

COMITÉ NATIONAL DE LA RECHERCHE SCIENTIFIQUE
CONSEIL SCIENTIFIQUE D'INSTITUT

Compte rendu

Conseil scientifique de l'IN2P3
23-24 juin 2022

Sommaire

1. Introduction à la séance du 23-24 juin	5
2. Physique nucléaire : l'utilisation du HPC pour la physique théorique	6
2.1 Présentation des activités	6
2.2 Avis et recommandations	6
3. Physique des particules : les activités pour le LHC et les développements pour le HL-LHC	7
3.1 Présentation des activités	7
3.2 Avis et recommandations	9
4. Physique des particules : Les développements innovants pour le traitement en ligne des données	10
4.1 Présentation	10
4.2 Avis et recommandations	11
5. Astroparticules : Les développements pour Rubin, Euclid, CTA	12
5.1 Description des développements pour Rubin/LSST	12
5.2 Description des développements pour Euclid	13
5.3 Description de l'analyse en temps réel pour CTAO	14
5.4 Avis et recommandations	14
6. Astroparticules : Les développements pour le traitement des alertes basses latences	16
6.1 Présentation	16
6.2 Avis et recommandations	17
7. Les développements logiciels et intergiciels	19
7.1 Présentation	19
7.2 Présentation des logiciels de simulation	19
7.3 Présentation des logiciels de traitement des données et analyse	20
7.4 Présentation des logiciels de gestion des jobs, données et métadonnées	21
7.5 Avis et recommandations	22
8. Développements logiciels et optimisation du calcul	24
8.1 Description du volet Développements, portabilité et performances des logiciels (Reprises)	24
8.2 Description du volet Optimisation du calcul : déploiement et ordonnancement	24
8.3 Avis et recommandations	25
9. Les développements en intelligence artificielle issus des projets IN2P3	27

9.1 Présentation	27
9.2 Avis et recommandations	27
10. L'informatique quantique	28
10.1 Présentation	28
10.2 Avis et recommandations	29
11. EOSC (« European Open Science Cloud »)	30
11.1 Présentation	30
11.2 Avis et recommandations	31
12. Discussions internes au CSI	33
12.1 Discussions avec la direction	33
12.2 Vie du Conseil	34

Présents : G. Brooijmans, H. Costantini, B. Cros, N. Chanon, O. Drapier, S. Escoffier-Martory, P. Janot, L. Fayard, B. Fernández Domínguez, V. Givaudan, S. Henrot-Versillé, D. Laporte, F. Marion, R. Maurice, N. Neyroud Gigleux, C. Nones, B. Ramstein, M. Rousseau, C. Smith, R. Trebossen, G. Verde, M. Yamouni

Invités : P. Balcou (CS CNRS), J.-J. Hernandez-Rey (CS CNRS)

Excusés : M. Lindroos, F. Yermia

Orateurs : E. Fede (CC-IN2P3), M. Jouvin (IJCLab), G. Hupin (IJCLab), L. Duflot (IJCLab), D. Vom Bruch (CPPM), D. Boutigny (LAPP), J. Peloton (IJCLab), L. Arrabito (LUPM), D. Chamont (IJCLab), M. Souchal (APC), D. Rousseau (IJCLab), F. Magniette (LLR), I. Bird (LAPP)

Rapporteurs : S. Campana (CERN), F. Carbognani (EGO), M. Dayde (IRIT), A. Di Meglio (CERN), T. Dorigo (INFN), G. Stewart (CERN)

Membres présents de la direction : S. Crépe-Renaudin, B. Giebels, M. Grasso, V. Poireau, L. Vacavant

1. Introduction à la séance du 23-24 juin

La séance du 23-24 juin 2022 était dédiée à l'examen des projets de Calcul et Données à l'IN2P3. Cette séance s'est tenue à la fois en présentiel et à distance.

Programme de la session ouverte :

1) Contexte : les ressources Calcul et Données à l'IN2P3 dans le contexte national et international

1.1) Panorama des infrastructures utilisées par les projets de l'institut et développements à venir (E. Fede)

1.2) Les défis, les contraintes et les opportunités (M. Jouvin)

2) Les contributions importantes par thématique scientifique et les logiciels

2.1) Physique nucléaire : l'utilisation du HPC pour la physique théorique (G. Hupin)

2.2) Physique des particules : les activités pour le LHC et les développements pour le HL-LHC (L. Duflot)

2.3) Physique des particules : les développements innovants pour le traitement en ligne des données (D. Vom Bruch)

2.4) Cosmologie : les développements pour LSST, Euclid (D. Boutigny)

2.5) Astroparticules : les développements pour le traitement des alertes basses latences (J. Peloton)

2.6) Les développements logiciels et interlogiciels : revue des logiciels ou suites de logiciels avec une forte participation de l'IN2P3 (L. Arrabito)

3) Les programmes transverses et de R&D

3.1) Développements, portabilité et performances des logiciels (D. Chamont)

3.2) Optimisation du calcul : déploiement et ordonnancement (M. Souchal)

3.3) Les développements en intelligence artificielle issus des projets IN2P3 (D. Rousseau)

3.4) L'informatique quantique (F. Magniette)

4) L'EOSC : objectifs, implications et enjeux pour nos disciplines et développements à l'IN2P3 (I. Bird)

Toutes les présentations sont accessibles depuis la page web du CSI IN2P3 :

<https://www.in2p3.cnrs.fr/fr/le-conseil-scientifique-de-lin2p3>

Le présent compte rendu aborde tout d'abord les contributions thématiques avec l'utilisation du HPC en physique théorique et particulièrement la physique théorique nucléaire (section 2), la physique des particules avec les activités pour le LHC et les développements pour le HL-LHC (section 3) ainsi que les développements pour le traitement des données en ligne (section 4), les astroparticules et la cosmologie avec les développements pour LSST, Euclid et CTA (section 5) ainsi que le traitement des alertes basses latences (section 6). Sont

ensuite abordés les thèmes des développements logiciels et intergiciels (section 7), puis les programmes transverses de R&D avec les développements, portabilité et performances des logiciels ainsi que l'optimisation du calcul (section 8), les développements en intelligence artificielle (section 9), l'informatique quantique (section 10), et enfin l'EOSC (section 11). Les discussions internes au Conseil font l'objet de la section 12.

2. Physique nucléaire : l'utilisation du HPC pour la physique théorique

2.1 Présentation des activités

Dans le domaine du calcul parallèle, on distingue le HPC (« High Performance Computing ») utilisant une interface pour du calcul parallèle avec mémoire partagée, par exemple avec le protocole MPI (« Message Passing Interface ») ou OpenMP (« Open Multi-Processing »), et le HTC (« High Throughput Computing ») utilisant une ferme de calcul optimisée pour des calculs indépendants.

Plusieurs équipes de l'IN2P3 développent des codes de calcul et utilisent des ressources HPC pour mener à bien des calculs de physique nucléaire, de physique des particules, ou d'astroparticules et de cosmologie. En fonction du domaine d'application et de la méthode utilisée, les besoins sont extrêmement différents : alors que les calculs DFT (« Density Functional Theory ») de structure nucléaire sont les moins consommateurs de ressources, les calculs de chromodynamique quantique (QCD) sur réseau sont extrêmement gourmands. Les personnels associés à ces activités émergent pour la plupart sur des projets IN2P3 ou ANR, et collaborent à l'international.

Différentes infrastructures sont utilisées : fermes de calcul locales, mésocentres régionaux, CC-IN2P3, calculateurs et supercalculateurs de GENCI (Grand Equipement National de Calcul Intensif) et PRACE (« Partnership for Advanced Computing in Europe »). Le CC-IN2P3 est spécialisé dans le calcul HTC, alors que le calcul HPC n'y dispose que de 16 nœuds dédiés. Cependant, l'accès au CC-IN2P3 demeure crucial en amont des projets GENCI, pour la préparation de soumission de calcul à grande échelle, ou pour l'analyse des résultats après calcul. On note donc une complémentarité des différents niveaux.

2.2 Avis et recommandations

Les équipes ayant des activités de HPC travaillent de manière indépendante dans les domaines de la physique nucléaire, physique des particules et astroparticules ou cosmologie, car les problématiques traitées et les perspectives scientifiques sont différentes. Cependant les besoins en HPC sont croissants, quelle que soit la discipline. Les développements se font aujourd'hui majoritairement sur des clusters locaux ou des mésocentres régionaux, car le CC-IN2P3 est spécialisé sur le calcul HTC.

Le Conseil note l'inquiétude des équipes quant au devenir des clusters locaux, de plus en plus intégrés dans des mésocentres régionaux. Les modalités d'accès à long terme de ces mésocentres doivent être clarifiées, notamment en ce qui concerne une potentielle participation financière des laboratoires. D'autre part, le transfert de larges fichiers de données depuis l'étranger vers les centres HPC français sont actuellement problématiques, ce qui freine l'utilisation du GENCI par les physiciens des particules.

De plus, le portage des codes sur architectures GPU (« Graphics Processing Unit », cartes graphiques) est un enjeu commun. Une adaptation ou une réécriture des codes existants est en effet nécessaire pour tirer pleinement profit de ces architectures, qui peuvent améliorer la rapidité d'exécution par rapport au calcul parallèle sur CPU.

Globalement, il est important de maintenir un niveau d'expertise suffisant pour soutenir les activités de l'institut en calcul HPC. Cependant, les équipes notent qu'il est de plus en plus difficile de recruter des CDD ingénieurs en informatique (qui se voient offrir des salaires bien plus élevés dans le privé). Le Conseil suggère donc comme alternative de recruter des post-doctorants et des chercheurs permanents ayant des compétences duales en physique et en informatique, sur les thématiques de développement de code et de calcul HPC.

Executive summary

The teams working with HPC resources independently work in the fields of nuclear physics, particle physics and astroparticles or cosmology, because the issues addressed and the scientific perspectives are different. However, HPC needs are growing, whatever the discipline. Developments are now mainly done on local clusters or regional mésocentres, whereas the CC-IN2P3 is specialized in HTC computing.

The Scientific Council notes the common concern about the future of local clusters, which are increasingly integrated into regional mésocentres. The terms and conditions for long-term access to these mésocentres must be clarified, particularly with regard to potential financial participation by laboratories. On the other hand, the transfer of large data files from abroad to GENCI HPC centers is currently problematic, which hinders their use by particle physicists.

Furthermore, the porting of codes on GPU (graphics processing unit) architectures is a common challenge. An adaptation or a rewriting of the existing codes is indeed necessary to take full advantage of those architectures, which can improve the speed of execution compared to parallel CPU calculations.

Overall, it is important to maintain a sufficient level of expertise to support the activities of the institute in HPC computing. However, the teams note that it is increasingly difficult to recruit computer engineers (who are offered much higher salaries in the private sector). The Scientific Council therefore suggests as an alternative to recruit post-docs and permanent researchers with dual skills in physics and computer science, on code development and HPC computing.

3. Physique des particules : les activités pour le LHC et les développements pour le HL-LHC

3.1 Présentation des activités

Le calcul et les données du LHC sont gérés en France par LCG-France, une organisation créée par l'IN2P3 et le CEA à l'intérieur de WLCG (« Worldwide LHC Computing Grid ») et qui assure la coordination technique d'une infrastructure comprenant neuf centres de

calculs, un “Tier1” (CC-IN2P3 à Lyon), sept “Tier2” (Annecy, Clermont-Ferrand, Grenoble, Paris, Palaiseau, Orsay et Saclay, Marseille, Nantes et Strasbourg) et un “Tier3” (Lyon).

