



# Le RI3 et le colloque de prospective « Calcul » IN2P3 *juin 2015*

*Remerciements à Luisa Arrabito, Jean-Michel Barbet, Arnaud Beck, Bernard Bouterin, Khalil Chawoshi, Jean-Claude Chevalyere, Michèle Detournay, Eric Fede, Muriel Gougerot, Fabio Hernandez, Dirk Hoffmann, Edith Knoops, Frédéric Magniette, Antoine Perus, Jean-René Rouet, Catherine Seznec pour leurs contributions.*

## Références

- Enquête « Le RI3 et le Colloque Calcul » : <http://cctools.in2p3.fr/survey/index.php?sid=95358>
- Colloque Calcul IN2P3 2015 : <https://indico.in2p3.fr/event/11590/>
- RI3 : <http://informatique.in2p3.fr> , <https://www.projet-plume.org/ressource/RI3>
- Présentation « RI3 : où veut-on aller ? »  
<https://indico.in2p3.fr/event/11590/contribution/10/material/slides/0.pdf>
- CHEP 2015 : <http://chep2015.kek.jp/>
- Ecole IN2P3 2013 « Big Data et No SQL » <http://www.in2p3.fr/actions/formation/Info13/BigData13.html>
- Journées Informatique IN2P3 Irfu (archives) : <http://ji.in2p3.fr>
- Groupes de travail RI3 ( : <http://informatique.in2p3.fr/?q=groupes>
- RGPP – Audit des fonctions support du CNRS (2012) : <http://intranet.cnrs.fr/intranet/dialogue-rgpp/>
- JI2014 <https://indico.in2p3.fr/event/9954/> , Outils collaboratifs gérés au CC pour l'IN2P3  
<https://indico.in2p3.fr/event/9954/session/4/contribution/68> (JR. Rouet)
- LCG-France <http://lcg.in2p3.fr>
- France Grilles <http://www.france-grilles.fr/>
- HSF : <http://hepsoftwarefoundation.org/>
- Programmation parallèle à l'IN2P3  
<https://indico.in2p3.fr/event/9720/contribution/3/material/slides/0.pdf> (David Chamont, Gilles Grasseau)
- CCRI du RI3 : <http://informatique.in2p3.fr/?q=ccri>

## Objet du document

Ce document est une synthèse des points de vue sur un certain nombre de problématiques de l'informatique à l'IN2P3 recueillis par le biais d'une enquête auprès de quelques représentants de la communauté des informaticiens, en amont du colloque de prospective "Calcul" organisé par la direction de l'IN2P3, le 1er et 2 juin 2015.

Les avis ont été collectés entre le 19 et le 28 mai 2015 via un questionnaire en ligne reprenant les grands thèmes du colloque\* et évoquant la vocation du Réseau des Informaticiens de l'IN2P3 et de l'IRFU. Ils ont permis de nourrir l'intervention « RI3 : où veut-on aller ? » faite à l'occasion du colloque. Ils apportent un éclairage intéressant, utile pour l'animation et la définition de l'identité du réseau.

## Finalités et prospectives de l'Informatique à l'IN2P3

### *Défis à venir*

Le défi le plus fréquemment cité est celui lié à l'utilisation de nouvelles architectures processeurs et l'évolution du code nécessaire pour faire face aux besoins de calcul futurs. Il tient à « *la complexité*

\*A noter : Certains sujets de l'enquête (évolution des métiers, outils collaboratifs notamment) évoqués ici, n'ont finalement pas été abordés lors du colloque.

*des architectures processeurs et des paradigmes de programmation associés* » et laisse entrevoir un autre enjeu, humain celui-là : celui d'acquérir et de développer de nouveaux savoir-faire au sein de la communauté des informaticiens. Il va falloir apprendre à « *utiliser des architectures hybrides, comprendre les modèles de programmation et d'exécution, et connaître les outils intégrant nativement ce type d'architectures* ».

