONNAISSANCE NATIONALE ET INTERNATIONALE

Le CINES est le centre archiveur du MESRI agréé par le Service Interministériel des Archives de France. Cet agrément fait l'objet d'un audit tous les 3 ans pour statuer sur son renouvellement ; il nécessite la validation du plan de gestion des risques et constitue une garantie de la qualité des environnements et des processus mis en place par le CINES pour la conservation à long terme de données. Initialement obtenu fin 2010, **l'agrément du CINES a été renouvelé en 2014, puis en janvier 2017 pour une période de trois ans.**

Le Centre bénéficie également d'une reconnaissance internationale pour la qualité de son service d'archivage à travers l'obtention et le renouvellement du **label DSA (Data Seal of Approval)**, une accréditation inscrite par la Commission européenne dans le « framework » pour l'audit et la certification des services d'archives), également renouvelé depuis 2010.

Par ailleurs **le CINES est membre des bureaux internationaux du DSA et du CoreTrustSeal** (en partenariat avec l'ISCU). À ce titre, il audite les services d'archives qui candidatent à l'obtention de cette accréditation (BnF, université d'Oxford, établissements de recherche et bibliothèques universitaires, infrastructures européennes de recherche, etc.) et a **participé à la rédaction du nouveau référentiel reconnu par la Commission européenne et par la RDA** (Research Data Alliance). Le test de ce nouveau référentiel a été réalisé en 2016 en partenariat avec le CCSD (HAL).

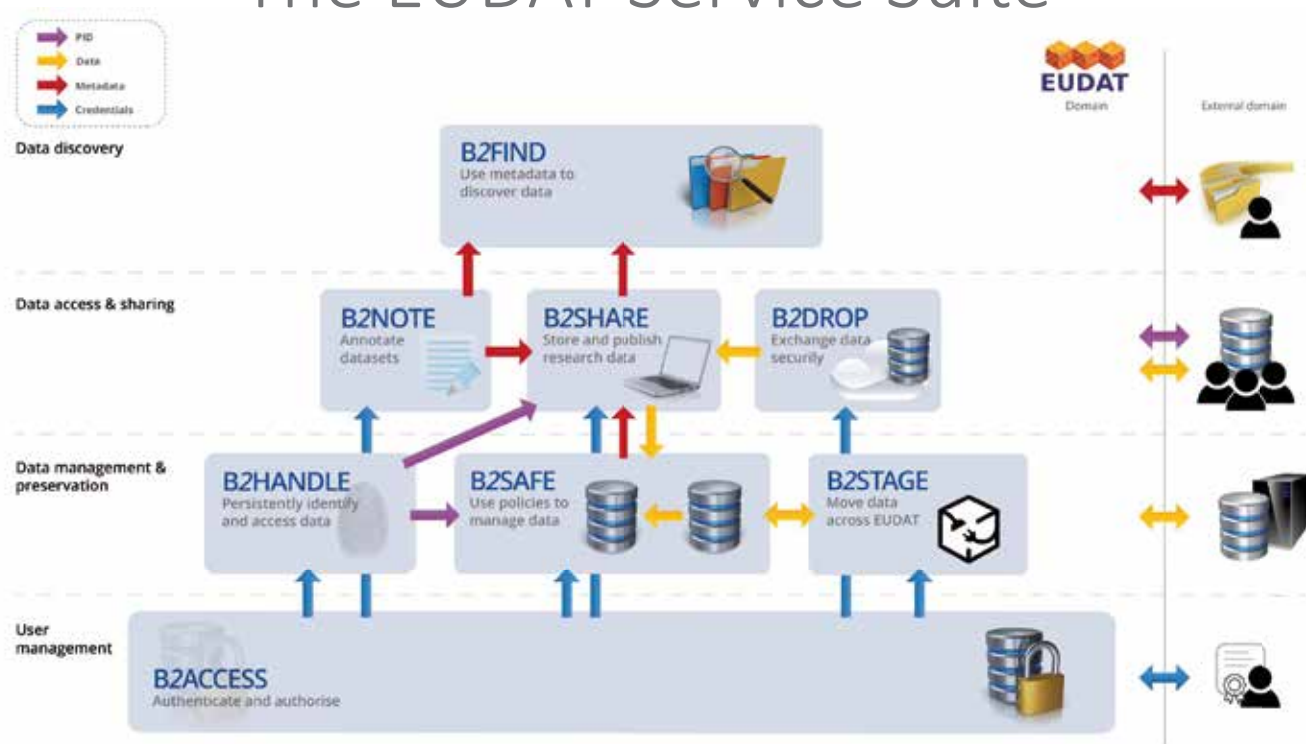
L'archivage au CINES et l'Europe

EUDAT 2020

Eudat 2020, signé en 2015 pour une durée de 3 ans, est la poursuite du projet Eudat, il regroupe 33 partenaires pour un budget de 19 M€, avec pour objectif d'assurer le **déploiement en production de l'infrastructure CDI (Common Data Infrastructure)**. Le CINES est plus particulièrement impliqué dans les actions suivantes :

- « WP2 - Policy and sustainability ». L'objectif de ce « work package » est de maintenir et renforcer les connexions entre centres au cœur du CDI du point de vue de la politique, de la stratégie et de la gouvernance. Dans le cadre de ce projet Horizon2020, **le CINES pilote au niveau européen le catalogue de services** de l'infrastructure européenne EUDAT CDI.
- « WP 5 - Service building ». Ce « work package » vise à consolider l'architecture et développer les services de l'infrastructure EUDAT CDI.
- « WP 6 - Operations ». Ce « work package » se préoccupe d'assurer la **conservation à long terme**, l'authenticité, l'intégrité et la confidentialité des données.
- « WP 7 - Cross infrastructure services ». Ce « work package » travaille sur l'interopérabilité entre Eudat et les autres e-infrastructures : PRACE, EGI,...

The EUDAT Service Suite



Développé dans le cadre d'EUDAT2020 le logiciel GEF (Generic Execution Framework) permet l'exécution d'outils logiciels conteneurisés sur des données stockées dans l'infrastructure CDI ; il utilise la technologie Docker avec des images spécifiquement annotés. *"By deploying these GEF services to Dockerized computational resources as close to the data as possible, data transfers can be minimized. GEF services can also be reused to reproduce results or toolchains. Data inputs to GEF services can be specified via PID or URL before they are automatically transferred to the processing location".*

Dans le cadre de son **implication dans les activités de certification au sein d'EUDAT2020**, le CINES a notamment accueilli en novembre 2015 un workshop dont l'objectif était l'évaluation et l'accompagnement des établissements de l'e-infrastructure EUDAT dans la mise en œuvre de standards d'archivage pérenne.

EUDAT CDI

Fin 2017, plus de 20 des principales organisations de recherche européenne, centres de calcul et autres centres de données ont signé un accord de consortium afin de soutenir l'e-infrastructure collaborative paneuropéenne EUDAT au moins pour les 10 prochaines années. La EUDAT Collaborative Data Infrastructure (ou EUDAT CDI) est ainsi née le 1^{er} janvier 2018.