Le protocole d'accord suivi au cours des dix dernières années (2012-2022) par LCG-France, conforté par l'implication et la détermination des équipes techniques dans chacun des centres, a couvert les besoins croissants des expériences (ATLAS, CMS, ALICE, LHCb) au prorata de la participation française jusqu'au début du Run3, tant pour le calcul que pour le stockage des données. Le budget supportant ce protocole d'accord a permis une croissance régulière (10% / an) des capacités du Tier1 et une contribution majoritaire à la maintenance des Tier2.

Les besoins des expériences pour le Run3 du LHC, en particulier pour LHCb, et les demandes encore plus importantes prévues pour le HL-LHC, ne pourront pas être satisfaits dans le cadre du modèle actuel et avec les solutions techniques en cours jusqu'à aujourd'hui, surtout avec le budget constant planifié dans de nombreux pays. Par exemple, le modèle actuel repose sur l'utilisation de bandes pour le stockage des données, en raison de son coût, mais ce choix pourrait être remis en cause par la réduction du marché international, le rendant à terme moins compétitif. Les difficultés financières de RENATER, qui pourvoit le réseau principal pour le transfert de données, mettent en danger la continuation et la jouvence du service, qui pourrait en conséquence ne pas pouvoir absorber l'augmentation très significative du trafic prévue pour la décennie à venir. Ces difficultés étaient en cours de résolution au plus haut niveau en juin 2022. L'arrêt prévu de deux Tier2 (Grenoble et Nantes) à la fin du Run3, motivé en particulier par le départ à la retraite du personnel technique, ainsi que les départs à la retraite prévus à Clermont-Ferrand et Marseille, auront également un impact négatif sur les évolutions à venir. En outre, la “feuille de route du numérique” préparée conjointement depuis quelques années par le Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI) et le CNRS préconise implicitement de réduire/supprimer le financement des Tier2, et de les remplacer par de grands centres régionaux éligibles pour un financement national via des Contrats Plan État-Région (CPER).

Des groupes de travail ont été mis sur pied par WLCG il y a cinq ans pour tenter de développer des solutions nouvelles pour le HL-LHC, et LCG-France a participé activement à plusieurs d'entre eux : modélisation des coûts pour différentes stratégies de calculs, accès - transfert - archivage - stockage des données (DOMA), évaluation des performances de différents matériels pour la génération, la simulation, la digitisation et la reconstruction d'événements, etc. Ce travail de R&D, maintenant terminé, n'a pas permis d'identifier de solutions “magiques” pour le HL-LHC en termes de réduction de coûts ou d'amélioration de performance, si ce n'est une réduction drastique de la taille des formats d'analyse pour les événements LHC.

La piste des calculateurs de haute performance (HPC), suivie par WLCG et qui pourrait offrir une solution possible pour les besoins en calculs du LHC, a été explorée en France par les équipes ATLAS de l'IN2P3, avec la participation du centre de calcul IDRIS (Institut du Développement et des Ressources en Informatique Scientifique). Malgré un certain succès aux USA et dans certains autres pays européens, les difficultés techniques rencontrées (qui de plus varient énormément d'un site HPC à l'autre) et les conditions d'accès complexes n'ont pour le moment pas permis d'établir une solution pérenne. Le projet FITS (« Federated

IT Services ») de fédérer le CC-IN2P3 et IDRIS via une infrastructure distribuée (2021-2029), pourrait offrir à terme une réponse aux besoins de calcul du HL-LHC basée sur HPC au HL-LHC et à WLCG (voir la présentation *Panorama des infrastructures des projets de l'institut*, sur le site web du CSI, <https://www.in2p3.cnrs.fr/fr/le-conseil-scientifique-de-lin2p3>).

3.2 Avis et recommandations

Le Conseil félicite les équipes impliquées dans LCG-France pour les performances obtenues sur la période 2012-2022. Celles-ci ont effectivement dépassé les engagements pris et sont supérieures à la moyenne des collaborateurs internationaux.

Le protocole d'accord actuel arrivant à sa fin, un nouveau protocole doit être négocié, mais l'incompatibilité entre les besoins des expériences et l'évolution attendue des capacités techniques et financières pose un sérieux problème. Si c'est avant tout aux collaborations de trouver des solutions communes (par exemple en réduisant leurs besoins en stockage et temps de simulation), l'absence de solutions magiques signifie qu'il faut poursuivre le plus d'options possibles, dont la piste HPC, où la France accuse un certain retard.

Le Conseil recommande à la direction de surveiller l'évolution de la situation concernant les difficultés financières de RENATER. Le changement de direction de RENATER et le fait que le ministère se préoccupe du problème permettent d'espérer une solution à plus ou moins court terme. Si cela ne se concrétise pas, le Conseil recommande à la direction de faire remonter au ministère l'importance de RENATER pour le (HL-)LHC (et bientôt pour LSST, etc.).

Le Conseil recommande également à la direction de surveiller la stratégie nationale des infrastructures de recherche du MESRI et en particulier la feuille de route du numérique, à laquelle il faudra s'adapter. La centralisation régionale mettra en difficulté les Tier2 actuels, et créera potentiellement des difficultés sur l'utilisation des nouveaux Tier2 vu que les besoins des centres de calculs LHC, principalement au niveau de la permanence du stockage, sont difficilement compatibles avec les configurations typiques de centres de calculs polyvalents. La hausse des coûts du matériel et de l'électricité, les départs à la retraite de bon nombre d'administrateurs de sites grille (Tier2) d'ici au démarrage du HL-LHC, ainsi que les difficultés de recrutement sur ces profils spécifiques, seront d'autres défis à relever.

En conclusion, l'avenir à court, moyen, et long terme des activités concernant le calcul et les données en France est incertain et risqué. La planification et la mise en œuvre de solutions pratiques pour pallier les effets de ces incertitudes et des risques associés devra être l'une des priorités essentielles de l'IN2P3 lors des années et décennies à venir : pour le Run3 du LHC dès à présent ; le HL-LHC à court terme (2029) ; le FCC-ee à moyen terme (2045), avec ses 10^{13} Z à collecter (ou simuler), stocker, et analyser; et le FCC-hh à long terme (2070), dont les problèmes de stockage de données et de temps de calculs seront au moins un ordre de grandeur plus compliqués que ceux du HL-LHC.

Executive summary

The Council congratulates the IN2P3 teams involved in LCG-France for the computing performance achieved over the period 2012-2022, exceeding both the initial objectives and the average performance reached by their international collaborators.

With the current MoU coming to an end, a new MoU needs to be negotiated, but the incompatibility between the needs of the experiments and the expected evolution of technical and financial capacities poses a serious problem. While it is mostly up to the collaborations to find common solutions (for example by reducing their storage needs and simulation time), the absence of magic solutions means that it will be necessary to pursue as many options as possible, including the deployment of HPC's, for which France is lagging.

The Council recommends that the direction monitor RENATER's financial difficulties. The change of management at RENATER, and the fact that the Ministry is concerned about the problem, make it possible to contemplate a rapid resolution. If a solution does not materialize, the Council recommends that the direction escalate to the Ministry the importance of RENATER for the (HL-)LHC (and soon LSST, etc.).

The Council also recommends that the direction monitor the national strategy for MESRI research infrastructures, and in particular the digital roadmap, which will require flexibility and adaptation. Regional centralization will put the current Tier2s in jeopardy, and might create issues in the use of the new Tier2s, given that the needs of the LHC computing centres (e.g., in terms of data storage) are not normally compatible with typical general-purpose data centre configurations. The rise in the cost of equipment and electricity, the retirement of a large number of grid site administrators (Tier 2) between now and the start of the HL-LHC, as well as the difficulties in recruiting for these specific profiles, are among the many challenges to be solved.

To conclude, the short, medium, and long-term future of computing and data activities in France is uncertain and risky. The planning and implementation of practical solutions to mitigate the effects of these uncertainties and the associated risks will have to be one of the essential priorities of IN2P3 in the years and decades to come, starting now with the LHC Run3, and followed by HL-LHC in the short term (2029); FCC-ee in the mid-term (2045), with its 10^{13} Z to collect (or simulate), store, and analyse; and FCC-hh in the long term (2070), whose data storage and computing time challenges will be at least an order of magnitude more complex than those of the HL-LHC.

4. Physique des particules : Les développements innovants pour le traitement en ligne des données

4.1 Présentation

Le traitement en ligne (et donc en temps réel) des données du LHC, en particulier pour ce qui concerne le système de déclenchement de l'acquisition des événements, doit également s'adapter aux défis posés par l'augmentation des taux de collisions prévus à partir de 2022, d'un facteur 5 pour ALICE et LHCb dès le Run3 (et d'un autre facteur 5 à partir de 2040), et d'un ordre de grandeur pour ATLAS et CMS dès le début du HL-LHC (2029). Plus précisément, le flux de données à traiter en temps réel pour LHCb sera de 5 To/s dès 2022,

et montera à 25 To/s en 2040. Le taux de déclenchement de premier niveau pour ATLAS et CMS sera accru de 100 kHz à 1 MHz environ en 2029.

Pour améliorer leur efficacité de déclenchement, les expériences innovent sur leurs schémas de traitement des données et augmentent la sophistication des algorithmes utilisés dans les premières étapes de sélection : des exemples d'algorithmes incluant l'intelligence artificielle dès le premier niveau de déclenchement, que ce soit sur des FPGA (ATLAS et CMS), des CPU ou des GPU (LHCb), ont été présentés. L'utilisation des plateformes spécialisées (FPGA et GPU) demande de porter les codes sur ces plateformes. Les équipes développent par exemple le cadre ALLEN (logiciel de déclenchement de haut niveau de LHCb sur GPU) qui fournit les outils pour exploiter l'architecture parallèle des GPUs, et permet d'exploiter le même code sur les architectures GPU et CPU. Les équipes contribuent aussi au développement d'outils spécifiques, par exemple HLS4ML (« High Level Synthesis language for Machine Learning »), pour l'utilisation de réseaux de neurones sur les FPGA. Dans ce contexte, le projet THINK (« Testing Hardware Instantiations of Neural Kernels ») est en cours pour fournir aux utilisateurs des outils permettant de déterminer la meilleure architecture pour chaque type d'intelligence artificielle (par exemple, réseau de neurones graphique ou convolutionnel).

4.2 Avis et recommandations

La contribution des équipes de l'IN2P3 dans les projets utilisant l'intelligence artificielle dès le premier niveau de déclenchement est en pointe à l'international. Le Conseil encourage les équipes concernées à conserver l'initiative et à développer leur expertise dans ces domaines.

Les développements sur FPGA bénéficieraient d'une mise en commun de l'expérience acquise indépendamment sur ATLAS et CMS (par exemple via le projet THINK), ainsi que d'un engagement plus visible, plus organisé, et plus hiérarchique de l'IN2P3, afin de progresser de manière centralisée avec plus d'efficacité (y compris en augmentant les ressources financières et humaines).

L'utilisation de GPUs par la collaboration LHCb est un succès indéniable et représente sans doute une solution d'avenir, en particulier si le logiciel ALLEN (logiciel de déclenchement de haut niveau de LHCb sur GPU) peut être adapté pour fonctionner sur des machines à haute performance de calcul.

Executive summary

The IN2P3 teams give leading contributions at the international level in projects using artificial intelligence in L1 Trigger systems. The Council encourages the teams concerned to maintain their initiative and further develop their expertise in these areas.

