



Groupe de travail Infrastructures de recherche

Note sur les particularités des infrastructures pour une recherche dans le numérique

Version internet - 30/04/2015

Les plates-formes numériques, outils et objets d'étude

Par leurs interactions avec d'autres disciplines et grâce à une technologie diffusante, les sciences du numérique couvrent maintenant un très large spectre et ont un impact sur l'ensemble des autres sciences. La technologie y joue un double rôle : à la fois objet d'étude et outil pour enrichir une production scientifique.

Les sciences du numérique ont de plus en plus besoin d'instruments et la part de l'expérimentation devient de plus en plus importante autour du triangle vertueux « modélisation – simulation – expérimentation ». La contrainte de passage à l'échelle à tous les niveaux s'est considérablement renforcée avec l'augmentation vertigineuse des moyens actuels de calcul, l'omniprésence des objets augmentés et la quantité de données expérimentales disponibles.

Certains des instruments de recherche numériques sont à la fois des outils absolument nécessaires pour les autres disciplines et des objets d'études en eux même au sein des sciences du numérique. Enfin, les sciences du numériques ont de plus en plus besoin d'expérimenter sur des environnements réels, en production, sur des objets qui ne sont plus des instruments de recherche : qui imagine qu'on puisse travailler sur les technologies de l'internet sans le faire sur l'Internet avec un grand I?

Alors que le numérique est depuis de nombreuses années un axe prioritaire, fondé sur des sciences dont la partie expérimentale est cruciale, la communauté de recherche dans le numérique a peine à faire financer de façon pérenne les infrastructures qui sont indispensables aux développements de ces disciplines et surtout aux transferts des résultats vers le monde socio-économique.

Les infrastructures nécessaires à la recherche des sciences du numérique ne peuvent pas être analysées avec les même critères que dans de nombreuses autres disciplines. Cette courte note a pour but d'éclairer cette problématique.

Le Groupe de travail des infrastructures de recherche du « numérique » de ALLISTENE a rédigé cette note pour illustrer ce propos.

Infrastructures de recherche numérique de la feuille de route TGIR

Il y a quelques années encore, la feuille de route des Très Grandes Infrastructures de Recherche (les TGIR) du MESR ne possédait pas de catégorie numérique, mais uniquement une catégorie « calcul et réseau », ce qui était très largement reflété par la

liste des infrastructures rattachées. On y trouvait que des moyens de calcul, certes vitaux pour la recherche nationale, mais essentiellement destinés aux disciplines autres que celles du numérique. Renater, seule TGIR non dédiée au calcul, est une TGIR clé pour la recherche française, quelque soit la discipline, car elle garantit aux scientifiques l'accès à un Internet maîtrisé (très haut débit, flexibilité, sécurité) nécessaire à de nombreuses recherches, qu'elles soient dans le numérique ou non.

Les moyens de calcul, regroupés sous la bannière GENCI (Grand Equipement National de calcul Intensif), sont également vitaux pour de nombreuses disciplines de par l'importance la simulation numérique, mais pas spécifiquement pour le numérique qui en fait un usage modéré. Même si quelques équipes de recherche travaillant dans les sciences du numérique sont utilisatrices de ces moyens de calcul, elles restent très minoritaires (quelques % de la ressource calcul). Ces moyens de calcul servent donc essentiellement à de la production (traitement de données, simulation), qui doivent être stables dans la durée et ne peuvent donc être des objets de recherche qu'à une échelle très limitée (e.g. difficile de réserver la totalité d'une machine pour effectuer des tests).

En marge de ces grandes ressources de calcul, deux infrastructures sont venues compléter le paysage dans la thématique des grilles : l'Institut des Grilles devenu entre temps Institut des Grilles et du Cloud (IDGC) et Grid5000 dédiée à l'expérimentation à grande échelle des systèmes distribués.