CDI members



Le CINES, un des fondateurs de l'e-infrastructure paneuropéenne EUDAT pour le soutien des communautés scientifiques dans la gestion des données de la recherche



Workshop Eudat 2020 organisé au CINES en 2015

EOSC Hub (European Open Science Cloud)

Le projet « EOSC-Hub » (01-2018/12-2020), composé de 74 partenaires et doté de 30 M€, réalise la convergence des projets EUDAT, EGI et INDIGO. En effet, son catalogue de services se base sur ceux de la fédération EGI, du consortium EUDAT et sur les services développés au travers du dispositif européen INDIGO-DataCloud ; les services thématiques sont, quant à eux, fournis par les communautés de recherche.

Le CINES contribue au projet en apportant son expérience dans la mise à disposition des services d'un centre national ayant un rôle de dépôt de données, avec un focus particulier sur les techniques et processus de conservation à long terme et sur les bonnes pratiques de gestion de ces services. Le CINES est **responsable du déploiement du service d'archivage à moyen et long terme sur l'e-infrastructure**, concrétisant par là son expérience dans l'exploitation de sa plateforme (PAC) et dans l'administration de l'outil B2SAFE d'EUDAT depuis 2010. Cet outil est associé à B2HANDLE, un autre outil EUDAT qui attribue un identifiant pérenne à chaque donnée stockée dans B2SAFE, et est utilisé notamment pour des données du CERFACS.

Pilote EUDAT/EOSC-pilot DPHEP

La validation des outils EUDAT dans le cadre d'EOSC hub est un enjeu majeur pour le consortium. Un pilote a donc été envisagé avec la collaboration du **CERN** (DPHEP EOSC-pilot) visant à utiliser l'infrastructure EUDAT entre les deux nœuds **CINES** et **CINECA** pour la **préservation et la diffusion de données volumineuses provenant du LHC** (Large Hadron Collider). L'objectif était d'utiliser des services génériques (i.e. non-spécifiques à cette communauté scientifique) combinés de manière simple et transparente (ex : pour l'attribution d'identifiants pérennes ou PID) afin de construire un système capable de stocker et de préserver à long terme de grands volumes de données en « open access » (100To ou plus).

ICEDIG ("Innovation and Consolidation for large Digitisation of natural heritage")

ICEDIG est un projet européen financé dans le cadre du programme Horizon2020. Il constitue la phase d'implémentation de la nouvelle infrastructure de recherche pan-européenne **DiSSCo** ("Distributed System of Scientific Collections"). Cette e-infrastructure est une importante initiative pour la **gestion distribuée des collections européennes en sciences naturelles**. Le projet ICEDIG vise donc à concevoir et traiter les aspects tant politiques, financiers, de gouvernance que les questions techniques de DiSSCo.

Le CINES est l'un des 2 partenaires français d'ICEDIG, au côté du Muséum National d'Histoire Naturelle. Après avoir généralisé les développements faits dans le cadre du projet **Herbadrop** (cf ci-dessous), le CINES intervient essentiellement sur l'évaluation et la validation des services de l'e-infrastructure EUDAT CDI (en particulier le service à forte valeur ajoutée ETDR pour l'archivage pérenne de données scientifiques et le service B2SAFE). L'objectif est, en effet, d'évaluer l'adéquation entre les services génériques d'EUDAT CDI et les besoins « métier » de la communauté de recherche DiSSCo.

Autres projets internationaux sur les données de recherche

En dehors des projets **Herbadrop** (voir ci-dessous) et EOSC-pilot DPHEP, le CINES a participé à des initiatives internationales visant à offrir des services de gestion de données de la recherche innovants en particulier dans le cadre d'**ESGF** (Earth System Grid Federation), une infrastructure internationale pour la distribution et l'analyse des données en climatologie.

FOCUS SUR QUELQUES PROJETS STRATÉGIQUES



HERBADROP (Herbiers numérisés)

Le projet HERBADROP, initié en 2015 avec le **Museum national d'Histoire naturelle** et piloté par celui-ci, s'intègre dans un programme plus vaste : **Recolnat**,

dont une des phases concerne les herbiers numérisés, soit environ 400 To de données (soit un espace de **1.2 Po** une fois archivés). Accepté et soutenu dans le cadre d'EUDAT pour la mise en œuvre d'un premier prototype s'appuyant sur l'infrastructure CDI d'EUDAT, le projet HERBADROP compte pour partenaires, en sus du CINES, 6 établissements-pilotes d'envergure nationale et qui sont représentatifs de la communauté européenne des sciences naturelles :

- MNHN (Museum national d'Histoire naturelle, Paris)
- RGBE (Royal Botanic Garden Edinburgh, Ecosse)
- BGBM (Botanischer Garten und Botanisches Museum Berlin, Allemagne)
- Naturalis Biodiversity Center (Leiden, The Netherlands)
- Jardin botanique de Meise (Belgique)
- Digitarium (Service Centre for HP digitisation in Finland, Helsinki).

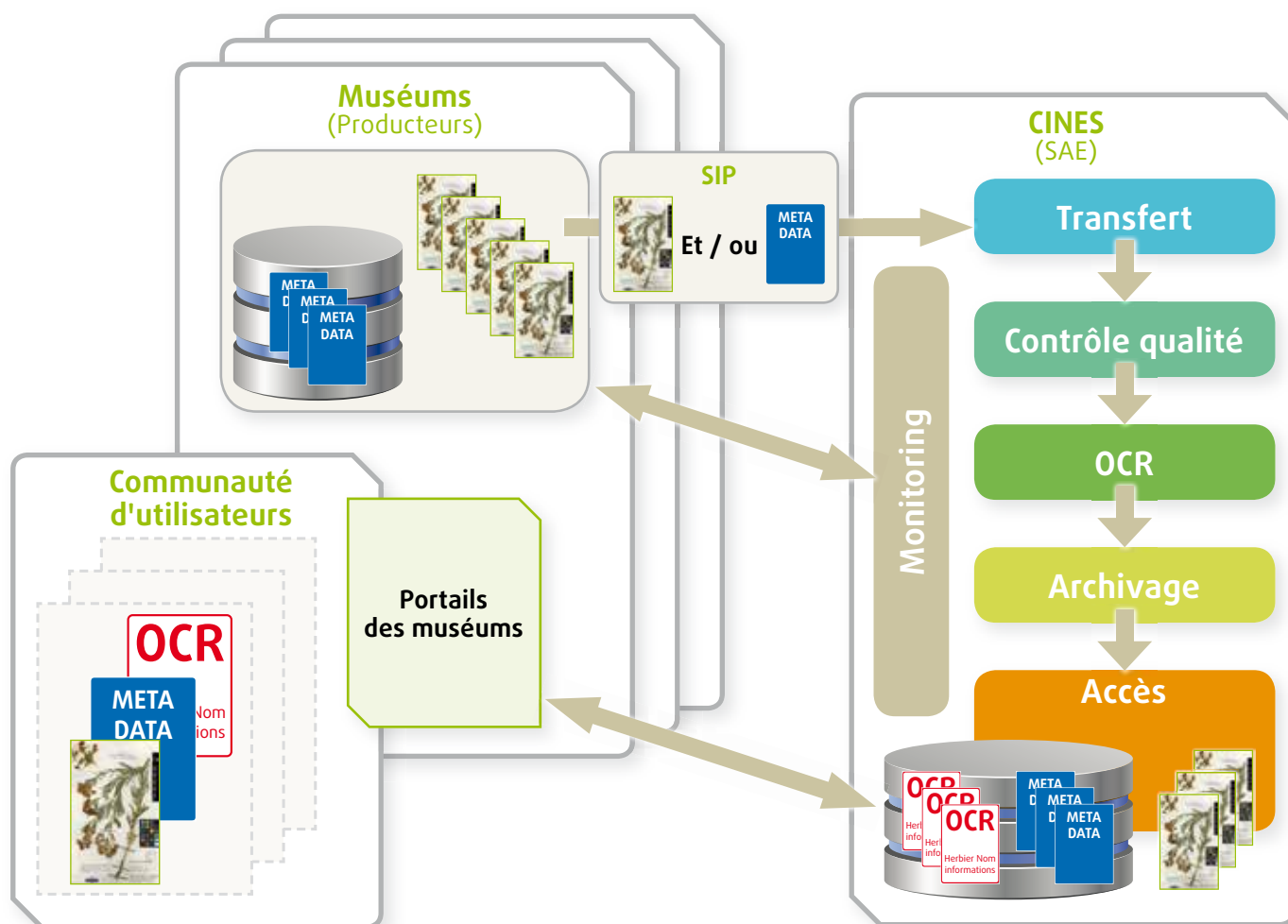
L'objectif est la **conservation à long terme des herbiers numérisés**, 22 millions de spécimens en Allemagne, 10 millions en France...