Certaines réponses se focalisent plutôt sur la gestion distribuée de nos données scientifiques. Quelques avis s'expriment même en faveur d'une vision globale de l'informatique IN2P3 orientée « Data », focalisée sur la gestion de nos données (gestion distribuée, accès aux données et évolution des infrastructures de stockage notamment). Des réponses prônent de façon assez naturelle de mettre la donnée, la « Data HEP », et pas le « Big Data », au centre de nos préoccupations car tout le reste en découle. Par extension et en dépit du poids du « Calcul », certains d'entre nous suggèrent de refocaliser l'informatique sur la donnée, plus précisément sur le cycle de vie de la donnée. C'est une approche intéressante, plus moderne peut-être ; le « Data center » n'a-t-il pas remplacé le « Centre de Calcul » ? C'est une vision fédératrice qui englobe facilement l'ensemble des thématiques informatiques allant du online jusqu'à la pérennisation et qui permettrait sans doute de mettre en lumière nos spécificités.

Une ou deux réponses abordent les choses sous un autre angle, évoquant des défis plutôt opérationnels comme « *tirer parti de processeurs moins performants mais moins énergivores, plus nombreux mais moins chers* », « *chercher à utiliser les ressources de façon plus efficiente* » quitte à « *agir en amont pour optimiser voire limiter le besoin* », « *éviter les infrastructures disproportionnées* ». Serait-ce le défi d'une informatique raisonnée ?

### *Place du « Big Data » à l'IN2P3*

Une majorité de réponses soulignent le fait que le « Big Data » « *ne correspond pas vraiment à notre problématique* » car il s'agit plutôt de la « *collecte d'un grand volume de données non structurées et/ou sans finalité bien précise, dont l'exploitation, les besoins d'analyse, combinaison, traitement viennent après* ». Rien à voir donc avec notre démarche parfaitement balisée d'analyse des données misant sur la qualité et la pertinence de la donnée et des métadonnées associées.

Il faut « *décliner cette expression générique dans le cadre qui nous concerne* » et analyser « *quelles sont les expériences futures éventuellement concernées* ». Notre domaine n'est pas moteur mais certaines technologies issues du « Big Data » pourraient sans doute nous être utiles. Il faut vraisemblablement collaborer avec d'autres. Aujourd'hui, seules quelques applications de gestion des logs ou de monitoring (grille par exemple) font appel à des technologies (NoSQL DB) apparentées au monde du « Big Data ». Les projets Euclid/LSST sont cités comme d'éventuels terrains d'exploration. Une réponse souligne l'émergence des techniques d'apprentissage et d'extraction automatiques ou « Machine Learning » dans le monde du online, remarquée lors de la conférence CHEP 2015.

En conclusion, on peut se demander s'il convient de faire quelque chose dans le cadre de l'animation RI3 suite à l'École IN2P3 d'informatique « Big Data et NoSQL » de 2013.

### *Interaction avec la « Data Science »*

La variété des réponses et des commentaires ne permet pas vraiment de synthèse. « Data Science », *un autre mot pour « Machine Learning », une spécialité qui se situe à la convergence de différentes compétences (statistique, visualisation de données, optimisation de codes...)*. Les « Data

*scientists* » ne peuvent-ils pas avoir un rôle au sein de l'IN2P3, sur le terrain du « Big Data » justement ?

Hormis lors des Journées Informatique, il n'y a sans doute pas eu suffisamment d'information sur le sujet ou d'occasions d'échanges entre les deux chercheurs en informatique présents au CC et du LAL et les ingénieurs de l'ensemble des laboratoires, pour pouvoir aborder ce thème.

La « Data Science » ou « Science des Données », en partie sous l'impulsion des besoins d'exploitation des Big Data par les géants de l'internet, explose depuis quelques années, en particulier dans sa branche « machine learning », qui pourrait avoir un impact profond sur la manière dont nous analysons nos données scientifiques.