Developments with FPGAs would benefit from exploiting potential synergies between ATLAS' and CMS' independent efforts (for example via the THINK project), as well as from a more visible, more organized, and more hierarchical commitment of IN2P3 (including increasing financial and human resources), in order to progress more efficiently.

The use of GPUs by the LHCb collaboration is an undeniable success, and undoubtedly represents a promising solution, in particular if the ALLEN software (LHCb high-level

triggering GPU software) can be adapted to run on machines with high computing performance.

5. Astroparticules et cosmologie : Les développements pour Rubin/LSST, Euclid, CTA

5.1 Description des développements pour Rubin/LSST

L'observatoire Rubin sera d'ici quelques années l'un des instruments phares pour la cosmologie. Doté d'un grand champ de vue et d'une caméra de 3,2 milliards de pixels, il effectuera le relevé du ciel austral le plus profond et le plus vaste jamais entrepris en enregistrant une image toutes les 40 secondes chaque nuit pendant dix ans. Le volume de données associé est à la mesure de l'ampleur du relevé LSST : environ 20 To enregistrés chaque nuit et 2000 Po de données au total. Le catalogue final répertoriera 37 milliards d'objets dans une base de données de 15 Po. L'IN2P3 est le seul institut du CNRS à être impliqué dans Rubin (construction et opération) et a des responsabilités majeures dans le traitement de ces données.

Chaque image sera immédiatement transférée et traitée sur la plateforme américaine (USDF à SLAC), ce qui permettra d'émettre des alertes à faible latence (~10 millions par nuit) à destination de la communauté scientifique. Une à deux fois par an, l'ensemble des données brutes sera traité par la chaîne de « Data Release Processing » (DRP) pour produire les catalogues d'objets et des images directement exploitables par les analyses de physique.

La France, à travers le CC-IN2P3, sera responsable de 50% du DRP. Celui-ci est adapté à des systèmes hétérogènes et l'orchestration sera faite depuis l'USDF. Le DRP s'appuie sur des technologies variées, comme le Google Cloud et l'outil de virtualisation Kubernetes, mais également des outils développés par la communauté HEP comme PanDA et Rucio. Le système de traitement des images est actuellement en test en parallèle aux USA et au CC-IN2P3.

Le CC-IN2P3 abritera par ailleurs une copie de l'intégralité des données, offrant à la communauté française un accès privilégié favorisant le retour scientifique. Le stockage et l'accès aux objets du catalogue se fera grâce à Qserv, une base de données optimisée pour l'astronomie et les grands volumes de données. L'IN2P3 a acquis une grande expertise avec cet outil, qui a été déployé au CC-IN2P3 sur une plateforme Kubernetes et est d'ores et déjà utilisé pour prototyper des analyses scientifiques. L'interface entre analyse et données se fait via la « Rubin Science Platform » (RSP), que l'IN2P3 a contribué à développer. La France est par ailleurs très impliquée dans le développement de l'agent de messages (« broker ») Fink, déployé au CC-IN2P3, pour la classification et la redistribution des alertes émises par Rubin/LSST (section 6).

Concernant l'exploitation scientifique, l'IN2P3 est principalement engagé dans la collaboration DESC (« LSST Dark Energy Science Collaboration »). Si celle-ci exploitera les données traitées de Rubin, elle aura également besoin de revenir aux données brutes pour atteindre la précision nécessaire aux applications cosmologiques. Elle devra aussi mettre en œuvre des traitements complexes pour extraire les paramètres cosmologiques à partir des

diverses sondes. Le supercalculateur NERSC à Berkeley fournira l'essentiel des ressources en calcul pour DESC. L'IN2P3 prévoit de maintenir les outils d'analyse également au CC-IN2P3 pour contribuer au calcul et optimiser le retour scientifique en France (notamment une instance de la RSP).

Le personnel de l'institut qui contribue à l'ensemble de ces développements informatiques se situe d'une part au CC-IN2P3, avec 7 ETP, et d'autre part dans les laboratoires (IJCLab, LAPP, LPC, LPSC) pour 4,4 ETP.

5.2 Description des développements pour Euclid

La mission spatiale Euclid sera également un instrument phare pour la cosmologie à partir de 2023. Elle est dotée d'un imageur dans le visible offrant une haute résolution angulaire pour la mesure du lentillage gravitationnel, ainsi que d'instruments dans le proche infrarouge (photomètre et spectrographe) dédiés à la cartographie de la distribution des galaxies dans l'Univers. Elle couvrira un tiers du ciel pendant six ans. Ces mesures devront être combinées avec des données externes afin d'atteindre la précision ultime sur les paramètres de l'équation d'état de l'énergie noire. Le volume de données brutes sera de 140 Go par jour et le volume total, incluant les données intermédiaires et externes, sera de 150 Po. Euclid mesurera deux milliards de galaxies, dont quelques pour cent avec des données spectroscopiques. La France (CNES, IN2P3, INSU, IRFU) contribue à 30% dans la construction d'Euclid et fournira notamment l'un des Centres de Données Scientifiques (SDC).

Le segment sol d'Euclid se compose d'une partie dédiée à la station sol et au centre des opérations de la mission (« Operation Ground Segment », OGS) et d'une partie dédiée à la science (« Science Ground Segment », SGS). L'objectif de cette dernière est de fournir les données traitées pour l'exploitation scientifique ainsi que les métadonnées de traçabilité et de qualité associées. Elle s'appuie sur une architecture distribuée centrée sur les données, les SDCs dans chaque pays participant étant en charge d'appliquer les chaînes de traitement aux données provenant d'une portion spécifique du ciel. Les SDCs sont également chargés du traitement des données instrumentales piloté par les équipes chargées de l'exploitation des instruments (IOT). Le SGS s'appuie sur des unités organisationnelles (OUs) chargées de concevoir les algorithmes de traitement et de développer des codes prototypes.

Les données d'Euclid sont organisées en plusieurs niveaux, des données brutes aux catalogues en passant par les données étalonnées pour chaque instrument, incluant également des données issues d'un programme approfondi de simulation nécessaire pour minimiser et maîtriser les incertitudes systématiques du relevé. Les équipes de l'IN2P3 contribuent aux activités de simulation et de traitement en lien avec le spectrographe infrarouge et avec les données externes, ainsi que dans la production des données de haut niveau pour l'exploitation scientifique. Elles sont des acteurs importants dans la série de tests de grande échelle (« scientific challenges ») menés par Euclid en préparation de la mission. Elles seront par ailleurs pleinement impliquées dans le fonctionnement, l'étalonnage et le contrôle du spectro imageur NISP.

Parmi la dizaine de SDCs d'Euclid, le centre français est hébergé par le CC-IN2P3 qui fournira 30% des capacités de stockage et de traitement des données de la mission. Le

traitement et le stockage des données sont partagés entre les différents SDCs grâce à une infrastructure de type grille, les méta-données étant centralisées dans une base de données hébergée à l'ESAC (Madrid). Le SDC français fournit également une infrastructure de développement de logiciels.

Le personnel de l'institut qui contribue à l'ensemble de ces efforts se situe d'une part au CC-IN2P3, avec 3,3 ETP, et d'autre part dans les laboratoires (CPPM, IP2I, APC, LPSC) pour 10,7 ETP. Il est à noter que 60% de ces ressources humaines correspondent à des CDD financés par le CNES.

5.3 Description de l'analyse en temps réel pour CTAO

CTAO sera à partir de 2028 le plus grand observatoire terrestre pour l'astronomie gamma, avec plusieurs dizaines de télescopes dans les hémisphères nord et sud, un grand champ de vue, une excellente résolution angulaire et une large gamme d'énergie (de 20 GeV à 300 TeV), ainsi qu'une capacité à pivoter rapidement. Il émettra de manière automatique des alertes en cas de détection de phénomènes transitoires ou variables, avec une latence inférieure à 30 secondes.

L'émission automatique d'alertes repose sur la chaîne d'analyse SAG (« pipeline science alert generation »), en cours de développement pour le premier télescope de grande taille (LST-1). A l'IN2P3, le LAPP (2,5 FTE) y contribue via le module de reconstruction qui traite les événements de bas niveau. Pour répondre aux exigences de latence, ce développement s'appuie sur une approche de calcul HPC.

5.4 Avis et recommandations

Le Conseil félicite les équipes de l'IN2P3 pour leurs contributions significatives dans le développement de l'infrastructure de traitement des données de Rubin/LSST. Celle-ci s'appuie sur des outils modernes et pointus, conçus pour être performants à l'échelle de grande ampleur du relevé. L'IN2P3 a investi des ressources pour maîtriser la complexité du système, ce qui est de bon augure pour maximiser les contributions françaises à la production scientifique ; cet effort doit être poursuivi.

Le Conseil salue également la contribution importante de l'IN2P3 au calcul d'Euclid. Celle-ci couvre un large spectre, de l'infrastructure de calcul jusqu'aux analyses de physique en passant par l'instrument, les simulations et les données externes. Une telle ampleur est adaptée à l'optimisation du retour scientifique de la mission.

Le Conseil attire l'attention de la direction de l'institut sur les points de vigilance suivants.

- Le transfert des données du LSST entre l'USDF et le CC-IN2P3 nécessite un réseau à haut débit (100 Go/s). Le CC-IN2P3 travaille avec RENATER et GEANT pour établir ce lien. Il est crucial que ce chantier aboutisse à temps.
- Suite à la réorganisation du projet, il est important que l'accord formel entre Rubin et l'IN2P3 établi en 2015 pour définir le rôle de l'institut dans le projet soit adapté afin d'affirmer la place et les contributions de l'IN2P3.
- Les technologies utilisées par Rubin/LSST reposent en grande partie sur des développements issus des GAFAMs qui demandent des compétences très pointues

et spécifiques à la fois pour ce qui est de leur implémentation et de leur maintenance. Le Conseil recommande que la direction soit particulièrement vigilante à ce que les choix du CC-IN2P3 (« on-premise system ») restent compatibles avec les développements sur le DRP. D'autre part, il est crucial de maintenir, suivre et développer les compétences sur ces technologies en évolution constante et rapide afin que l'IN2P3 garde un rôle central dans Rubin.

- L'effort de calcul à prévoir au CC-IN2P3 pour DESC ne doit pas être négligé, même s'il reste modeste par rapport au traitement des données de LSST, car il est le garant d'une contribution forte à la science.
- Le Conseil rappelle sa recommandation d'octobre 2020 d'appuyer fortement l'élaboration d'un accord de partage de données entre Rubin/LSST et Euclid, pour concrétiser la synergie scientifique prometteuse entre les deux projets.
- Le financement des ressources (calcul et stockage) allouées à Euclid au CC-IN2P3 est assuré par le CNES via une convention avec l'IN2P3 qui couvre les besoins estimés jusqu'à la première publication des données de la mission. Une estimation actualisée des besoins est en cours ainsi qu'une discussion pour renouveler la convention afin de couvrir toute la mission. Les ressources humaines financées par le CNES sont également déterminantes.
- Le calcul et le stockage nécessaires à l'exploitation scientifique d'Euclid pour produire les résultats de cosmologie ne sont pas pris en compte dans les besoins du SGS. Il faudra rester attentif à leur évaluation qui est en cours.

Executive summary

The Council is pleased with the very significant contributions of the IN2P3 teams to the development of the Rubin/LSST data processing infrastructure. The latter relies on modern and specialized tools, designed to perform well at the large scale of the survey. IN2P3 has invested resources to get an expert knowledge of the complex system, which bodes well for maximizing French contributions to science; this effort must be continued.