Au-delà de la volumétrie traitée, l'intérêt du projet réside dans le **couplage de l'archivage et du calcul de haute performance, qui peut être généralisé à d'autres thématiques que les herbiers** (textes ou manuscrits anciens numérisés par exemple). En effet un certain nombre de métadonnées sont incluses dans les images et nécessitent l'utilisation d'un OCR (dispositif de reconnaissance automatisé de caractères) afin de les extraire et/ou de dispositifs d'intelligence artificielle spécifiques. Or, la puissance de calcul requise pour effectuer ces traitements est hors de portée des établissements partenaires. Cette partie (OCR et IA) est donc déportée automatiquement sur les dispositifs de calcul de haute performance du CINES. Les outils d'EUDAT sont alors interfacés avec la plateforme d'archivage du CINES et avec un moteur de recherche puissant de plus en plus utilisé pour indexer les « big data ».

Le prototype a été validé fin 2017 et mis en production début 2018.

Considéré comme un des projets-phares du projet EUDAT2020 et de l'e-infrastructure pérenne EUDAT CDI, la qualité des résultats d'HERBADROP participe grandement, en dehors de l'attribution à la France de financements européens supplémentaires (+177 % du budget initialement alloué aux partenaires français), au rayonnement national et international du CINES et à la valorisation de son offre de services.

Enfin, le dispositif technique et organisationnel mis en place constitue **le socle générique de la prochaine infrastructure nationale d'archivage pérenne du CINES (PAC-V4)**. Il participe au positionnement central de la France au sein de l'infrastructure paneuropéenne de recherche DiSSCo.



Le dispositif Herbadrop du CINES, une solution innovante entre archivage pérenne, calcul intensif, intelligence artificielle et diffusion de données.

« Software Heritage »

Soutenu par l'UNESCO et porté par l'INRIA, le projet a pour objectif d'archiver les codes sources pour les générations futures, considérant que les logiciels font partie du patrimoine scientifique.

« Parce qu'ils sont le support indissociable des connaissances techniques et scientifiques de l'humanité tout entière et que nous ne pouvons pas prendre le risque de les perdre, les logiciels doivent être **collectés** et **préservés**. Nous avons fait le choix de les sauvegarder tous sous leur forme fondamentale : leur code source. » Le CINES est associé au projet en tant que centre archiver et fournisseur d'expertise en pérennisation.

Démarche générique d'archivage des données environnementales

Dans le domaine des géosciences et sciences de l'environnement, des « pôles nationaux de données et de services » ont été créés. Ils ont pour vocation la fourniture d'informations et de services sur les observations

réalisés par les laboratoires et les observatoires.

Un schéma directeur a été proposé en 2014 pour ces pôles qui sont confrontés, pour la gestion des données, au défi apporté par la forte croissance des données générées, la diversité des formats et la nécessité de conservation de tout ou partie comme partie intégrante du cycle de vie de la donnée. Au-delà des traitements et de la diffusion, le problème de l'archivage à long terme nécessite une prise en compte particulière pour sélectionner les « bonnes données » à conserver, les méthodes et processus à utiliser, les outils informatiques à mobiliser et le modèle économique basé sur une étude des coûts complets.

Suite à une rencontre avec le directeur adjoint du domaine « Surfaces et interfaces continentales » (CNRS/INSU), faisant part de sa volonté de sensibiliser les pôles rattachés à l'Infrastructure de Recherche « Observation de la Terre (ou système Terre) » (regroupement des pôles Theia, AERIS, Form@ter, OCEAN, ...) à la conservation à long terme des données, le CINES, statutairement mandaté pour offrir nationalement ce type de services à la communauté scientifique, a proposé de travailler à la **mise en place d'une démarche générique du processus de pérennisation des données**. La proposition a

été approuvée par les membres du Comité d'orientation stratégique et le Conseil d'administration du CINES. La démarche comporte des actions de : sensibilisation, information, formations, études techniques, prototypes, mises en production, définition de modèle économique, etc. Elle sera élaborée de façon itérative en collaboration avec les partenaires, validée pour cette thématique scientifique et par la suite reproductible sur d'autres pôles et d'autres domaines scientifiques, moyennant quelques adaptations liées aux spécificités de chaque cas d'utilisation.

Le CINES via GEOSUD est partenaire du pôle THEIA, créé en 2012, qui s'intéresse aux surfaces continentales. Dans ce cadre, le CINES archive déjà les données satellitaires

produites par GEOSUD et pourra donc s'appuyer sur cette expérience pour travailler avec les partenaires à **définir la démarche la plus efficace et enclencher pour les données environnementales un processus d'archivage généralisé.**

La première étape, en cours, a consisté à étudier avec les pôles la stratégie de gestion de leurs données notamment sur le long terme, les méthodes de sélection et de documentation des données à conserver et les coûts complets engendrés suivant les solutions adoptées. Des réunions de travail ont eu lieu en 2017 et une première ébauche méthodologique avancée.

Zoom sur... géosud



L'archivage au CINES des **images satellitaires en haute définition** de l'Equipex GEOSUD (Infrastructure de Recherche « système Terre » - pôle de données et de services « THEIA ») est entré en production le 23 novembre 2017.

Cela a nécessité de nombreux travaux d'adaptation de la plateforme d'archivage du CINES en particulier en ce qui concerne la validation des images au format GeoTIFF et celle des fichiers de métadonnées métiers associées aux images. Les premières images archivées sont l'historique des couvertures nationales SPOT6/7 de 2013 à 2016 en mode « image brute ». Les travaux de mise en géométrie réalisés par l'IGN permettront ensuite d'archiver la version orthorectifiée, conforme RGE (Référentiel à Grande Echelle), de ces mêmes images.