### *HPC à l'IN2P3*

Peu de réponses sur ce sujet. Voici quelques remarques ou suggestions relevées :

- utiliser de manière opportuniste la fraction non négligeable des ressources HPC (~20 %) qui ne sont pas allouées
- recruter des experts HPC à l'IN2P3, ce qui aurait aussi pour effet de favoriser les collaborations avec le monde HPC
- favoriser la formation et notamment les participations aux formations dites PATC (non financées par le CNRS) organisées par PRACE, la structure européenne coordonnant le HPC en Europe et relayées par GENCI, l'acteur HPC français.

### **Collaborations en France et au-delà**

L'intérêt de collaborer au sein de notre communauté IN2P3-IRFU s'impose comme une évidence pour de multiples raisons : « *mutualiser au mieux les ressources, services et infrastructures, définir des priorités et avoir une masse critique sur les sujets importants, développer les compétences en calcul/optimisation, parallélisme, GPU...* ». Une fois admis qu'« *il est indispensable que l'ensemble des forces travaillant en informatique, dans les laboratoires et au CC-IN2P3, travaillent de façon concertée* », les réponses soulignent la nécessité d'une « *coordination de l'informatique au niveau national* » et d'une « *stratégie commune* ».

Collaborer entre instituts Inria, INSI, CNES en France est jugé difficile en « *l'absence d'objectifs, d'outils, de procédures ou standards communs* ». Ce serait possible moyennant « *la mise en place de projets transverses bien identifiés* ». A la lumière d'une expérience positive, une réponse évoque la possibilité de collaborer par le biais de la mobilité des personnels. Une autre souligne les possibilités de collaboration IN2P3-Inria sur les problématiques de performance du code.

A l'international, la priorité est accordée aux collaborations ou organisations comme WLCG, HEPiX et bientôt HSF. Au-delà, l'investissement humain est souvent jugé trop important pour un laboratoire. Il en va de même pour le montage de projets européens, très coûteux en temps ; toujours la règle d'or : « *Avoir la taille critique* » pour avoir des chances de succès. Pour avoir un impact dans les collaborations européennes, il faut avoir une stratégie nationale et, dans certains cas, une sous-structure au niveau français à l'image de LCG-France et France-Grille vis-à-vis de WLCG et EGI.

La participation à des projets Européens semble indispensable, « *puisque une partie significative des budgets arrivent par ce biais-là* ». Cette participation devrait être concertée, et visible comme une contribution IN2P3, et non pas uniquement au niveau d'un laboratoire ou du centre de calcul. L'IN2P3 dispose de personnels avec des profils très techniques, voire trop techniques pour la

rédaction et le management de projets européens. On pense bien évidemment à une « certaine lourdeur ». On note la suggestion selon laquelle « il faut peut-être recruter des profils adaptés ».

## Communauté des informaticiens

### *Evolution des métiers*

L'évolution des métiers ASR (Administration Systèmes Réseaux) est liée à l'optimisation du support des services IT dans les laboratoires associée à une baisse des fonctions support SI, pour reprendre le vocabulaire de l'audit RGPP (Révision Générale des Politiques Publiques) du CNRS.

La mise à disposition des services IT par le CC-IN2P3 ou les Universités permet aux ASR des laboratoires de s'investir dans le déploiement des infrastructures (ressources et services) distribuées de type grille ou cloud. Il est également possible « de réorienter les personnels vers d'autres tâches de soutien », notamment aux utilisateurs de ces moyens informatiques ou aux différents outils collaboratifs et de production de logiciel etc. Externaliser au CC-IN2P3 les services informatiques (messagerie, sauvegarde etc) permet de dégager du temps pour faire de « l'administration collaborative de ressources et services distribués ». Dans un modèle d'externalisation plus poussé », la mutation serait tout autre avec l'obligation de prendre en charge des activités de gestion (sous-traitance, budget) et pilotage (SLA, OLA). On peut y lire l'aveu d'un attachement à la technicité des activités actuelles.