The Council also recognizes IN2P3's major contribution to Euclid computing. This covers a broad spectrum, from the computing infrastructure to science analyzes, including the instrument, simulations and external data. Such a breadth is appropriate to optimize the scientific return of the mission.

The Council wishes to draw the attention of the institute management to the following critical issues.

- LSST data transfer between the USDF and the CC-IN2P3 requires a high-bandwidth network (100 Go/s). CC-IN2P3 is working with RENATER and GEANT to set up such a connection. It is crucial that this comes to a successful conclusion.
- Following the reorganization of the project, the formal agreement between Rubin and IN2P3 established in 2015 to define the role of the institute should be adapted to affirm the role and contributions of IN2P3.
- Technologies used by Rubin rely primarily on developments from GAFAMs that require very specialized expertise, both for their implementation and maintenance. The Council recommends that the management makes sure that choices made at the CC-IN2P3 (on-premise system) remain compatible with DRP developments.

Moreover, it is crucial to monitor, maintain and develop expertise on those quickly evolving technologies, in order for IN2P3 to remain a key player in Rubin.

- The DESC computing needs at the CC-IN2P3 should not be overlooked, even if they represent a low fraction of the LSST needs, as they condition a strong contribution to the science.
- The Council reiterates the recommendation made in October 2020 to push for a data-sharing agreement between Rubin/LSST and Euclid, to realize the promising scientific synergy between the two projects.
- Computing and storage resources for Euclid at the CC-IN2P3 are funded by CNES, through an agreement with IN2P3 that covers the estimated needs up to the first data release. The needs are currently being reassessed, and there is an on-going discussion to extend the agreement in order to cover the full mission lifetime. Human resources funded by CNES are also essential.
- The computing and storage resources required to produce Euclid's cosmology results have not been taken into account in the SGS requirements. It is important to pay close attention to their on-going evaluation.

6. Astroparticules : Les développements pour le traitement des alertes basses latences

6.1 Présentation

L'approche multi-messagers et multi longueurs d'onde pour l'étude des phénomènes transitoires est une discipline en rapide évolution. Les dernières années ont été caractérisées par les premières observations conjointes de sources en utilisant différents messagers tels que les photons, les ondes gravitationnelles et les neutrinos. Des efforts importants sont déployés pour développer des systèmes de suivi afin d'identifier rapidement les phénomènes transitoires pour étudier et comprendre la nature des sources et les processus physiques sous-jacents.

De nouveaux instruments plus puissants seront mis en œuvre dans les prochaines années et beaucoup d'efforts ont été déjà fournis dans les différentes collaborations pour la mise en place de systèmes de production, gestion et envoi d'alertes qui devront faire face à un taux d'alertes très élevé (1 To/nuit dans le cas du LSST). Parmi les principaux projets en astroparticules, dans lesquels l'IN2P3 est impliqué et qui permettent de faire ces analyses temporelles, on peut citer Virgo/LIGO/KAGRA, KM3NET, Rubin, SVOM, CTAO et CMB-S4. L'IN2P3 joue un rôle clé avec le développement de systèmes robustes dans ces expériences et dans la mise à disposition des compétences et ressources du CC-IN2P3 qui assurent un retour scientifique fort.

Pour combiner de manière efficace les données venant de ces différents instruments, le principal défi passe par la structuration des communautés au-delà des expériences et le développement d'infrastructures partagées. Ce défi commun inclut la gestion de la réception d'un grand volume d'alertes venant des différents observatoires, accompagnée de

l'extraction et de la classification des sources pour un éventuel déclenchement de nouvelles observations.

Dans ce contexte, et à l'initiative de l'IN2P3 (IJCLab et LPC), l'agent de messages (« broker ») Fink a été conçu dans le but principal de filtrer et de distribuer les alertes Rubin. Il a ensuite été sélectionné par la collaboration comme l'un des sept agents de messages officiels en 2021. Il s'appuie sur des technologies de pointe (Apache Spark, Kafka, Kubernetes...) et sur les services et outils largement utilisés dans la communauté pour bénéficier des standards en place et assurer l'interopérabilité avec différents acteurs multi-messagers. Un prototype a été développé et déployé sur le « cloud » VirtualData à Orsay pour l'analyse des données du précurseur de Rubin, ZTF, et est en déploiement au CC-IN2P3 pour Rubin. Fink sera également utilisé par la mission SVOM, en particulier pour déclencher son segment sol et par le réseau GRANDMA. Fink compte de nombreux experts venant des différents domaines scientifiques reliés à l'étude du ciel transitoire.

Il existe plusieurs initiatives travaillant à la structuration de communautés au-delà des projets auxquels l'IN2P3 participe. On peut en particulier citer GRANDMA (« Global Rapid Advanced Network Devoted to the Multi-messengers Addicts »), un réseau de télescopes avec l'objectif de découvrir et caractériser les contreparties électromagnétiques des ondes gravitationnelles. Le réseau est géré par un système central qui distribue les alertes et les séquences d'observation souhaitées à chaque observatoire de manière coordonnée, afin de maximiser la couverture.

Au niveau du partage d'information sur les phénomènes transitoires, il existe des initiatives intéressantes telles que ASTRO-Colibri, initié par l'IRFU/CEA Paris-Saclay. Ce pipeline automatique de partage de l'information des détections multi-messagers et multi longueurs d'onde à travers différentes interfaces, permet par exemple des notifications en temps réel sur téléphone, informant un grand nombre de personnes. Des efforts sont aussi faits au sein de l'IN2P3 pour faciliter l'analyse de ces données multi-messagers et multi longueurs d'onde avec par exemple l'interface web MMODA (« Multi-Messenger Online Data Analysis »).

L'interopérabilité indispensable des différents instruments et télescopes n'est rendue possible que grâce à la définition et à l'adoption de normes communes sur le langage et la technologie. Des organisations telles que le « Virtual Observatory » (VO) cherchent à créer des standards techniques afin de fournir un cadre pour les lots de données astronomiques, les outils et les services permettant de travailler ensemble de manière transparente. Plusieurs expériences majeures ont adopté les normes VO, mais il existe de nouveaux défis liés à la croissance de la taille (à l'échelle du pétaoctet) et de la complexité des lots de données.

6.2 Avis et recommandations

Le Conseil félicite les équipes pour leur positionnement judicieux dans le développement d'outils comme Fink. La sélection de Fink par Rubin est très encourageante et son utilisation par SVOM témoigne de son potentiel. Le Conseil recommande à la direction de veiller à garder un positionnement de premier plan dans ces développements en rapide évolution.

Pour soutenir ces développements et leur assurer une visibilité sur le long terme, le Conseil recommande que soient soutenues des initiatives permettant plus d'échanges et de

développements communs entre les différents acteurs impliqués dans l'étude du ciel transitoire. Il s'agit de maximiser le retour scientifique des développements informatiques pour lesquels l'institut joue un rôle majeur aujourd'hui (sur l'exemple de ce qui a été mis en place ces dernières années par différents acteurs du domaine, la « Cosmostatistics Initiative », ou les ateliers de travail « Transient Sky 2020 » et « Low-latency alerts and data analysis for Multi-messenger Astrophysics »). Il est notamment important que les équipes de l'IN2P3 soient impliquées dans les discussions sur la définition des normes nécessaires à la mise en place de l'interopérabilité, au moment où les questions se posent pour passer à l'échelle du petaoctet.

Les besoins en HPC sont croissants pour les activités multi-messagers comme pour les grandes expériences d'astroparticules et cosmologie à venir. L'utilisation des mésocentres locaux permet de réaliser les développements nécessaires à la mise en place d'outils pour l'ensemble de ces projets. Cependant, l'accès aux mésocentres par appels d'offre n'est pas compatible avec des travaux de simulation ou de traitement des données sur le long terme, risquant de limiter l'implication et la visibilité des équipes de l'institut dans l'analyse. Des problématiques similaires se présentent également pour d'autres équipes de l'IN2P3 (voir section 2.2). Au-delà de l'identification des besoins à venir et du développement de plateformes d'accès communs à l'ensemble des ressources, travaux déjà engagés, le Conseil recommande que soient menées des négociations avec les mésocentres pour garantir aux équipes de l'IN2P3 un accès minimal aux machines de calcul HPC.

Executive summary

The Council felicitates the groups for their wise positioning in the development of Fink. Fink's selection by Rubin is very encouraging and its usage by SVOM demonstrates even more its potential. The Council recommends that the IN2P3 direction pays attention that a leading position is maintained in these rapidly evolving developments.

To support these developments and ensure their long-term visibility, the Council recommends to support initiatives that encourage exchanges and joint developments between the various groups involved in the study of the transitory sky. This involves maximizing the scientific return from computer developments for which the institute plays a major role today (following the example of what has been implemented in recent years by various players in the field: the "Cosmostatistics Initiative ", the workshops "Transient Sky 2020" and "Low-latency alerts and data analysis for Multi-messenger Astrophysics"). It is particularly important that the IN2P3 teams be involved in the discussions on the definition of the standards necessary for the implementation of interoperability, at a time when questions arise about moving to the petabyte scale.

HPC needs are growing for multi-messenger activities as well as for future large astroparticle and cosmology experiments. The use of local mesocentres makes it possible to carry out the developments necessary for the implementation of tools for all these different projects. However, access to mesocentres through calls for tenders is not compatible with long-term simulation or data processing work, risking to limit the involvement and visibility of the institute's teams in the data analysis. Similar problems are present also for other teams in IN2P3 (see section 2.2). Beyond the identification of future needs and the development of platforms for common access to all resources, work that is already underway, the Council

recommends that negotiations should be conducted with the mesocentres to guarantee to the IN2P3 teams a minimal access to HPC computing machines.

7. Les développements logiciels et intergiciels

7.1 Présentation

Les développements logiciels et intergiciels¹ discutés dans cette section concernent les principaux projets dans lesquels l'IN2P3 s'implique de manière importante, dans les différents domaines scientifiques de l'institut : physique des particules, astroparticules, physique nucléaire et analyse biomédicale. Ces projets sont regroupés selon trois catégories : 1) Simulation ; 2) Traitement de données et analyse ; et 3) Gestion des travaux, données et métadonnées.

7.2 Présentation des logiciels de simulation

Simulation Geant4 (« GEometry ANd Tracking »)

L'IN2P3 est impliqué depuis plus de 25 ans dans les développements de la boîte à outils Geant4 pour la simulation du passage des particules dans la matière. Ce logiciel est largement utilisé en physique des hautes énergies et en physique nucléaire, mais également dans le domaine médical, spatial et dans d'autres domaines. Ce projet, initié en 1994 suite au projet RD44 du CERN consistant à réécrire GEANT3 en C++, voit l'implication de l'IN2P3 dès l'origine. Les défis du futur sont essentiellement liés au HL-LHC et nécessitent des évolutions à la fois pour améliorer le code sur CPU (évolution adiabatique), mais également sur GPU, dans le cadre de deux projets de R&D au CERN et aux USA. L'implication de l'IN2P3 est aujourd'hui de 10 personnes dans quatre laboratoires, pour un total d'environ 5 ETP (sur 30 ETP au niveau international) avec une forte implication du CERN. Les activités des équipes de l'IN2P3 concernent les modules de physique électromagnétique, la simulation rapide, la visualisation. Elles contribuent à la géométrie et dirigent la partie exemples. Un tutoriel annuel est organisé chaque année avec les autres partenaires français.