Les couvertures nationales seront archivées tous les ans, dans leurs deux versions (i.e. brute et orthorectifiée), au fur et à mesure de leur intégration dans l'IDS GEOSUD, ce qui représentera d'ici 5 ans plus de 110 To de données d'imagerie satellitaire archivées au CINES pour ce projet. La conservation pérenne de ces images satellites dans leur version brute et orthorectifiée est **une première au niveau national**. Elle permet de garantir à

la communauté scientifique et aux acteurs publics leur réutilisation dans les décennies à venir. En effet, cette banque d'images en haute définition constitue un **riche ensemble de clichés de la France (métropole + DOM-TOM) régulièrement pris depuis 2005.**

L'archivage pérenne des données produites a été envisagé dès la création du projet compte tenu des coûts d'acquisition et de traitements nécessaires à l'obtention d'images de qualité, exploitables à des fins scientifiques ou opérationnelles. Par ailleurs, ces données sont amenées à être réutilisées sur plusieurs décennies car les études environnementales nécessitent souvent de **remobiliser des images anciennes afin d'évaluer les changements...**

Une **réflexion conjointe entre l'infrastructure de recherche « système Terre » et le CINES a été initiée en 2017 sur les données obtenues par l'observation in situ** et associées à ces images satellitaires. Un des principaux objectifs de ce projet consiste à mettre au point une méthodologie de travail innovante pour une **meilleure prise en charge de ces données de la recherche** tout au long de leur cycle de vie (production, publication, gestion, pérennisation, etc.).

HÉBERGEMENT

› Une mission à part entière	41
› Offre de haut niveau	41
› Hébergés nationaux et locaux	41
› Projet de data center	41

SERVICES INFORMATIQUES TRANSVERSES

› Serveurs	42
› Capacité de stockage	42
› Sauvegardes	42
› Réseaux	43
› Métrologie	43
› Sécurité	43
› Informatique et libertés	43
› Supervision et astreintes	43

UNE MISSION À PART ENTIÈRE

Héberger des plates-formes informatiques de partenaires appartenant à la communauté de l'Enseignement Supérieur et de la Recherche est une activité « historique » du CINES dans le cadre de la mutualisation d'infrastructures et des économies financières et énergétiques associées. Cette activité est devenue une mission dans les nouveaux statuts de 2014, sous condition de ne pas contrecarrer les évolutions des deux autres missions relatives au calcul et à l'archivage.

OFFRE DE HAUT NIVEAU

En tant que centre national le CINES a participé au groupe de travail sur les datacenters mis en place dans le cadre d'INFRANUM. Répondant aux critères élaborés par INFRANUM, sur les bases des recommandations de la DINSIC, le CINES est apte à accueillir dans ses salles machines avec un niveau d'infrastructure Tier3+, c'est-à-dire hautement sécurisées 24 h/24 7j/7, tout type de plates-formes informatiques. Le CINES s'appuie sur une longue expérience de plus de 20 ans d'hébergement de plates-formes sensibles et sur une offre de services au plus près des besoins des hébergés (gestes de proximité, accompagnement et gestion en mode projet, ...). Pour assurer cette offre le CINES a dédié à la mission hébergement sa salle machine N°3 dont il a augmenté les capacités électriques et de climatisation (ajout de 215 kW en juillet 2016). Afin de maîtriser l'urbanisation et la consommation d'énergie de cette salle machine le CINES fournit les baies sécurisées par digi-code et assure le confinement des allées froides.

HÉBERGÉS NATIONAUX ET LOCAUX

Sont visées en priorité les plates-formes à caractère national. À cet égard le CINES héberge depuis plusieurs années le supercalculateur de GENCI, les serveurs de l'ABES (Agence Bibliographique de l'Enseignement Supérieur), ceux de la DSI de l'INSERM et le cluster de l'IRSTEA (depuis 2016), ainsi que le nœud Renater. Le centre a également hébergé entre 2013 et le début 2017 l'application FUN MOOC.

Cela n'exclut pas la possibilité d'héberger des plates-formes locales ou régionales. Ainsi le CINES héberge les nœuds du réseau régional et du réseau métropolitain, le calculateur du méso-centre Meso@LR (300 nœuds de calcul pour 300 Tflops) et depuis l'été 2017, les DSI du rectorat de Montpellier et de l'université de Montpellier. Au total une centaine de baies sont hébergées pour des organismes extérieurs en sus des équipements propres du CINES.

L'ensemble des hébergés bénéficie d'atouts liés à une mutualisation des infrastructures électriques, de climatisation, de sécurité du site et de la proximité et de la disponibilité des nœuds des réseaux nationaux et régionaux installés au CINES.



Baies de l'Université de Montpellier

PROJET DE DATA CENTER

L'existence en 2014 d'un projet H2M, piloté par RENATER, de construction d'un data-center ex-nihilo sur Montpellier, résultant d'une volonté de mutualisation d'infrastructures par des partenaires ayant des sites locaux (Rectorat, Universités de Montpellier I, II, III, CROUS, Ecole Nationale Supérieure de Chimie Montpellier, CIRAD, IRD) a conduit le CINES à proposer une solution alternative basée sur la stratégie suivante :

L'utilisation dans un premier temps de la place disponible dans sa salle machine N°3 dédiée à l'hébergement, ce dont ont profité le Rectorat et l'Université en 2017.

Dans un second temps si la demande, ouverte à de nouveaux partenaires locaux et nationaux hors du périmètre initial, dépasse la capacité d'accueil, d'envisager la possibilité de construction d'une nouvelle salle machine. Solution qui, en s'appuyant sur une mutualisation d'infrastructures existantes, serait moins onéreuse que la construction ou l'aménagement d'un nouveau data-center

La mise en place de la nouvelle région OCCITANIE a réorienté la réflexion vers le projet d'un data-center régional éventuellement réparti sur 2 sites : Toulousain et Montpelliérain.



Calculateur du mésocentre Meso@LR hébergé au CINES

SERVICES INFORMATIQUES TRANSVERSES

SERVEURS

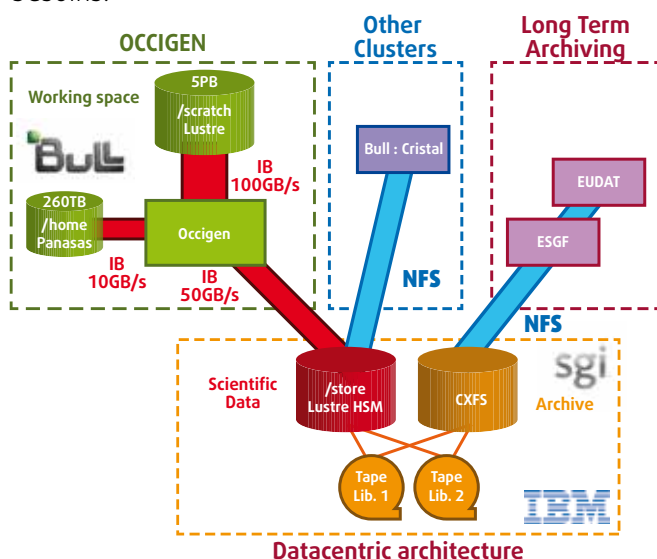
L'administration des systèmes (en majorité Linux) des serveurs du CINES à l'exception de ceux des supercalculateurs est effectuée par le DS2I. Toutes les plates-formes « sensibles » qui gèrent des données ainsi que les **espaces de stockage sont sécurisés par une double alimentation de bout en bout.**

La multiplication du nombre de serveurs virtualisés entre 2015 et 2017 a nécessité l'augmentation de la plate-forme de virtualisation en termes de puissance et de capacité de stockage associée.