France Grilles représente la France dans le consortium EGI (European Grid Infrastructure) qui offre un environnement de production basé sur un modèle décentralisé permettant des expérimentations dans un contexte original mais sans donner la main aux chercheurs sur l'instrument lui même.

Seul Grid5000 est dédiée aux expérimentations dans le domaine du numérique, et plus particulièrement pour les systèmes distribués et l'informatique dans les nuages. Grid5000 permet aux chercheurs de prendre la main sur l'instrument et d'y conduire des expériences originales en configurant et instrumentant complètement l'ensemble de l'équipement (protocoles d'échange, virtualisation des systèmes et des réseaux).

Enfin, notons que l'arrivée de ces deux nouveaux instruments, dédiés aux sciences du numériques s'est faite sous le terme d'Infrastructure de Recherche (IR) et non plus de Très Grande Infrastructure de Recherche (TGIR), la différence majeure étant que le financement des IR est du ressort des Alliances, donc ALLISTENE pour l'IDGC et Grid5000.

Dans le périmètre scientifique d'ALLISTENE, notons que la liste des IR du MESR fait référence à Net-Robotique comme une « IR en projet » correspondant au réseau national de robotique. Il s'agit d'une infrastructure de recherche, regroupant différents plateaux techniques répartis sur toute la France. Elle permet un pilotage concerté, des économies d'échelle, la mise en commun de technologies.

Dans le périmètre scientifique d'ALLISTENE, mais au niveau du MESR classé dans Sciences de la matière et ingénierie, RENATECH est également dans la catégorie des IR. Il s'agit d'un réseau national de grandes centrales technologiques dédiées aux nanotechnologies et nanosciences.

Les nouvelles infrastructures de recherche numérique issues du Plan Investissement d'Avenir (PIA)

Dans le cadre des investissements d'avenir, le programme Equipex a pour objectif de doter la France d'équipements scientifiques de taille intermédiaire (c'est-à-dire entre 1 et 20 millions d'euros) dans l'ensemble des domaines de recherche. Bien que la notion d'infrastructures de tailles intermédiaires soit bien adaptée aux sciences du numérique, seuls 8 projets peuvent réellement être rattachés à la thématique du numérique.

Deux d'entre eux viennent renforcer des infrastructures pré existantes :

- Equipex@meso, permet de renforcer un certain nombre de centres de calcul régionaux et de les coordonner avec les équipements nationaux de calcul, le tout piloté par GENCI (Grand Equipement National de Calcul Intensif).
- Leaf vient développer et renforcer certains dispositifs de l'infrastructure Renatech.

L'Equipex Robotex est en quelque sorte une reconnaissance du projet d'IR Net-robotique au sein de la feuille de route du ministère.

Seuls cinq Equipex permettent le développement d'infrastructures numériques de tailles intermédiaires et sont complètement nouvelles. Le spectre thématique de ces nouveaux équipements est également étroit. On trouve deux Equipex, très différents, mais dans le domaine de la vision :

- Digiscope qui permet le développement et la mise en production d'un instrument pour l'étude de la visualisation interactive et collaborative ;
- Kinovis qui permet le déploiement d'une double plateforme de capture et d'analyse avancée des formes en mouvement.

Enfin, le dernier groupe est dans le domaine des réseaux de capteurs communicants et internet des objets :

- FIT qui a permis la mise au point et le déploiement d'une infrastructure pour tester en vraie grandeur les futures technologies de l'Internet, plus spécialement sans fil.
- Amiqua4Home qui est une plate-forme ouverte pour la convergence logicielle et matérielle dans le domaine de l'intelligence ambiante et les services aux citoyens.
- Sense-city permet de travailler sur une chaîne de développement de capteurs innovants impliquant une instrumentation et la communication entre ces capteurs.

On peut donc constater que la liste des infrastructures de recherche numériques reste limitée. L'une des causes est la spécificité des infrastructures dédiées à la recherche dans les sciences du numérique, qui ne correspondent pas au schéma standard des infrastructures de recherche dans les sciences de la matière et du vivant, ou bien d'autres sciences technologiques (e.g. secteur de la micro-électronique). Ces demandes peinent ensuite pour être comprises des pouvoirs publics.