En tant que projet fortement intégré à Geant4, le projet **Geant4-DNA**, pour la simulation des effets biologiques des radiations ionisantes, initié en 2001 par l'ESA, est dirigé par l'IN2P3 depuis 2008. C'est le premier code entièrement accessible dans ce domaine où de nombreux codes propriétaires existent, et c'est la raison pour laquelle il est devenu une plateforme de référence pour la communauté. Cette communauté participe largement aux tests du code, dont les retours sont une source d'amélioration et d'idées pour le développement de nouvelles fonctionnalités. Il s'appuie à l'IN2P3 sur une communauté de 10 permanents, 6 ETP et 3 thèses dans 4 laboratoires sur un total de 60 collaborateurs au niveau international. Le retour scientifique est fort avec plus de 120 publications.

Le projet **GATE** (« Geant4 Application for Emission Tomography ») pour les simulations en imagerie médicale et radiothérapie, qui s'appuie sur la boîte à outils GEANT4, est la plateforme de simulation Monte-Carlo ouverte la plus populaire en imagerie TEP et SPECT

¹ On appelle "intergiciel" un logiciel utilisé par différentes applications pour communiquer entre elles.

avec plus de 2000 utilisateurs enregistrés. L'IN2P3 est le contributeur principal, avec 8.4 FTE distribués dans 6 laboratoires (sur 19 FTE au niveau français incluant aussi l'INSERM et le CEA). L'IN2P3 en est le porte-parole depuis 2008. De nombreux développements sont à prévoir dans les 5 ans à venir afin de garder la plateforme compétitive, en particulier pour une encapsulation en python avec les autres projets GEANT4 et l'intégration des techniques d'apprentissage automatique.

Simulation nptool (« Nuclear Physics Tool »)

Ce projet initié en 2008 est destiné à la simulation et l'analyse des expériences de transfert à basse énergie. C'est un logiciel générique, indépendant de la physique et du détecteur, qui s'appuie sur les possibilités de root et de GEANT4, mais qui ne nécessite qu'une formation minimum des utilisateurs. Chaque détecteur (plus de 70 détecteurs possibles) est mis en œuvre en tant que module externe et de nombreuses bibliothèques sont fournies pour aider le physicien dans son analyse. Ce logiciel est largement utilisé au niveau international dans toutes les installations ; le seul logiciel concurrent est dédié à l'installation FAIR en Allemagne. Il y a 6 développeurs dont 1 ETP à l'IN2P3 (2 FTE pour la collaboration internationale). Des évolutions sont prévues pour faciliter la gestion des modules détecteurs et la maintenance du code.

Simulation Smilei (« Simulating Matter Irradiated by Light at Extreme Intensities »)

Le logiciel ouvert Smilei, qui simule la physique de l'interaction laser plasma, est basé sur la méthode standard « Particle-In-Cell » (PIC). La description complète de la physique des faisceaux accélérés est un défi scientifique qui nécessite des codes massivement parallèles et les ressources les plus avancées en calcul HPC. Ce projet, qui a débuté en 2013, bénéficie d'un partenariat avec les plus grands centres HPC français et d'une collaboration avec des experts en calcul HPC (Maison de la Simulation, CINES, IDRIS). Le contexte des simulations PIC est dominé par l'émergence de nouvelles architectures matérielles liées aux centres exascale mais également par la nécessité de produire des simulations presque en temps réel. Les développements sont effectués par une équipe de 7 permanents de différents instituts français (INP, INSIS, IN2P3, Ecole Polytechnique et CEA) avec une contribution IN2P3 de 0.9 FTE. Les évolutions envisagées sont le portage de tout le code sur GPU mais également l'optimisation du code pour l'électronique embarquée.

7.3 Présentation des logiciels de traitement des données et analyse

Le projet ACTS (« A Common Tracking Software »)

La reconstruction des traces des particules chargées constitue la partie la plus complexe et la plus consommatrice en CPU du traitement des événements des expériences du LHC. Ceci constitue le défi le plus important dans le cadre de futurs collisionneurs tels le HL-LHC. En s'appuyant sur l'expérience passée, ce projet constitue une tentative d'encapsuler le logiciel ATLAS actuel pour en faire une version générique indépendante de l'expérience et de la structure logicielle utilisée. Il a été initié par le groupe de trajectographie d'ATLAS au CERN dans les années 2010 et aujourd'hui l'équipe compte 10 personnes, principalement au CERN et aux US, dont 1.8 ETP à l'IN2P3 dans deux laboratoires. Le logiciel a été testé sur plusieurs détecteurs, pour Belle II et pour le trajectographe interne d'ATLAS (ITK). En parallèle, des activités de R&D sont en cours selon deux axes : le portage sur GPU, avec une première version prévue pour la fin 2022, l'apprentissage automatique pour la reconstruction des traces (ANR ATRAPP). Deux post-doctorants supplémentaires sont en

cours de recrutement ainsi que les travaux d'un troisième laboratoire sur les algorithmes de « Graph Neural Networks » en collaboration avec l'université de Stanford. Sur la partie portage sur GPU, d'autres développements concurrents sont en cours mais spécifiques à une expérience : le projet Allen pour LHCb et O2 pour ALICE. Les essais de collaboration en particulier avec LHCb se sont heurtés à une difficulté d'utilisation dans la structure logicielle de LHCb.

Analyse astrophysique Gammapy

Gammapy est un logiciel ouvert, écrit en python, d'analyse en astrophysique pour des données de haut niveau produites par des télescopes à rayon gamma du GeV au TeV tels HESS, CTA, HAWC ou Fermi-LAT. Ce logiciel a été retenu par l'observatoire CTA (CTAO) comme librairie principale qui sera fournie à la communauté. C'est une librairie autonome, complètement intégrée à l'écosystème python, qui utilise de nombreuses bibliothèques de la communauté et qui peut s'exécuter à la fois sur un PC ou un cluster. Deux chercheurs permanents de l'IN2P3 assurent la gestion de projet.

Traitement des données AGATA

AGATA est un projet européen, initié il y a 10 ans, de spectrométrie gamma visant la construction d'un détecteur couvrant un angle solide de 4π . Deux algorithmes principaux sont en cours de développement : le premier appelé PSA pour « Pulse Shape Analysis » (très consommateur de CPU) est destiné à fournir l'énergie déposée dans les cristaux de Germanium, alors que le second est un algorithme de reconstruction de traces. L'objectif, qui n'est pas encore atteint, est de permettre le traitement en ligne et en temps réel des données. Une des solutions évoquées pour l'atteindre serait l'apprentissage automatique. En tant que détecteur nomade ayant déjà collecté des données, avec une configuration qui évolue régulièrement, le détecteur AGATA nécessite une évolution régulière du logiciel qui devra prévoir l'introduction de GPUs et peut-être de FPGA dans le futur. Trois laboratoires de l'IN2P3 travaillent sur ces développements pour un total de 5 ETP qui représentent environ 30% des ressources humaines engagées.

Boîte à outil d'analyse ions lourds Kaliveda

Le projet Kaliveda, initié en 2002, est une structure logicielle ouverte basée sur ROOT et écrite en C++, qui fournit tous les outils logiciels pour la reconstruction des événements, l'identification des particules, la calibration du détecteur et l'analyse des données produites par INDRA et FAZIA. Une équipe de trois développeurs distribués dans trois laboratoires fournissent 0.7 FTE.

7.4 Présentation des logiciels de gestion des tâches de calcul, données et métadonnées

Cette section rapporte les travaux de deux projets, l'un qui concerne la branche française du logiciel de gestion de données et des tâches DIRAC@IN2P3 et l'autre le logiciel de gestion des métadonnées scientifiques pour ATLAS : AMI.

DIRAC@IN2P3 (« Distributed Infrastructure with Remote Agent Control »)

L'interlogiciel DIRAC a été initié en 2002 par le CPPM en tant que solution pour la gestion de données et des tâches dans le calcul distribué de l'expérience LHCb. Le succès de ce logiciel a permis son adoption par d'autres grandes collaborations en physique des

particules (BelleII, BESII et autres...). En 2014 un consortium a été créé pour gérer les développements futurs et la promotion du logiciel avec six partenaires internationaux dont le CERN et le CNRS. Des services ont alors été proposés à différentes communautés dans le cadre d'infrastructures de grille nationales ou européennes (France Grille, l'initiative de Grille Européenne EGI, GridPP en Grande Bretagne, JINR en Russie...). Dans ce contexte, la solution DIRAC a été retenue pour assurer le service de gestion et distribution des tâches de calcul (« EGI Workload Manager », 20 communautés scientifiques et 700 utilisateurs enregistrés), service aujourd'hui intégré dans le catalogue EOSC.

Le projet DIRAC@IN2P3 regroupe 5 laboratoires IN2P3 avec 12 ingénieurs de recherche pour un total de 2.3 FTE ainsi qu'un laboratoire de l'INSERM. Ce projet, en plus d'avoir la responsabilité des opérations du service fourni par EGI, est aussi chargé des évolutions et de la maintenance du logiciel, ainsi que de la généralisation des développements effectués pour des collaborations spécifiques telles LHCb, CTA, FCC et ILC/CLIC.

Gestion des métadonnées AMI (« ATLAS Metadata Interface »)

Bien qu'initialement développé dès 2000 au LPSC pour l'expérience ATLAS, AMI est un écosystème générique et unique dédié aux métadonnées scientifiques, facilement adaptable et compatible avec tout type de base de données. Avec plus de 2300 utilisateurs actifs, ATLAS reste la communauté principale, mais d'autres expériences commencent à l'adopter ou sont en cours d'évaluation du logiciel (comme SuperNEMO). L'équipe est uniquement localisée au LPSC avec 3 ingénieurs de recherche et un coordinateur scientifique pour un total de 3 FTE. Les développements en cours concernent les besoins d'ATLAS pour le Run3 et la préparation du HL-LHC. Un enjeu supplémentaire est la promotion d'AMI vers de nouvelles communautés d'utilisateurs et la recherche de nouveaux développeurs français ou internationaux.

7.5 Avis et recommandations

Les équipes de l'IN2P3 contribuent largement aux projets GEANT4, GEANT4-DNA et GATE, qui sont des codes critiques nécessitant un support appuyé. Cependant, ces codes vieillissent et nécessitent une modernisation dans leur usage du C++ face aux défis du HL-LHC et des autres domaines scientifiques concernés, notamment pour s'adapter aux futures ressources telles les GPU en environnement hybride. Le Conseil recommande aux équipes d'envisager la participation au projet AdePT (« Accelerated demonstrator of electromagnetic Particle Transport ») du CERN sur le portage sur GPU en environnement hétérogène, pour à la fois former les jeunes chercheurs, se préparer aux prochains défis mais également rendre le projet plus attractif. Enfin, les membres du projet de l'IN2P3 vieillissent et sont peu remplacés ou par des membres non permanents, ce qui est incompatible avec un projet de cette durée et les défis à venir. Pour faire face au problème des futurs départs en retraite, le Conseil recommande à la direction de planifier le renouvellement des équipes et la pérennisation des compétences existantes ou nécessaires à l'évolution des codes.

Le nombre d'expériences utilisant le détecteur AGATA avec d'autres détecteurs inclus dans le logiciel nptool augmente progressivement : il serait opportun pour la communauté de Physique Nucléaire de consolider l'interactivité entre les deux logiciels.