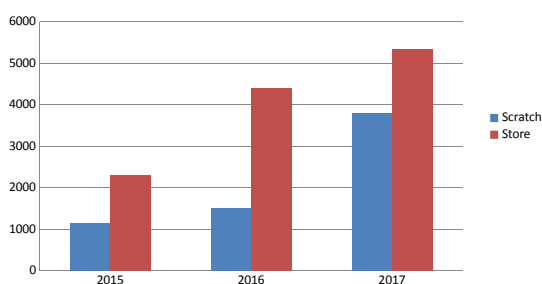
CAPACITÉ DE STOCKAGE

L'augmentation du nombre de données traitées et générées par les supercalculateurs, le volume de « big data » scientifiques nécessitant une conservation à plus ou moins long terme, conduisent à ajuster en permanence la capacité de stockage sur disques et cartouches du Centre.

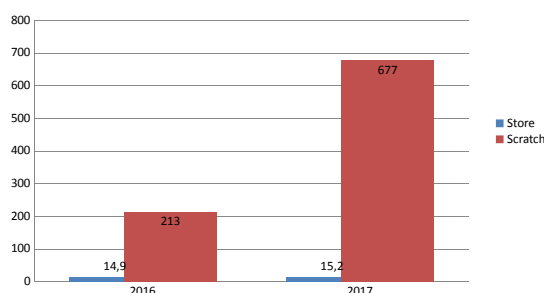
L'architecture « data centrique » adoptée par le CINES permet de répondre avec une certaine souplesse aux besoins.



Occupation maximum journalière en To

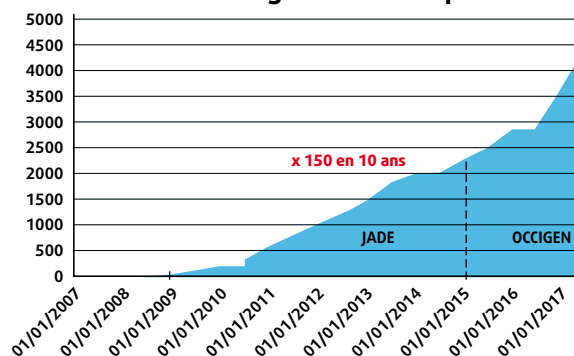


Nombre maximum journalier de fichiers



La volumétrie du Store représente le nombre de To utiles, elle inclue l'espace de niveau 2 sur cartouches.

Évolution des données scientifiques stockage datacentrique



L'archivage « pérenne », conservation à long terme des données, nécessite la prise en compte de contraintes supplémentaires, qui justifie l'agrément du SIAF. Parmi ces contraintes l'obligation d'avoir des **copies multiples réparties dans différents lieux dont un site à distance**, a nécessité la mise en œuvre de processus de copies automatiques des données sur des supports répartis dans quatre salles machines du centre et une copie à l'IN2P3 à Lyon. Les baies hébergeant les serveurs et les disques, sont individuellement protégées par des **digi-codes à l'intérieur de salles machines** elles-mêmes hautement sécurisées.

SAUVEGARDES

Les données sauvegardées bénéficient d'une **double copie sur cartouches** dans deux robotiques installées dans **deux salles machines distinctes**. En cohérence avec l'augmentation des volumes traités, ces robotiques ont bénéficié d'une augmentation de leurs capacités par l'ajout de lecteurs pour cartouches capacitatives Jaguar 4 et LTO 6.



RÉSEAUX

Les réseaux sont en perpétuelle évolution, les volumes de données échangés intra et inter centres tant pour le calcul que pour le transfert des données à archiver, conduit à adapter la capacité des « tuyaux ». Ainsi le CINES a passé à **10 Gb/s sa prise sur le nœud Renater**. **Le réseau CCFR à 10 Gb/s entre les centres nationaux** pour les aspects calcul est aujourd'hui opérationnel permettant aux chercheurs partageant leurs calculs sur deux centres de transférer les données utiles.

Parallèlement le réseau interne du CINES vers les données du supercalculateur a lui aussi évolué vers le haut débit, afin de garantir le débit de bout en bout.

MÉTROLOGIE

Dans un souci d'amélioration de la maîtrise de ses consommations énergétiques le Centre a fait l'acquisition et mis en place en 2017 des compteurs, à la norme MID, permettant la **mesure des consommations électriques de chaque équipement** d'infrastructure et des différentes plates-formes et équipements hébergés. Le suivi détaillé des différents éléments pris en compte permettra de mieux appréhender les gains énergétiques possibles et d'identifier les coûts réels par environnement.

SÉCURITÉ

La sécurité a toujours été une préoccupation majeure du CINES qui nécessite une remise en cause permanente prenant en compte les nouveaux risques inhérents à l'évolution technologique. Ainsi dans le cadre de la protection du potentiel scientifique et technique de l'Etat un certain nombre de mesures concernant la protection des systèmes informatiques ont été adoptées entre 2015 et 2017. Elles ont été élaborées avec le RSSI et le FSD en étroite collaboration avec l'ANSSI (Agence Nationale de la Sécurité des Systèmes d'Information), la DGSI, le HFDS et pour le calcul, avec GENCI et les autres centres nationaux.

Une première étape a consisté à confier à l'**ANSSI l'audit de l'environnement de calcul**. Cet audit a été réalisé entre le 19 et le 23 janvier 2015. La plate-forme de calcul OCCIGEN a ensuite été homologuée en juin 2016 après une étude poussée des risques et la mise en œuvre de réponses adéquates dans le cadre du SMSI (Système de Management de la Sécurité de l'Information).

Parallèlement ont été mis en œuvre :

- les organes de suivi, de conseil et d'audit, en particulier le **comité opérationnel de sécurité (COSec)** composé

du Directeur (AQSSI), du FSD, du RSSI, du responsable sûreté du site, des chefs de département et de leur correspondant sécurité,

- la conformité de l'établissement avec la PSSIe (Politique de sécurité des systèmes d'information de l'état), avec notamment
 - le passage à des mots de passe robustes d'au moins 12 caractères,
 - le filtrage des adresses IP,
 - l'étude de la possibilité d'adopter une procédure d'authentification forte,
 - des processus communs de validation des ouvertures de compte entre les centres nationaux,
 - le découplage du réseau des postes d'administration des systèmes, des autres réseaux internes,
 - le chiffrement des échanges hautement confidentiels.

INFORMATIQUE ET LIBERTÉS

Le CINES avait l'obligation de créer officiellement la fonction de CIL (Correspondant Informatique et Libertés) qui était auparavant partiellement assurée par un chef de département. La nomination d'un agent sur cette fonction en 2016 a permis de réaliser un état des lieux avec pour objectif la préparation de la mise en œuvre du **Règlement Général sur la Protection des Données (RGPD)** adopté en avril 2016, et **applicable en mai 2018**. Ce règlement vise le renforcement des droits des personnes sur leurs données personnelles. Tous les traitements (passés et futurs) devront être conformes aux dispositions du règlement. Celui-ci pointe la responsabilisation des acteurs traitant les données (y compris l'hébergeur) et participe à la crédibilisation de la régulation via une coopération entre autorités de protection des données et une augmentation du niveau de sanction.