Critères spécifiques des infrastructures de recherche pour le numérique

De la nécessaire évolution de la notion de TGIR

La première constatation est que la définition actuelle d'une infrastructure de recherche, dans sa définition formelle et dans l'esprit avec lequel elle est souvent comprise, n'est pas en adéquation avec les besoins de la communauté des sciences du numérique.

En effet, dans le monde des sciences du numérique, les infrastructures de recherche sont très souvent distribuées – notion qui est maintenant acceptée dans la notion de TGIR – mais surtout très hétérogènes, combinant des objets techniques très variables. Ceci provient du fait que les chercheurs de ces disciplines sont souvent amenés à **étudier des systèmes**, dont le support provient de l'intégration de différents éléments.

Travailler sur l'Internet des objets ne peut pas se faire uniquement sous l'angle des protocoles de communication, mais nécessite une véritable approche systémique à la fois sur les aspects matériels et systèmes embarqués, les aspects communications et protocoles jusqu'aux aspects services et usages.

Une deuxième constatation très importante est le fait que **ces infrastructures de recherche sont elles-mêmes l'objet de la recherche** et non un instrument pour observer, manipuler, fabriquer, etc. Il est difficile d'avoir un catalogue de services bien circonscrits à présenter, à l'instar de ce que proposent les TGIR "standards", chaque expérimentation pouvant nécessiter plusieurs semaines d'adaptation de l'infrastructure de recherche. Dans un tel contexte, il est tout à fait possible d'avoir une approche « service » (au sens capacité pour faire) offert à une large communauté scientifique, mais cela revêt un caractère très différent.

Travailler sur l'assistance à la personne par des moyens informatiques nécessite un lieu de vie que l'on peut configurer en fonction de l'objet de l'expérience. Cela implique des équipements très hétérogènes combinés entre eux. Cette infrastructure est réellement l'objet de l'étude et pas uniquement le moyen d'obtenir des données.

La troisième constatation porte sur la partie immatérielle inhérente aux infrastructures de recherche dans les sciences du numérique. En effet, pratiquement **toute infrastructure de recherche dans ce domaine possède une très forte composante logicielle**, cette composante faisant intégralement partie de la plateforme. De plus, ces logiciels doivent très souvent être développés (ou au moins adaptés à la plateforme) avant que des services puissent être offerts aux équipes de recherche. A l'heure où l'industrie du logiciel est vue comme l'un des secteurs où l'on peut espérer un fort impact économique, il est absolument nécessaire de prendre cela en compte.

Même dans le cas où l'instrument possède une véritable partie matérielle visible, le logiciel reste primordial. Prenons l'exemple d'une salle de réalité virtuelle où l'on combine des équipements achetés sur étagère dans une configuration particulière. Celle-ci nécessite tout de même un environnement logiciel pour opérer l'ensemble du dispositif, mener les expérimentations, qui n'existe pas dans le catalogue des fournisseurs des équipements et qui est l'objet même de la recherche. Si l'on prend le cas d'étude sur la sécurité de l'internet, travailler sur les systèmes de détection d'intrusion nécessite une plateforme à 90% logicielle.

La quatrième constatation porte sur le coût d'investissement et de fonctionnement. En effet, le coût du matériel est nettement plus faible que dans beaucoup d'autres disciplines. Les TGIR acceptent maintenant les infrastructures de petites tailles, mais la

granularité est encore trop grosse pour beaucoup de travaux dans les sciences du numérique. Dans le domaine des TGIR, on considère un centre de recherche clinique ou un laser de forte puissance comme un équipement de petite taille. Dans le monde du numérique, on a très souvent des infrastructures encore moins coûteuses. Mais bien plus que le coût d'investissement en lui-même, il faut souligner que **la part du coût des moyens humains par rapport aux investissements matériels est très importante.**