Le projet Smilei est un outil généraliste de l'interaction laser plasma. L'intérêt pour l'IN2P3 est de développer les modules et modes de fonctionnement adaptés à la modélisation de la physique des accélérateurs laser plasma, et à terme contribuer à leur conception et leur

pilotage. Pour ce faire, la contribution en ETP de l'institut à ces développements doit être renforcée.

L'implication des équipes de l'IN2P3 dans les projets tels ACTS ou Gammapy, qui consistent en des développements récents prenant en compte un environnement hétérogène, disposant d'un interfaçage avec le langage python largement utilisé par les nouvelles générations (peut-être à intégrer également dans nptool et KaliVeda), est de grande valeur avec une reconnaissance importante au niveau international. Dans le domaine de la gestion des tâches, données et métadonnées, DIRAC est la solution naturelle dans le contexte de l'EOSC et a déjà une large communauté d'utilisateurs. De manière similaire, le Conseil recommande au projet AMI d'explorer la possibilité de jouer un rôle similaire pour la gestion des métadonnées scientifiques au-delà de l'expérience ATLAS.

Executive summary

The IN2P3 teams contribute significantly to the GEANT4, GEANT4-DNA and GATE projects, which are critical codes requiring strong support. However, these codes are ageing and need to be modernized in their use of C++ in order to meet the challenges of the HL-LHC and other relevant scientific fields, in particular to adapt them to future resources such as GPUs in a hybrid environment. The Council recommends that the teams consider participating in CERN's AdePT ("Accelerated demonstrator of electromagnetic Particle Transport") project on porting to GPUs in a heterogeneous environment, in order to train young researchers, prepare for future challenges and make the project more attractive. Finally, the IN2P3 project members are getting older and are rarely replaced or replaced by non-permanent members, which is incompatible with a project of this duration and the challenges to come. To mitigate the impact of future retirements, the Council recommends that the direction be able to plan the renewal of the teams and the preservation of existing skills or those needed for the evolution of the codes.

The number of experiments using the AGATA detector with other detectors included in the nptool software is gradually increasing: it would be appropriate for the Nuclear Physics community to consolidate the interactivity between the two software.

The Smilei project is a general tool for laser plasma interaction. The interest for IN2P3 is to develop the modules and operating modes adapted to the modelling of the physics of laser plasma accelerators, and ultimately to contribute to their design and operation. To do this, the Institute's FTE contribution to these developments must be strengthened.

The involvement of IN2P3 teams in projects such as ACTS or Gammapy, which consist of recent developments taking into account a heterogeneous environment, with interfacing with the Python language widely used by the new generations (perhaps also to be integrated into nptool and KaliVeda), is of great value with significant recognition at the international level. In the area of job, data and metadata management, DIRAC is the natural solution in the context of the EOSC and already has a large user community. Similarly, the Council recommends that the AMI project should explore the possibility of playing a similar role for the management of scientific metadata beyond the ATLAS experiment.

8. Développements logiciels et optimisation du calcul

Suite à une école IN2P3 en 2016 sur le parallélisme en environnement matériel hétérogène, et en partageant le constat du « HEP Software Foundation » (HSF) en 2017, un groupe d'ingénieurs et de chercheurs ont proposé la création du master projet DECALOG avec deux sous-projets, l'un sur la performance, portabilité en environnement hétérogène et la précision (Reprises), et l'autre sur la technologie des containers (ComputeOps).

8.1 Description du volet Développements, portabilité et performances des logiciels (Reprises)

L'objectif principal de ce projet est d'étudier les différentes technologies possibles pour le développement des logiciels à l'IN2P3 et de favoriser celles qui préservent la portabilité, la pérennité et la performance, sans compromettre la précision des résultats. Les différentes actions visent à collaborer avec des chercheurs en informatique et en mathématiques appliquées, à favoriser la participation de physiciens dans le projet, et à contribuer à la création de tutoriaux qui seront utilisés pour le transfert de compétence vers une audience élargie d'ingénieurs et de physiciens. Trois thèmes techniques sont étudiés : 1) la vectorisation des codes pour l'utilisation optimisée des cœurs CPU, qui pose de nombreux problèmes liés aux possibilités de calcul parallèle, aux structures de données, aux versions différentes de CPU et à la complexité du déploiement dans un environnement d'utilisation de sites hétérogènes ; 2) le portage des codes sur architecture GPU et FPGA, avec l'adoption de langages de haut niveau tels SYCL, mais orientés C++ ; 3) la précision des codes, qui est souvent sous-estimée par les physiciens.

Le projet a permis des échanges internes à travers des réunions mensuelles, une contribution importante aux exercices de prospectives de l'IN2P3 avec la création d'un guide en version web qui devrait être disponible dès cet automne, la première démonstration d'un serveur destiné aux formations, des collaborations avec le CEA et l'EDF et finalement 3 thèses en co-direction avec des laboratoires d'informatique. Ce projet regroupe 10 laboratoires, 11 ingénieurs et 1,5 ETP.

8.2 Description du volet Optimisation du calcul : déploiement et ordonnancement

Créé en 2018, ce projet étudie les technologies de « conteneurisation » en environnement hétérogène (CPU, GPU, FPGA, ...) dans un but de portabilité et de reproductibilité, en évaluant les différentes solutions techniques, intégrant également un aspect de formation et de transfert de compétence vers la communauté IN2P3. Les différents travaux en cours sont la comparaison des différentes technologies de conteneurs (Singularity, docker, Apptainer), des orchestrateurs de conteneurs (qui permettent l'automatisation de la plupart des efforts opérationnels nécessaires pour déployer et exécuter des travaux et services « conteneurisés », comme Kubernetes, Nomad), l'interopérabilité des technologies, la compatibilité avec la grille et les bénéfices d'une organisation de type « DevOps » qui rapproche les développeurs des administrateurs système en charge de l'exploitation dans le milieu de la recherche. Ce projet a organisé ou participé à plus d'une dizaine d'ateliers et formations sur le sujet. Il a également conteneurisé des applications pilotes et déployé un service de « Singularity Hub », c'est à dire un catalogue de conteneurs, pour distribuer des applications en conteneur pour la recherche, projet aujourd'hui remplacé par le projet CSAN

(« Comprehensive Software Archive Network ») qui a pour objectif de fournir à la communauté de l'Enseignement Supérieur et de la Recherche un catalogue d'applications scientifiques ouvert clé en main, avec des codes prêts à l'exécution sur les centres de calcul nationaux ou sur les mésocentres. Après un premier déploiement à l'APC en tant que preuve de concept, une version de production est en cours de déploiement dans deux sites (IPHC et mésocentre de Montpellier). Ce projet regroupe 14 membres dans 7 laboratoires de l'IN2P3, des partenaires extérieurs (IAS, INRAE, CEA, INS2I, CNES et des universités) et du secteur privé (Sylabs, INTEL).

8.3 Avis et recommandations

Les activités scientifiques et de R&D transverses du projet **Reprises** telles l'utilisation des accélérateurs GPU ou FPGA, la portabilité, l'optimisation et l'Intelligence Artificielle (IA) sont pertinentes et montrent un haut niveau d'expertise clairement reconnu au vu de l'implication de chercheurs dans des instances internationales de haut niveau. Le Conseil recommande que les équipes s'impliquent dans des collaborations plus importantes encore avec le « HEP Software Foundation » (HSF), mais également avec les grands projets de la Commission européenne, les initiatives internationales en HPC, IA et calcul quantique, et avec des collaborations entre différents domaines scientifiques. Le Conseil recommande que la direction apporte également un soutien au financement de ces activités et veille à la reconnaissance des équipes.

Le passage du CPU au GPU (ou aux architectures accélérées en général) et aux architectures hybrides pose un certain nombre de problèmes et soulève des inquiétudes dans la communauté : compatibilité avec les futurs centres de calcul, nécessité d'accéder à des mésocentres, risque de perte ou de réorientation des financements, manque ou perte éventuelle de compétences, divergence possible entre les outils HEP "classiques" et les cadres de science des données plus récents. Le Conseil recommande à la direction de se saisir de ces sujets, qui devraient probablement être abordés à un niveau plus général et global.

Il semble au Conseil que les activités de R&D transverses ne sont actuellement pas suffisamment soutenues pour pouvoir gérer les différents défis du futur. Le nombre de ressources affectées à ces projets, en particulier le projet Reprises, semble aujourd'hui très limité (1.5 FTE). Il semble qu'une plus grande coordination, au besoin planifiée par la direction, serait bénéfique dans certains domaines. Par exemple, comment soutenir des études synergiques sur la durabilité et la persistance des logiciels sur des plateformes qui se développent rapidement, un sujet qui a un impact sur la recherche en sciences fondamentales dans tous les domaines ? Bien qu'un effort significatif ait été produit en ce sens, le manque d'incitation à investir des ETP dans ce domaine est inquiétant. Le Conseil recommande à la direction de soutenir plus fermement les activités de R&D transverses, notamment auprès des directions d'unités qui tendent parfois à privilégier les expériences. Retenir les expertises demandera d'offrir des évolutions de carrière attractives pour les ingénieurs et jeunes chercheurs qui s'impliquent dans les activités de logiciel, en particulier avec des postes permanents mais également avec une juste reconnaissance de l'importance de leurs travaux.

Les liens avec les laboratoires de recherche en informatique sont un des points forts de la France et doivent être encouragés. L'offre en formation doit être forte et en lien avec les

défis futurs. Elle devra en particulier prendre en compte une évolution vers le python, même si le C++ restera une expertise de spécialistes.

Le partage de connaissances permis par le projet **ComputeOps** pour le déploiement des ressources dans les centres de calcul est important. Cependant, il serait également judicieux d'étudier une complémentarité avec des techniques telles CVMFS (« CERN Virtual Machine File System ») afin d'offrir une meilleure mise à l'échelle face au choix de la solution du conteneur par l'utilisateur ; les travaux suivants devraient se consacrer à l'orchestration des conteneurs. La collaboration avec d'autres communautés est également à encourager (par exemple l'IDRIS). Le Conseil recommande à la direction le recrutement d'un CDD pour permettre une avancée des travaux plus rapide. Dans le domaine des dépôts de codes, le Conseil recommande l'intégration plus importante ou même complète du projet CSAN dans les initiatives de la commission européenne (projet ESCAPE) en lien avec EOSC.

Executive summary

The scientific and transverse R&D activities of the Reprises project, such as the use of GPU or FPGA accelerators, portability, optimisation and Artificial Intelligence (AI), are relevant and show a high level of expertise that is clearly recognised by the involvement of researchers in high-level international bodies. The Council recommends that the teams get involved in even more important collaborations with the HEP Software Foundation (HSF), but also with the major projects of the European Commission, international initiatives in HPC, AI and quantum computing, and with collaborations between different scientific fields. The Council recommends that the direction also supports the funding of these activities and ensures the recognition of the teams.

The move from CPU to GPU (or accelerated architectures in general) and hybrid architectures raises a number of issues and concerns in the community: compatibility with future computing centres, the need to access mesocentres, the risk of loss or redirection of funding, the lack or possible loss of skills, the possible divergence between "classical" HEP tools and newer data science frameworks. The Council recommends that management take up these issues, which should probably be addressed at a more general and comprehensive level.

It seems to the Council that cross-cutting R&D activities are currently not sufficiently supported to be able to handle the different challenges of the future. The Council recommends that the direction give stronger support to cross-disciplinary R&D activities, especially in accordance with the laboratory directors, which sometimes tends to favour experiments. Retaining expertise will require offering attractive career paths for engineers and young researchers who are involved in software activities, in particular with permanent positions but also with fair recognition of the importance of their work.