SUPERVISION ET ASTREINTES

En journée l'ensemble des équipements est sous surveillance et maintenance par les équipes du centre. Un service SVP est disponible auprès des utilisateurs pour signaler tout problème qui est alors répertorié dans un ticket d'incident. Le problème est alors soit directement traité par la cellule SVP soit transmis vers le support spécialisé.

Le CINES offre un **service de surveillance des infrastructures 24 h/24 7j/7** grâce à la présence d'un gardien de nuit et à une astreinte des équipes du DS2I pour les nuits et les week-end.

RESSOURCES HUMAINES

› Postes et effectifs	45
› Promotions	45
› Hygiène et sécurité	45
› Mutualisation de la gestion des pensions (Adhésion à PETREL)	45

GESTION FINANCIÈRE ET ADMINISTRATION

› Missions et déplacements des agents	47
› Finances	47
› Dématérialisation	48

POSTES ET EFFECTIFS

Une constatation s'impose à la lecture des effectifs, la proportion des personnels temporaires est trop importante par rapport aux personnels titulaires pour un centre de production comme le CINES. D'autant que la loi stipule l'impossibilité de confier à des personnels temporaires des tâches de fond, stratégiques pour le centre, ce qui par ailleurs n'est pas souhaitable pour des raisons de pérennité et de sécurité. Parallèlement ces tâches de fond augmentent en nombre et en intensité, sur la sécurité ou l'importance des infrastructures par exemple.

Effectifs disponibles au CINES

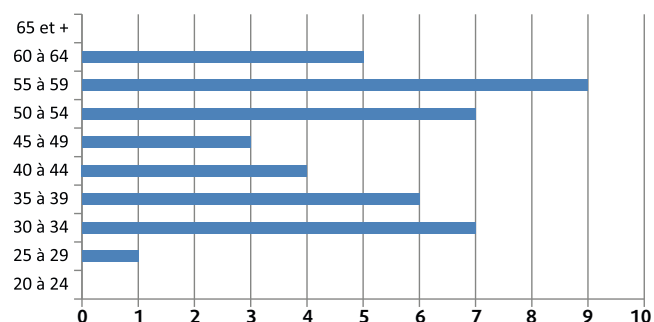
	2014	2015	2016	2017
Titulaires	45	43	43	43
CDD ou apprentissage	8	5	6	6
Régies	7	9	14	12

En 2016, afin de pallier l'absence de longue durée de 2 agents, pour raisons de santé, cela a conduit le CINES à faire appel à deux personnels en régie pour assurer les tâches stratégiques correspondantes.

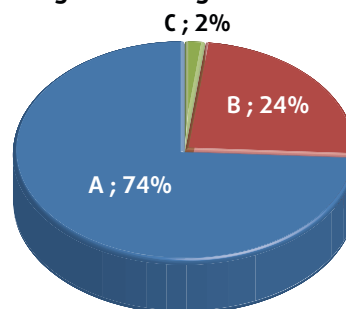
Il faut ajouter aux effectifs du tableau les personnes des sociétés de services chargées des tâches externalisées : accueil, gardiennage, ménage.

Le renforcement des effectifs de titulaires fait l'objet d'une demande récurrente du directeur du CINES auprès du Conseil d'Administration du CINES, de la DGESIP et de la DGRI.

Pyramide des âges des titulaires au 1^{er} janvier 2018



Catégories des agents titulaires

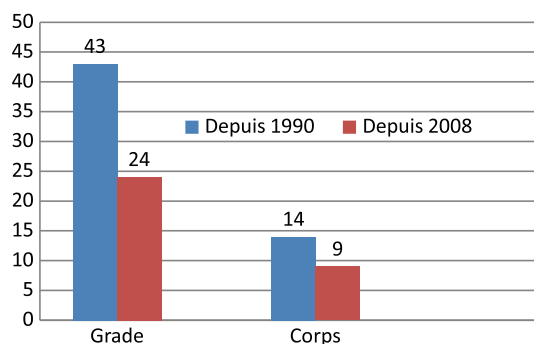


30 agents titulaires hommes et 13 femmes

PROMOTIONS

Le taux de promotions est satisfaisant pour un établissement comme le CINES, compte tenu des effectifs et de la structure des emplois.

Promotions : état en 2017



Entre 2015 et 2017, 9 agents ont bénéficié d'un changement de grade et 1 agent a changé de corps.

HYGIÈNE ET SÉCURITÉ

Une attention particulière a été portée sur les Risques Psycho- Sociaux. Dans un premier temps un audit externe a été réalisé en 2016 qui a débouché sur un certain nombre de préconisations. Un comité de pilotage intégrant le CHSCT et les responsables de département a été mis en place et une démarche définie pour prendre en compte les préconisations de l'audit et définir les indicateurs de suivi.

Des formations des personnels ont été programmées par le conseiller de prévention concernant les extinctions incendie, le secourisme, les habilitations électriques et des exercices d'évacuation régulièrement organisés. Des visites mensuelles d'inspection des locaux ont été également réalisées.

Une convention a été signée avec la médecine du travail de l'Université de Montpellier pour le suivi des personnels du CINES.

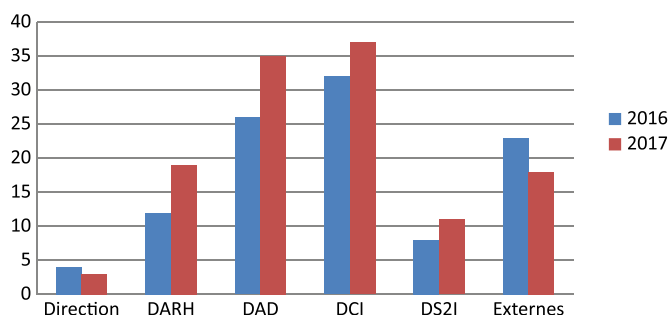
MUTUALISATION DE LA GESTION DES PENSIONS (ADHÉSION À PETREL)

Le CINES a adhéré au service inter établissements mutualisé de gestion des pensions, PETREL, qui est opérationnel au 1^{er} janvier 2018 et dont les missions sont :

- d'informer et de former les gestionnaires RH des établissements partenaires sur la réglementation et son évolution en matière de pension pour permettre une information locale de 1^{er} niveau des usagers,
- d'informer et de conseiller les agents sur leurs droits à pension,
- d'enregistrer les données liées à la retraite dans le logiciel Pension puis à terme de vérifier les informations accessibles dans le portail PETREL, au moment de la demande de départ en retraite et de réaliser des simulations financières permettant l'aide à la décision des agents,
- d'instruire les demandes de pension et de constituer les dossiers de pension,
- de constituer les dossiers d'Estimation Indicative Globale (EIG) dans le logiciel Pension puis à terme dans le portail PETREL,
- de traiter les fiches de liaison inter-régimes CARSAT (Caisse d'Assurance Retraite et de la Santé Au Travail),
- d'être le correspondant unique du Service des Retraites de l'Etat (SRE) en tant qu'interface entre le SRE et les services de gestion RH des établissements partenaires notamment en matière de complétude des Comptes Individuels Retraites, de disponibilité des pièces justificatives et de qualité des données saisies dans les SIRH.