Enfin, dans le domaine du online, est mentionné le rapprochement des compétences en informatique embarquée et en électronique qui est en train de s'opérer. Le monde de l'embarqué rassemble désormais des processeurs et de la logique programmable FPGA sur le même matériel. La distinction entre matériel et logiciel s'estompe au profit de la notion de système embarqué, il faut proposer des profils présentant une double compétence en informatique embarquée et électronique et vice versa !

### *RI3*

Le RI3, réseau des informaticiens de l'IN2P3 et de l'IRFU est d'abord perçu comme un réseau de métier, un « réseau d'entraide et de co-formation », un « organe d'animation de la communauté », ou encore comme une organisation qui « fédère les agents qui ont des activités et compétences similaires et transcende les unités ». Pour certains, le RI3 est une « structure floue » dont on admet bien volontiers ne pas bien connaître les actions ou les rouages, « sans impact clairement identifié », ce qui se justifie en partie par le fait qu'« il n'y a pas d'obligation de s'impliquer ».

On attribue, à juste titre, au RI3 l'organisation des Journées informatique et leur succès. Les attentes portent essentiellement sur l'information et la formation. Concernant la dimension thématique du RI3, les avis sont partagés. Plusieurs réponses font le lien avec la liste des groupes de travail cités sur le site informatique.in2p3.fr en suggérant de les redynamiser. D'autres craignent une redondance avec certains réseaux nationaux.

Tout le monde ne répond pas aux questions « Faut-il plus de concertation et de suivi au sein de l'IN2P3 ? Le RI3 a-t-il un rôle à jouer sur ce terrain ? » La tendance va au « oui mais » le RI3 n'a « pas de rôle opérationnel », « pas de pouvoir de décision », souvent « pas d'objectifs clairs ». Certaines réponses rappellent que le RI3 ne fait pas office de direction de l'informatique et soulignent l'importance d'avoir des contacts fréquents avec le Chargé de Mission Informatique, l'importance de la coordination / direction informatique.

## **Outils collaboratifs à l'IN2P3, « Services nationaux »**

Il y a un consensus quant à l'importance grandissante des outils collaboratifs. Il existe un certain nombre d'outils collaboratifs disponibles au CC-IN2P3 qui ont été présentés lors des dernières Journées Informatique et qui s'apparentent à des « Services Nationaux » (cf. présentation de J.R. Rouet lors des JI2014).

Beaucoup à dire sur ce sujet car il y a des besoins multiples : gestion de temps, organisation projet, réservation de ressources, qualité, production logicielle et dissémination de nos développements, partage de licences, notamment pour des compilateurs, outils d'analyse de performance etc. *« Ces outils doivent répondre à nos besoins mais aussi être utilisables au-delà de notre communauté dans le cadre de collaborations ou projets internationaux ».*

*« Il faut mutualiser les services en production le plus possible au CC pour éviter de disperser les forces ». « Il faut identifier les services que l'IN2P3 doit mettre en place en interne en complément de ceux qui existent par ailleurs CNRS, Renater et qui peuvent être utilisés, repréciser la spécificité et la portée de ces « outils IN2P3 Irfu » et discuter, si il y a lieu, les modalités de responsabilités conjointes entre le CC et les laboratoires ». « Ne pas oublier le modèle pratiqué avec succès : l'IN2P3 (CC-IN2P3) peut être pionnier (web, visioconférence...), force de proposition ou d'évaluation et opérer un transfert vers Renater ou la DSI pour la gestion à grande échelle, le moment venu ».*

*« Il faut adopter une approche transparente et collaborative, les promouvoir au sein de la communauté ».* La mise en place d'une « Task force » proposée par le CC-IN2P3 est considérée comme une bonne idée mais avec des objectifs à préciser, entre veille, force de proposition, évaluation et rôle plus opérationnel pour des services en production.