The links with computer science research laboratories are one of France's strong points and must be encouraged. The training offer must be strong and in line with future challenges. In particular, it should take into account an evolution towards Python, even if C++ will remain a specialist expertise.

The knowledge sharing enabled by the ComputeOps project for the deployment of resources in the computing centres is important. However, it would also be wise to explore

complementarity with techniques such as CVMFS (CERN Virtual Machine File System) to provide better scaling to the user's choice of container solution; further work should focus on container orchestration. Collaboration with other communities should also be encouraged (e.g. IDRIS). The Council recommends to the direction the recruitment of a fixed-term contract to enable the work to progress more quickly. In the field of code repositories, the Council recommends that the CSAN project be more fully integrated into the European Commission's initiatives (ESCAPE project) in conjunction with EOSC.

9. Les développements en intelligence artificielle issus des projets IN2P3

9.1 Présentation

Une enquête a été menée en mai 2022 afin d'avoir une compréhension globale de l'utilisation de l'Intelligence Artificielle (IA) parmi les équipes de l'IN2P3. D'après cette enquête, 55 projets en cours ont été recensés comme se servant de ce type d'outils. Sont concernés principalement la recherche auprès de collisionneurs et en astroparticules, mais aussi la physique des neutrinos, des accélérateurs, etc. Typiquement, ces utilisations ne nécessitent pas d'importantes ressources de calcul : les ordinateurs portables sont souvent suffisants. Toutefois l'utilisation de GPUs, qu'il s'agisse de ressources locales ou en centres de calcul - par exemple au CC-IN2P3 ou au CERN - permet d'améliorer les performances. Une exception concerne l'utilisation de réseaux génératifs pour remplacer la simulation détaillée (par exemple avec Geant4) : si l'apprentissage automatique permet de réduire le temps de simulation de plusieurs ordres de grandeur, à terme la génération devra tout de même se faire sur des installations conséquentes (mais beaucoup plus d'événements simulés pourront être produits avec le même nombre de cycles CPU).

Les outils IA sont employés à plusieurs stades dans une chaîne d'analyse typique : depuis la simulation, en passant par la reconstruction de particules individuelles sur base des données brutes, jusqu'à l'analyse d'événements complets dans l'extraction des mesures physiques. En fonction du type de données et du but de l'analyse, différents outils sont mieux adaptés : modèles génératifs pour la simulation, réseaux de neurones convolutionnels ou graphiques pour la reconstruction, auto-encodeurs variationnels ou réseaux de neurones profonds pour l'analyse finale.

Les laboratoires de l'IN2P3 (parfois en collaboration avec l'IRFU) organisent des écoles et ateliers sur le sujet, permettant aux personnes intéressées d'acquérir les compétences nécessaires à l'utilisation de ces outils. Le projet THINK (voir section 4) est en place pour aider les utilisateurs à choisir l'architecture la mieux adaptée à leur cas et déterminer la plateforme optimale de développement (CPU ou GPU par exemple). Sont aussi organisées des compétitions amicales (« challenges ») permettant la comparaison directe de différents algorithmes dans un environnement réaliste.

9.2 Avis et recommandations

Ces dernières années, les efforts en IA à l'IN2P3 se sont développés de manière essentiellement "organique," comme un peu partout ailleurs : durant le travail d'optimisation de la simulation, de reconstruction et d'analyse, les physiciens et ingénieurs se tournent vers les outils les plus performants, qui sont aujourd'hui basés sur l'IA. Le moment est

probablement venu d'évaluer si, pour certains aspects du moins, une coordination de ces efforts pourrait être bénéfique. L'exemple évident est celui de l'accélération de la simulation grâce à l'utilisation de réseaux génératifs, en particulier pour le HL-LHC, où les ressources attendues ne suffiront pas aux besoins actuellement prévus par les collaborations (voir section 3).

L'organisation d'écoles et d'ateliers est une initiative très utile qui permet aux utilisateurs d'acquérir les compétences nécessaires plus rapidement. Le Conseil recommande de soutenir cet effort à long terme en particulier pour les nouveaux doctorants et post-doctorants, mais aussi pour les permanents.

Le Conseil note que l'activité des équipes à l'IN2P3 correspond souvent à l'utilisation de techniques d'IA établies, et que les développements originaux en IA ne sont pas majoritaires. D'autre part, les activités IA en tant que telles ne sont pas obligatoirement publiées, et leur visibilité gagnerait à être élargie. Le Conseil recommande de valoriser ces travaux sur le plan académique, ce qui pourrait permettre d'attirer davantage les étudiants, et offrir aux nouveaux docteurs des perspectives de carrière industrielle ou plus appliquées. Le Master « computational physics » en gestation au LLR est une initiative à encourager, de même que les thèses en co-direction ou collaboration avec des spécialistes de l'IA.

Executive summary

Artificial Intelligence (AI) developments within IN2P3, as in most places, have developed mostly “organically” over the past few years: during the optimization phases for simulation, reconstruction and analysis, physicists and engineers turn to the most powerful tools, currently based on AI. The time has now probably come to evaluate how beneficial a more coordinated effort would be, at least for certain aspects. The obvious example is accelerating event simulation by using generative networks, in particular for HL-LHC, where the expected computing resources will not meet the collaborations' expected needs (see section 3).

Organizing schools and workshops is a very useful initiative which allows users to acquire the required expertise quickly. The Council recommends sustaining this effort in the long term, targeting in particular new PhD students and postdocs, while being open to permanent staff.

The Council observes that the IN2P3 teams' activities are often based on established AI techniques, while original AI developments are in the minority. In addition, the AI activities themselves are not often published, and they would benefit from increased visibility. The Council recommends valorizing this work at the academic level, as this may be attractive to students, and open doors in industry or applied science for recent PhDs. A Master's degree in “computational physics” which is being developed by LLR should be encouraged, as well as theses that are co-directed or performed in collaboration with AI experts.

10. L'informatique quantique

10.1 Présentation

L'informatique quantique (QC pour « Quantum Computing ») représente une technologie à haut potentiel pour les domaines d'activité de l'IN2P3, que ce soit pour l'analyse de données ou pour la simulation. De par leur complexité en termes de calcul, ou leurs propriétés quantiques, nombre de domaines de recherche centraux à l'IN2P3 pourraient directement

bénéficiaire de l'avènement de l'informatique quantique, voire en être révolutionnés. Même si les ordinateurs quantiques disponibles ne permettent pas encore de répondre aux besoins réels, l'évolution est rapide. Il est donc plus qu'opportun d'effectuer une veille technologique, et d'explorer ce que ces technologies peuvent apporter aux projets de l'institut.

Les activités liées au QC discutées lors du Conseil sont assez récentes à l'IN2P3. Elles ont été initiées en 2019 et sont structurées depuis 2021 dans le master projet QC2I (« Quantum Computing for the two Infinities »). Celui-ci s'organise autour de trois thèmes : simulation de systèmes complexes (responsable à l'IJCLab), préparer la révolution de l'informatique quantique (responsable au LPC Clermont), l'apprentissage automatique quantique (« Quantum Machine Learning », responsable au LLR). Ce projet réunit actuellement 22 membres de l'IN2P3 (rassemblant chercheurs déjà actifs dans le domaine ou bien intéressés), ce qui représente un groupe déjà conséquent, dont la diversité des expertises est notable puisqu'elle couvre un large spectre de domaines : physique nucléaire, physique des particules, théorie... Outre l'objectif important et générique d'accroître l'expertise intra-IN2P3 sur les questions liées au QC, deux des groupes de travail explorent des algorithmes dédiés : avec des algorithmes d'apprentissage automatique tirant profit des propriétés des ordinateurs quantiques, dont le but est d'améliorer la performance des algorithmes d'apprentissage automatique classiques et leur rapidité ; et avec une approche par calcul quantique du problème à N corps (« Many body ») notamment pour la simulation des noyaux en physique nucléaire, historiquement pionnière à l'IN2P3.

10.2 Avis et recommandations

Actuellement, les ordinateurs quantiques disponibles gratuitement (par exemple ordinateurs quantiques à 5 qubits d'IBM) semblent répondre aux besoins exprimés ; cependant il sera nécessaire le temps venu de financer l'accès à des ordinateurs quantiques plus performants, ou d'établir des collaborations ouvrant l'accès vers les ressources informatiques du secteur privé. Du point de vue du financement, le budget alloué par l'IN2P3 au master projet QC2I semble adéquat ; le Conseil recommande à la direction de l'adapter en fonction de l'évolution de l'engagement de membres IN2P3 sur cette thématique qui est en croissance. Le Conseil note également que ces thématiques ont un fort potentiel de financements extérieurs, de par leur intérêt stratégique et interdisciplinaire, comme en témoigne le financement 80PRIME de la MITI (SNAC-QC, Simuler les Noyaux Atomiques et leurs Constituants avec des ordinateurs quantiques), obtenu conjointement avec l'INS2I pour les activités sur le problème à N corps.

Même si ce projet est relativement jeune, et les besoins de l'IN2P3 sont plutôt spécifiques, les recherches en QC sont en plein essor actuellement. De nombreux acteurs nationaux et internationaux sont très actifs, et font progresser la discipline à un rythme élevé. Il existe par exemple de nombreuses fédérations au niveau régional (comme le « Paris Center for Quantum Computing », ou le Domaine d'intérêt majeur en technologie quantique (SIRTEQ) pour la région Ile-de-France), au niveau national avec le GDR IQFA (Ingénierie Quantique, des Aspects Fondamentaux aux Applications), ou bien au niveau européen avec l'Initiative Technologique Quantique du CERN. Dans ce contexte, et dans la lignée des conclusions aux exercices de perspectives de l'IN2P3, le Conseil encourage les équipes IN2P3 à poursuivre leur ouverture vers ces initiatives afin d'identifier de nouveaux axes de recherche prometteurs, bénéficier de l'expertise acquise dans ces réseaux, et garantir une visibilité tant aux recherches menées à l'Institut qu'aux jeunes chercheurs impliqués.

Executive summary

Currently, quantum computers freely available (e.g. 5-qubit quantum computers from IBM) seem to meet the needs expressed; however, it will be necessary when the time comes to finance access to more efficient quantum computers, or to establish collaborations opening access to computing resources from the private sector. From a funding point of view, the budget allocated by IN2P3 to the QC2I master project seems adequate; the Board recommends to the management to adapt it according to the evolution of the commitment of IN2P3 members on this theme which is growing. The Board also notes that these themes have strong potential for external funding, due to their strategic and interdisciplinary interest, as evidenced by the 80PRIME funding from the MITI (SNAC-QC, Simulating Atomic Nuclei and their Constituents with quantum computers), obtained jointly with the INS2I for the activities on the many-body problem.

Even if this project is relatively young, and the needs of IN2P3 are rather specific, research in QC is currently booming. Many national and international players are very active, and are advancing the discipline at a high pace. For example, there are many federations at the regional level (such as the "Paris Center for Quantum Computing", or the Domain of Major Interest in Quantum Technology (SIRTEQ) for the Ile-de-France region), at the national level with the GDR IQFA (Quantum Engineering, from Fundamental Aspects to Applications), or at European level with CERN's Quantum Technology Initiative. In this context, and in line with the conclusions of the IN2P3 forecasting exercises, the Board encourages the IN2P3 teams to continue their openness to these initiatives in order to identify new promising areas of research, benefit from the expertise acquired in these networks, and guarantee visibility both for the research carried out at the Institute and for the young researchers involved.