MISSIONS ET DÉPLACEMENTS DES AGENTS

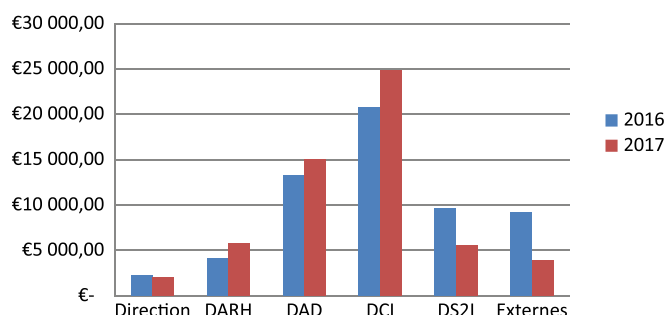
Nombre de missions



Les missions "Externes" correspondent notamment à la prise en charge des déplacements :

- des représentants aux différents conseils et comités du centre ;

Coût des missions

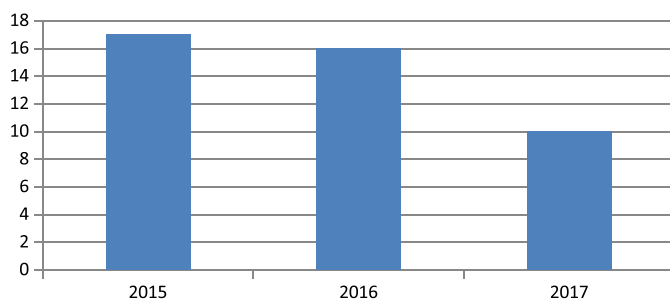


- des intervenants externes pour les formations et séminaires organisés par le CINES.

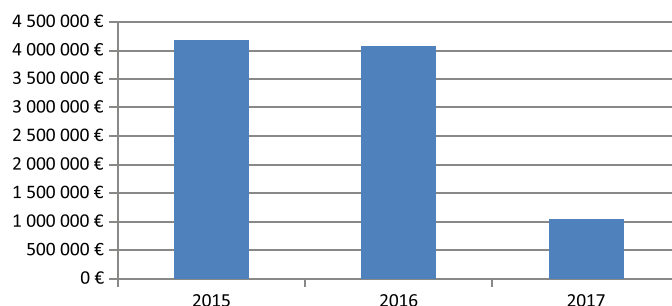
FINANCES

Marchés

Nombre de marchés



Total TTC des marchés



Mise en place de la GBCP

La « réforme GBCP » (Gestion Budgétaire et Comptable Publique) porte sur une révision en profondeur du cadre général qui structure la gestion budgétaire et comptable de l'État et de ses opérateurs.

C'est au 1^{er} janvier 2016 que se sont appliquées les dispositions ayant le plus lourd impact sur le cadre budgétaire et comptable des établissements. La GBCP a entraîné un bouleversement de l'organisation interne des finances des établissements publics.

La GBCP permet une gestion prévisionnelle du budget plus fine et fondée sur un ensemble d'alerte et d'indicateur qui a pour but d'éviter une obstruction à la consommation des Crédits de Paiements (CP). L'introduction de la notion d'Autorisation d'Engagement (AE) permet de retracer de manière pluriannuelle les engagements

pris sur l'ensemble des dépenses. La pluri annualité est ainsi confirmée comme axe essentiel de préparation et d'exécution du budget, ce qui permet le croisement entre la stratégie exprimée dans le projet d'établissement et la programmation budgétaire.

Le CINES a été un des sites pilotes pour le déploiement de la GBCP et il fait partie de la vague 1 des établissements utilisant le logiciel comptable WIN M9 de GFI. Depuis le 1^{er} janvier 2016, il utilise le logiciel comptable en mode GBCP. En 2017, le traitement logiciel de la GBCP a intégré :

- La certification du service fait,
- L'Activation des dépenses par origines,
- La modification de la fiche d'acquisition pour répondre aux nouvelles normes budgétaires et comptables,
- Le suivi mensuel à l'aide de tableaux de bord de la consommation d'AE et de CP.
- Le suivi individuel des marchés en cours

Contrôle de gestion

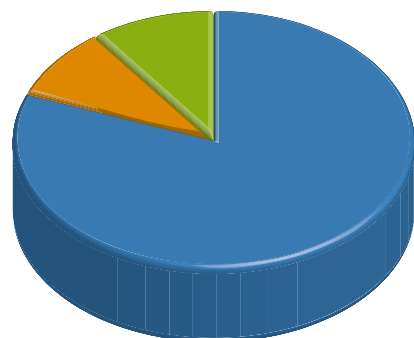
Il se compose du contrôle interne budgétaire et du contrôle interne comptable (CIB et CIC).

Le contrôle interne comptable recouvre l'ensemble des dispositifs organisés, formalisés et permanents mis en œuvre pour maîtriser le fonctionnement des activités financières et donner ainsi une assurance raisonnable sur la qualité des comptes, c'est-à-dire, leur fidélité à la réalité économique, patrimoniale et financière. Le contrôle interne intègre un audit interne comptable et financier permettant d'évaluer périodiquement l'efficacité du dispositif de contrôle interne. Il repose sur une démarche permanente et itérative dans le sens où, pour maîtriser les risques, il doit être constamment adapté aux évolutions de l'environnement de travail (normatives, informatiques, organisationnelles...) et réajusté en fonction de l'évaluation des risques par nature mouvants. Il s'inscrit dans un cadre normatif : le référentiel de contrôle interne comptable de l'État.

Structure du budget 2017 de fonctionnement

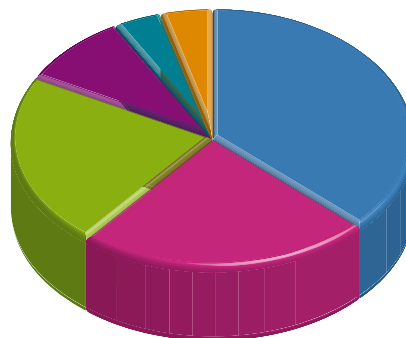
Dépenses de fonctionnement		Recettes	
Fluides (majorité l'électricité)	37 %	Subvention pour charges de service public	81 %
Sous-traitance (personnels en régie)	24 %	Subventions européennes	9 %
Maintenances et logiciels	22 %	Recettes propres	10 %
Prestations extérieures (accueil, ménage, gardiennage, ...)	9 %		
Personnels sous contrat	4 %		
Autres	4 %		

Recettes



- Subvention pour charges de service public
- Subventions européennes
- Recettes propres

Dépenses de fonctionnement



- Fluides (majorité l'électricité)
- Sous-traitance (personnels en régie)
- Maintenances et logiciels
- Prestations extérieures (accueil, ménage, gardiennage, ...)
- Personnels sous contrat
- Autres

DÉMATÉRIALISATION

Elle constitue un levier important d'amélioration de l'efficacité des processus au cours desquels des pièces justificatives doivent être prises en compte ou produites. Elle apporte une réponse adaptée aux enjeux de fiabilité, de fluidité, de traçabilité et de sécurité de l'information, et participe à l'amélioration de l'environnement, en limitant la consommation de papier.