Regardons en arrière et prenons l'exemple de la robotique : que de progrès techniques réalisés en quelques années. Travailler sur la robotique nécessite maintenant des plateaux techniques où le « robot » n'est que l'un des éléments d'un dispositif plus complexe. Dès lors, on comprend bien que le maintien en conditions opérationnelles de tels plateaux est extrêmement complexe et nécessite des équipes techniques conséquentes. Le coût d'exploitation n'est absolument pas proportionnel au coût de l'infrastructure mais à la complexité de celle-ci et aux logiciels associés.

Les infrastructures de recherche dans les sciences du numérique doivent **s'adapter à un environnement technologique extrêmement dynamique.** Cela nécessite une jouvence très régulière (typiquement tous les 3 ans) voire continue. Au-delà de la simple notion de jouvence, la forme que peut prendre une plateforme technologique varie beaucoup avec le temps. Il ne s'agit pas de durée de vie, qui elle est basée sur l'intérêt de l'objet de l'étude, mais de configuration matérielle et logicielle.

On imagine bien qu'une infrastructure de recherche construite pour étudier les réseaux et l'internet, n'aurait pas la même forme aujourd'hui que celle définie dans les années 2000. Pourtant l'objet de l'étude reste le même et l'évolution peut être réalisée dans un processus continu. De même, pour étudier les systèmes d'intelligence ambiante, il n'est pas forcément nécessaire d'avoir un équipement du dernier modèle, mais il est primordial d'intégrer dans les dispositifs les éléments de rupture technologique au fur et à mesure qu'ils apparaissent. Par exemple, une très forte augmentation de l'autonomie énergétique de certains équipements peut rendre possible certains usages jusque là inenvisageables, nécessitant de revoir les logiciels, l'environnement technologique.

Conclusion

Actuellement, pour une très grande partie des objets d'étude des sciences du numérique, il n'existe aucun moyen de pérenniser un plateau technique de grande ampleur, et donc de le rendre visible pour l'ensemble des acteurs (chercheurs, étudiants, entreprises) – ampleur se mesurant par l'impact attendu de la recherche qu'il facilite – et notamment des équipes d'ingénieurs qui doivent y être associées. Dans le monde du logiciel, on dit souvent que les développeurs ont plus de valeur que le code lui-même. Cette maxime est parfaitement adaptée aux plateformes technologiques des sciences du numérique.

Si l'on souhaite que la recherche dans les sciences du numérique puisse augmenter son impact économique et social, elle a besoin d'instruments pérennes et visibles pour appuyer son expérimentation, sans que ces instruments dépendent de programmes de recherche à très court terme.

Caractéristiques des TGIR pour le numérique

<ul style="list-style-type: none">• Il s'agit de grand réseau d'infrastructures, souvent hétérogènes.
<ul style="list-style-type: none">• Le rapport entre coûts de fonctionnement, de mise à jour d'une part et coûts d'investissement d'autre part est très largement supérieur à ce que l'on peut trouver dans d'autres domaines.
<ul style="list-style-type: none">• Elles sont presque toujours « objets de l'étude » et non outil d'observation uniquement.
<ul style="list-style-type: none">• Elles nécessitent des équipes techniques pointues, en formation permanente, travaillant au sein de réseaux de compétences .
<ul style="list-style-type: none">• Elles ont presque toujours une composante immatérielle forte : le logiciel !
<ul style="list-style-type: none">• Elles nécessitent un travail intégratif important avant de pouvoir être mises à disposition de la communauté.
<ul style="list-style-type: none">• Elles nécessitent presque toujours du développement logiciel conséquent avant la mise en production.
<ul style="list-style-type: none">• Les expériences que l'on peut mener sur ces infrastructures sont rarement « standard » et le coût humain de la mise en œuvre d'une expérimentation est important.
<ul style="list-style-type: none">• Elles doivent absorber une évolution technologique permanente.