11. EOSC (« European Open Science Cloud »)

11.1 Présentation

L'EOSC (« European Open Science Cloud ») est un cadre mis en place en 2018, issu d'une série d'appels à projets européens et appuyé par une volonté politique au niveau de la Commission européenne, pour aider au développement de la science ouverte. Pour répondre aux exigences de la science ouverte, les données doivent être « FAIR » (« Findable, Accessible, Interoperable, Reusable data »), c'est-à-dire de données faciles à trouver, accessibles, interopérables et réutilisables. L'ambition d'EOSC est de fournir un environnement ouvert, bien structuré, destiné aux chercheurs mais aussi aux industriels et plus généralement aux citoyens, qui puisse favoriser la production de science ouverte au service de la recherche, l'innovation et l'éducation. L'échange de données entre les communautés scientifiques est un des objectifs annoncés par EOSC, avec un bénéfice attendu sur des sujets ayant un fort impact sociétal (comme cela est déjà clairement apparu pendant la crise de la COVID-19), comme l'énergie, la santé, le climat, l'environnement. Les objectifs de l'EOSC sont aussi de favoriser l'intérêt des citoyens envers la science, ainsi que leur contribution à la production de résultats (science « citoyenne » ou « participative »).

L'IN2P3 s'est impliqué dans les projets de grilles de calculs et d'infrastructures à l'échelle nationale ou européenne qui ont émergé dans les années 2000, motivés par les besoins d'échange de données et de moyens de calculs de la physique des particules, mais aussi

des sciences du vivant, par exemple. L'IN2P3 a ensuite été partie prenante dans la fédération EGI (« European Grid Infrastructure »), créée en 2010 et regroupant différentes infrastructures européennes (France Grilles pour la France). Il est prévu que ces services soient intégrés dans EOSC. EOSC bénéficiera aussi des actions menées par le projet OpenAire (« Open Access Infrastructure for Research in Europe ») qui a été créé pour améliorer la transparence et l'accessibilité des données académiques.

Récemment, la construction d'EOSC s'est appuyée sur plusieurs projets financés par la Commission européenne, qui ont servi de démonstrateurs pour les aspects d'interopérabilité de différents services et ont permis de définir une architecture pour EOSC. En 2021, le projet EOSC-Future a été lancé pour regrouper les réalisations précédentes et proposer une première réalisation du projet global, associant d'une part les infrastructures européennes de calcul et de partage de données (EGI, « EUropean DATa », OpenAire, GEANT (<https://geant.org>) et « Research Data Alliance »), d'autre part les communautés scientifiques organisées autour de cinq regroupements thématiques. EOSC se présente ainsi comme une fédération de services et de catalogues de données provenant de chaque communauté, se basant sur les principes et les protocoles FAIR. L'architecture d'EOSC est organisée autour de quatre tâches principales : (i) développer les services nécessaires pour faire fonctionner EOSC (« EOSC Core »), (ii) répertorier et intégrer dans EOSC les services et données proposées par les infrastructures de recherche ou les regroupements scientifiques (« EOSC exchange »), (iii) fournir des instructions pour leur dépôt dans EOSC (« EOSC interoperability framework »), (iv) fournir support et formations. L'IN2P3 est impliqué dans le regroupement thématique ESCAPE (pour « European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures ») qui regroupe des infrastructures de recherches européennes dans les domaines de l'astrophysique, des astroparticules, des ondes gravitationnelles, de la physique des particules et nucléaire. L'objectif d'ESCAPE est de construire des services qui s'appuient sur un lac de données (« data lake », un système de gestion et stockage de grands lots de données hétérogènes à l'échelle Exa-Byte) et un répertoire de logiciels. Dans la pratique, ESCAPE est organisé avec cinq grandes structures : un répertoire des logiciels et services fournis par les différentes infrastructures de recherche, une plateforme flexible pour l'analyse des données en accès libre, une archive des données astronomiques, une infrastructure de stockage de données ouvertes prenant en compte le stockage à l'échelle Exa-Byte et un portail dédié à la communication et aux actions de science citoyenne.

Le regroupement thématique ESCAPE pourra fournir des services informatiques, comme par exemple son lac de données. Les collaborations d'analyse scientifique pourront utiliser des outils fournis par EOSC Core, comme des services d'AAI (« Autorisation, Authentification Infrastructure »). ESCAPE va aussi publier ses standards, ce qui peut être important pour l'utilisation des données dans d'autres domaines.

La gouvernance d'EOSC est tripartite, avec une représentation de la Commission européenne, le comité de pilotage composé des États-membres ainsi que l'association EOSC, composée de plus de 200 membres et observateurs (organismes de recherche, fournisseurs de services, organisations non gouvernementales...), qui coordonne les projets financés pour la construction d'EOSC.

11.2 Avis et recommandations

EOSC est un projet de très grande envergure, dont la pérennité n'est cependant pas assurée. Il a été financé jusqu'ici dans le cadre des projets Horizon 2020 et Horizon Europe, mais le modèle financier à long terme reste à définir. Le Conseil recommande à la direction de rester vigilante sur l'adéquation entre les ressources à investir dans un tel projet en évolution et le bénéfice attendu pour la communauté scientifique de l'IN2P3.

Certains bénéfices attendus d'EOSC sont clairs. L'aspect collaboratif d'EOSC devrait favoriser les échanges d'expertise et conduire à une utilisation plus efficace des outils informatiques de stockage et gestion de données scientifiques. La mise en commun de ressources peut aussi permettre plus facilement l'émergence de nouveaux projets. EOSC mettra aussi à la disposition des utilisateurs des services communs, comme par exemple des services de « notebook » comme « Jupyter », ou des services d'AAI. EOSC semble aussi être un outil nécessaire pour l'harmonisation des stratégies vers la science ouverte. Il y a cependant un risque réel que la communauté de l'IN2P3 n'adopte pas EOSC. En effet, le projet n'émane pas de la communauté scientifique mais de la Commission européenne et doit donc démontrer qu'il répondra à ses besoins spécifiques. Pour garantir que ce sera le cas, une implication dans le pilotage de la construction d'EOSC est nécessaire. Pour le moment, à part quelques exemples (notamment LSST et CTA), peu d'utilisateurs de l'IN2P3 se tournent vers EOSC, via ESCAPE. Le Conseil recommande donc de renforcer l'information et la pédagogie autour des initiatives intégrant et utilisant EOSC et d'encourager une plus forte participation à ces projets. Cependant, pour éviter que l'intégration dans EOSC ne s'accompagne d'un surcroît de travail, il faudra que la production de nouveaux logiciels, de services ou de données soit prévue dès l'origine pour s'intégrer dans ce cadre.

Le Conseil recommande également de veiller à la flexibilité du cadre proposé pour EOSC. Il doit être en effet capable de suivre les évolutions très rapides des techniques et des besoins des utilisateurs.

Executive summary

EOSC (European Open Science Cloud) is a very large-scale project, but its sustainability is not guaranteed. It has been funded so far under the Horizon 2020 and Horizon Europe projects, but the long-term financial model remains to be defined. The Council recommends that the IN2P3 management remain vigilant about the adequacy between the resources to be invested in such an evolving project and the expected benefit for the IN2P3 scientific community.

Some expected benefits of EOSC are clear. The collaborative aspect of EOSC should allow exchanges of expertise and lead to more efficient use of computer tools for storing and managing scientific data. The pooling of resources can also facilitate the emergence of new projects. EOSC will also make common services available to users, such as “notebook” services like “Jupyter”, or AAI services. EOSC also seems to be a necessary tool for the harmonization of strategies towards open science. However, there is a real risk that the IN2P3 community will not embrace EOSC. Indeed, the project does not come from the scientific community but from the European Commission and must therefore ensure that it will meet its specific needs. To ensure that this will be the case, involvement in steering the construction of EOSC is necessary. For the moment, apart from a few examples (in particular LSST and CTA), few IN2P3 users turn to EOSC, via ESCAPE. The Council

therefore recommends strengthening information and education around initiatives integrating and using EOSC and encouraging greater participation in these projects. However, to prevent integration into EOSC from being accompanied by additional work, the production of new software, services or data will have to be planned from the outset to fit into this framework.

The Council also recommends ensuring the flexibility of the framework proposed for EOSC. It must indeed be able to follow the very rapid changes in techniques and user needs.

12. Discussions internes au CSI

12.1 Discussions avec la direction

Discussion sur la séance présente :

La direction souligne que les discussions lors de cette séance étaient très intéressantes et correspondaient bien aux questions posées. Le Conseil note que le paysage des projets a déjà évolué, certains projets ayant avancé par rapport aux plans annoncés lors des exercices de prospectives de l'Institut (GT09).

Une politique de l'IN2P3 est déjà en place concernant les parties thématiques et R&D transverse examinées en séance, mais il y a une réflexion à mener concernant la politique de l'Institut sur le calcul hors de ces projets. La direction note qu'il faut continuer de favoriser la diffusion de la culture informatique dans l'Institut. Le Conseil demande à la direction d'encourager l'ouverture des équipes de l'IN2P3 vers les autres laboratoires du CNRS à l'international. Enfin, la question de comment mieux valoriser la R&D effectuée dans les expériences, par des publications, se pose et est importante.

Prochaine séance du CSI IN2P3, 27-28 octobre 2022 :

La prochaine séance comportera trois sujets pour avis du Conseil et un point d'information :

- Contributions de l'IN2P3 à Hyper-Kamiokande, faisant suite au point d'information lors de la séance d'octobre 2021,
- Implication de l'IN2P3 sur la jouvence du détecteur de vertex de Belle-2 (et participation à la gouvernance de l'expérience),
- Participation à la jouvence du trajectographe d'ALICE pour le Run3 du LHC (pas d'examen des projets de jouvence pour LHCb et ALICE au HL-LHC pour le moment),
- Point d'information sur la contribution française à EIC à Brookhaven.

Lors de la prochaine séance, la direction donnera aussi un retour sur les actions de l'IN2P3 suite à la séance de février 2021 sur les accélérateurs, ainsi que sur le document de synthèse de l'exercice de prospectives de l'Institut, qui est en train d'être rédigé.

Questions diverses :

Le recrutement de quatre postes de chaire de professeur junior a eu lieu cette année à l'IN2P3 : deux chaires où le candidat serait recruté comme professeur à l'issue des 4 ans de CDD, et deux chaires où le candidat serait recruté comme directeur de recherche à l'issue des 4 ans. La nature "Université" ou "CNRS" des chaires est discutée avec les directeurs d'unité et avec universités pour les chaires universitaires.

Sur un autre plan, la direction signale qu'elle encourage et soutient l'embauche d'apprentis dans les laboratoires.

12.2 Vie du Conseil

Remplacements :

Suite au départ à la retraite de Valérie Givaudan (collège C), membre élu du Conseil, un siège vacant a été affiché au mois de juillet. Après examen des candidatures, Luisa Arrabito a été cooptée par les autres membres élus du Conseil en octobre. La même procédure sera suivie pour le remplacement de Louis Fayard (collège A1) suite à son départ à la retraite en août 2022.

Prochaines discussions avec la direction :

Lors de la prochaine séance, le Conseil aimerait évoquer trois points de discussion avec la direction : la question du nouveau régime indemnitaire des personnels enseignants et chercheurs (dispositif RIPEC), les postes et particulièrement une discussion plus approfondie sur les chaires de professeur junior, ainsi que les conséquences pour l'IN2P3 de la guerre en Ukraine.