La dématérialisation est un long processus qui a été depuis longtemps pris en compte au CINES, notamment sur les documentations techniques, les échanges d'informations ou le travail collaboratif. Sur les échanges de documents administratifs légaux elle a été mise en œuvre à compter du 1^{er} janvier 2017. Le CINES a l'obligation de récupérer les factures des « gros » fournisseurs sur la passerelle CHORUS Pro. Il met également à la disposition de ses clients sur cette même plateforme les factures émises. La dématérialisation concerne également les marchés publics avec la gestion des appels d'offres. Les nouveaux textes régissant les marchés publics imposent de donner un accès libre et complet aux données essentielles des contrats et des modifications de marchés (avenants, décisions de poursuivre, etc.).

› Formations organisées par le CINES	50
› Formations continues suivies par le personnel CINES	50
› Adhésion aux formations « communes »	50

FORMATIONS ORGANISÉES PAR LE CINES

Le CINES propose des formations à la communauté scientifique dans ses deux missions stratégiques : le calcul de haute performance et l'archivage pérenne de données numériques :

Archivage

- Introduction aux méthodes d'archivage numérique
- Formats d'archivage des données numériques
- Journées archivage du CINES
- Participation à diverses formations d'organismes tels que ENSIB, École des Chartes, Institut National du patrimoine, Direction générale des Patrimoines - Ministère de la Culture, Association Aristote - groupe PIN.

Calcul

- Débogage et optimisation
- Introduction aux architectures parallèles et au calculateur OCCIGEN
- Introduction au calcul parallèle : Open MP et MPI
- Utilisation du logiciel Gaussian sur OCCIGEN
- « Parallel linear algebra »
- « Tuning workshop »

FORMATIONS CONTINUES SUIVIES PAR LE PERSONNEL CINES

Entre 2015 et 2017 un peu plus de 115 sessions de formation ont été organisées au CINES pour les personnels dans les différents thèmes suivants :

- Ateliers du CNIL, formation du CIL
- Habilitations électrique H0B0 et BS BE
- Certifications Sécurité (homologation RGS, ISO 27005, cursus pour les correspondants RSSI)
- Sécurité réseau
- Sécurité Java/JEE
- Sécurité des applications Web
- Prévention des risques d'attentat
- Chargé d'évacuation des locaux
- Extinction incendie
- Sauveteurs Secouristes du travail
- Rôle du CHSCT
- Modules GFI : mise en œuvre de la GBCP, recettes, dépenses
- Actualités des marchés publics
- Contrôle comptable interne
- Gestion de projets
- Contrôle de version avec GIT
- Photoshop

- Système décisionnel TALEND
- Ingénierie et web des données, production du code
- Publication de rapports avec JASPER
- Programmation avec Python
- ElasticSearch
- Cloud computing la synthèse
- Linux l'essentiel
- Certification RedHat
- Administration d'un serveur d'applications Java EE avec JBoss WildFly
- TCP/IP mise en œuvre
- Préparation aux concours de l'ESR
- Informations sur la retraite

Ces formations ont été complétées par des cours d'anglais (collectifs et individuels) et de nombreuses participations des agents techniques à des MOOC ou aux divers congrès nationaux ou internationaux, et séminaires thématiques dans leurs domaines : JDEV (développement), journées Linux, « super computing », Teratec (Calcul), JRES (réseau), RDA (archivage), ...

ADHÉSION AUX FORMATIONS « COMMUNES »

Le CINES a adhéré à la convention cadre inter-établissements, regroupant les universités de Montpellier, de Nîmes, de Perpignan, l'école nationale de chimie de Montpellier et l'Abes pour la formation des personnels. L'objectif de cet accord est la mutualisation de certaines actions de formation et la mise à disposition collaborative de locaux et de matériels. Les agents du centre pourront ainsi bénéficier de l'offre de formations du réseau mais également proposer des actions de formation.

► Juin 1982 inauguration du Centre (ex Centre National Universitaire Sud de Calcul)

Midi ^{M 1} Régio

DES ALPES AUX PYRENEES ♦ SAMEDI 5 JUIN 1982 ♦ DE L'Auvergne

Alain Savary : l'informatique vocation montpelliéraine

La visite du ministre de l'Education Nationale, hier à Montpellier, fut résolument « pratique ». A aucun moment, en effet, Alain Savary n'évoqua les grands thèmes de sa politique et plus particulièrement le problème aigu de l'enseignement privé.

Venu pour inaugurer le Centre National Universitaire Sud de Calcul et pour célébrer le centenaire du lycée « Clemenceau », le ministre de l'Education mit essentiellement l'accent sur la vocation de Montpellier pour l'informatique appliquée.

A ce jour, rappela Alain Savary, deux centres nationaux dont la fonction prioritaire est de délivrer une puissance de calcul importante en traitement par lot à distance sont installés dans le pays. Le Centre International de Calcul Electronique d'Orsay et le C.N.U.S.C., que nous inaugurons aujourd'hui. L'installation de ce centre dans la région Languedoc-Roussillon marque le prix que nous avons attaché à l'effort important ainsi consenti. C'est une raison suffisante pour que ce centre joue la carte régionale et de la décentralisation.

Employant 45 personnes, le C.N.U.S.C. a pour rôle de fournir aux communautés scientifiques de la moitié sud de la France les outils dont elles ont besoin en traitement de l'information et d'aider à la mise en place et au développement des applications informatiques préconisées par les responsables des collectivités locales. Après cette présentation du nouvel établissement, le ministre de l'Education termina en beauté. Autrement dit par l'annonce de la création d'un Centre National en Informatique Appliquée, dès la rentrée prochaine. Ce centre occupera les locaux, initialement prévus



Le directeur du Centre national de calcul, M. Ippolito au cours de la visite inaugurale, en compagnie de MM. Vincent, préfet de région ; Frêche, maire de Montpellier ; Alain Savary et, au deuxième plan le recteur, M. Farran.

pour accueillir un institut d'informatique d'entreprise du Conservatoire National des Arts et Métiers (C.N.A.M.).

Pourquoi ce changement de programme ? Alain Savary fut très clair. Si l'idée était séduisante, le choix retenu n'était pas le plus heureux. Avant de former des informaticiens, il faut penser à former des formateurs ! C'est ce à quoi nous allons nous appliquer. De ce fait plusieurs postes de professeurs seront très prochainement créés.

Tout à son sujet Montpellier point d'ancrage de la filière électronique, le ministre de l'Education n'eut guère le loisir, durant cette visite jalonnée de nombreuses audiences, d'aborder dans le détail d'autres questions. A noter toutefois l'annonce de la publication, dans quinze jours, d'une déclaration qui ne devrait pas passer inaperçue en Languedoc-Roussillon. Elle fixera en effet les grandes lignes de la politique gouvernementale en matière d'enseignement des langues minoritaires.

Centre Informatique National de l'Enseignement Supérieur

950, rue de Saint-Priest - 34097 Montpellier Cedex 5 - France

Tél. : 04 67 14 14 14 - Fax : 04 67 52 37 63

www.cines.fr - services utilisateurs : svp@cines.fr

Directeur de la publication : Francis Dumas

Co-rédacteurs : Marie Galez, Marion Massol, Marcel Pontillon, Olivier Rouchon