## **Infrastructures distribuées**

LCG-France, France Grilles : Faut-il deux structures ? Quel est le futur des mésocentres / noeud de grille ? Voici deux questions qui partagent par excellence !

Tous les avis existent concernant LCG-France et France-Grille. *« Une seule structure serait en théorie suffisante mais deux structures se justifient car France Grille s'adresse à d'autres communautés » « Il faut 2 structures mais la partie opération de France Grilles devrait être réorganisée ». « Quid du centre d'opération commun France Grilles/LCG dont il avait été question ? ».* Un avis est *« favorable à une seule structure pour favoriser les échanges, mutualiser les efforts et les compétences au-delà des seules expériences LHC »* avec l'astro par exemple, faisant par là écho aux discussions actuelles concernant l'ouverture de WLCG à des expériences comme BELLE2 ou HESS.

Concernant le futur des mésocentres et/ou nœud de grilles, un avis souligne *« l'erreur de continuer à promouvoir les mésocentres au sein de l'IN2P3 »*, une autre réponse précise que *« cela dépend si on va vers une nouvelle génération de processeurs et si le CC-IN2P3, en tant que futur Tier-1, peut ou non absorber les ressources Tier-2 en termes énergétique et froid ».*

Les autres avis s'expriment en faveur des mésocentres et/ou nœuds de grilles avec différents arguments.

- *« Le besoin subsistera de disposer de fermes orientées analyse et facilité d'accès au plus proches des utilisateurs c.a.d dans les laboratoires ».*

- « Les mésocentres dans les laboratoires fournissent des ressources additionnelles (IT, RH, financières) qui ne seraient pas mobilisées autrement. Ils constituent une communauté d'informaticiens, une somme de compétences avec un dynamisme et un potentiel d'évolution concertée, à même également de collaborer avec le CC ».
- « Les infrastructures distribuées sont préférables et offrent une évolutivité et fiabilité globale bien supérieure, la difficulté de mise en œuvre et le surcoût mis en avant n'étant par ailleurs pas vraiment quantifiables ni quantifiés ».

Une seule réponse tend à différencier les mésocentres des nœuds de grille. « Les mésocentres de calcul locaux (HPC) dont les accès ne sont pas conditionnés par des acceptations de dossier comme GENCI ou PRACE présentent un intérêt. Il ne faut pas négliger le calcul intensif sur supercalculateur et privilégier le calcul sur grille plus que nécessaire à l'avenir ».

### **Animation RI3 : perspectives**

Voici quelques actions possibles et perspectives en faveur de l'animation du RI3, à discuter dans le cadre du CCRI, conseil de coordination du RI3, à envisager en concertation avec l'ensemble des personnes et acteurs concernés.

- Aborder la question de gouvernance de l'informatique lors du colloque et veiller à faire un compte-rendu à la communauté RI3 sous différentes formes (lettre informatique, diffusion du document de synthèse, réunion physique...);
- Veiller à la fréquence des échanges avec le CMI, la direction du CC et le DAS chargé du calcul (en commençant par leur transmettre le présent document avec l'accord des contributeurs);
- Soutenir en priorité l'animation et le fonctionnement du groupe de travail « Parallélisme et Vectorisation » initié début 2014 et repasser à travers l'ensemble des groupes mentionnés sur le site informatique.in2p3.fr ; voir si il faut faire quelque chose dans le cadre de l'animation RI3 suite à l'Ecole IN2P3 d'informatique « Big Data et NoSQL » de 2013 ;
- Proposer dès sept. 2015 la mise en place en concertation d'une Task Force « Outils collaboratifs à l'IN2P3 » avec un mandat, des objectifs, et une échéance ;
- Suivre ce qui se passe au niveau de HSF et en faire écho au niveau de la communauté RI3;
- Valoriser le matériel des écoles informatiques (réfléchir au sein de la cellule « Formation et Evénements » du RI3 à la façon de);
- Etablir des contacts avec les « Data Scientists » ;