



# Rapport de Prospective du Conseil Scientifique de l'Institut des Sciences de l'Information et leurs Interactions (INS2I)

Christian Barillot, Inbar Fijalkow, Isabelle Queinnec, Fabrice Theoleyre,  
Hélène Touzet, Michel Beaudouin-Lafon, Marie-Paule Cani, François  
Chaumette, Gérald Conreur, Véronique Cortier, et al.

## ► To cite this version:

Christian Barillot, Inbar Fijalkow, Isabelle Queinnec, Fabrice Theoleyre, Hélène Touzet, et al.. Rapport de Prospective du Conseil Scientifique de l'Institut des Sciences de l'Information et leurs Interactions (INS2I). [Autre] CNRS. 2018. <hal-01956087>

**HAL Id: hal-01956087**

**<https://hal.archives-ouvertes.fr/hal-01956087>**

Submitted on 21 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

*Ce rapport de prospective est dédié à notre collègue et amie Isabelle Tellier qui nous a quittée au cours de son mandat au sein du conseil*

# Rapport de Prospective du Conseil Scientifique de l'Institut des Sciences de l'Information et leurs Interactions (INS2I)

Christian BARILLOT *Président* ; **Bureau** : Inbar FIJALKOW, Isabelle QUEINNEC, Fabrice THEOLEYRE, *Secrétaire du conseil*, Hélène TOUZET ; **Membres** : Michel BEAUDOUIN-LAFON, Marie-Paule CANI, François CHAUMETTE, Gérald CONREUR, Véronique CORTIER, Cyril GAVOILLE, Guy GOGNIAT, Julien GOSSA, Andreas HERZIG, Jean KRIVINE, Philippe LAMARRE, Françoise LAMNABHI, Anne-Catherine LETOURNEL, Lionel SEINTURIER, Isabelle TELLIER, Serge TORRES, Charlotte TRUCHET, Alexandros TSOUKIAS, Michel VERLEYSSEN

## 1. Objectif du document, méthodologie

La rédaction d'un rapport de prospective scientifique correspond à une des missions des conseils scientifiques d'instituts (CSI) au CNRS. Un tel rapport est à la fois à usage interne au CNRS (directions d'établissement, d'instituts, comité national), et à destination de la communauté scientifique. Dans ce cadre et en concertation avec la direction de l'institut, le conseil scientifique de l'institut INS2I du CNRS a souhaité orienter ce rapport de prospective autour du positionnement des sciences de l'information et leurs interactions dans le paysage scientifique global. Nous souhaitons ainsi pointer l'importance et le rôle de notre discipline dans une période où notre domaine est traversé par de profondes évolutions technologiques, sociétales et scientifiques.

Pour porter cette réflexion, nous avons fait le choix de nous appuyer, entre autres, sur des auditions de personnalités de nos domaines. Ces auditions, menées en petits groupes de discussion, ont donné lieu à des échanges extrêmement riches qui ont nourri la réflexion du conseil dans l'élaboration de ce rapport. Elles ont été conduites sur la base d'une trame commune articulée autour de cinq thèmes :

1. Comment définir notre discipline ? Quelles en sont les grandes réalisations ?
2. Quelles sont les évolutions envisageables dans les 10 à 20 prochaines années ?
3. Quels sont les plus grands enjeux et risques ?
4. Comment notre discipline va-t-elle impacter les autres sciences ? Comment gérer la mobilité thématique dans la recherche aux interfaces avec les autres disciplines ?
5. Comment notre discipline va-t-elle impacter la société au sens large, concernant notamment l'enseignement et l'éducation populaire ?

Nous tenons à remercier vivement les personnes ayant participé à ces ateliers de discussion, dont la liste est fournie à la fin de ce document.



Enfin, la prospective réalisée dans ce document se nourrit également du travail du conseil lors de ces quatre dernières années, en particulier autour de l'organisation de séminaires thématiques. Ces séminaires nous ont permis de porter une réflexion plus spécifique sur une dizaine de thèmes transversaux ou disciplinaires. Beaucoup des enjeux abordés lors de ces séminaires ont également été discutés à travers les ateliers préparatoires à l'élaboration de ce rapport, notamment autour de la place des femmes dans nos disciplines ou bien encore autour des enjeux d'éthique. Les synthèses de ces séminaires sont données en annexe de ce document et les documents de travail sont disponibles en accès libre sur le site internet du conseil (<https://csins2i.irisa.fr/>).

## 2. Introduction : Comment définir notre discipline ?

Les sciences de l'information et leurs interactions couvrent un spectre scientifique très large et très riche. Elles comprennent la « science des ordinateurs » et la transcendent. Le traitement de l'information et la modélisation en constituent des éléments fondamentaux. Nous pourrions même affirmer que les sciences de l'information consistent à modéliser, simuler, observer, analyser et réguler le monde.

Les sciences de l'information représentent également de façon naturelle une « science à l'interface », qui se nourrit des interactions avec les autres disciplines telles que les mathématiques, la physique, la biologie, les sciences humaines et cognitives. Les frontières sont de plus en plus perméables. Ainsi, les probabilités-statistiques sont nécessaires pour l'informatique<sup>1</sup> et inversement. Les mathématiciens conçoivent des outils et un ensemble de théories sur lesquelles ils s'appuient. Les informaticiens doivent eux concevoir des outils, tout en garantissant leurs performances. Si les mathématiques sont le langage historique des sciences, au 21<sup>e</sup> siècle, l'informatique est en passe de devenir aussi un nouveau langage universel des sciences.

L'informatique s'attache enfin à répondre à des besoins concrets, qu'ils soient scientifiques, industriels ou sociétaux en proposant des modèles, méthodologies et outils permettant de mettre en œuvre des solutions innovantes. Grâce en partie à ces avancées, la technologie évolue très vite. Cela amène notre science à interagir avec toutes les facettes de nos vies et avec une rapidité d'évolution unique dans le monde scientifique. Cela engendre des défis "à débits rapides" pour lesquels nous devons développer des stratégies nouvelles pour y apporter des réponses. Cela change la manière traditionnelle de poser les problèmes scientifiques, en pose de nouveaux et conduit notre discipline à être particulièrement réactive et à évoluer très vite.

## 3. Quelles en sont les grandes réalisations ?

### 3.1. Au sein de notre domaine

En raison de sa diversité et de la variété de ses applications, l'informatique se décline en un grand nombre de sous-domaines ou d'écoles de pensées. Les dernières décennies ont été

---

<sup>1</sup> Par souci de concision, dans ce rapport, "informaticien-ne" et "informatique" sont pris dans une acceptation large déclinant tout le spectre disciplinaire de l'INS2I.



marquées par des créations majeures qui impactent aujourd'hui nos champs disciplinaires, mais aussi les autres sciences et la société au sens large. A titre d'exemple, nous pouvons citer :

- Certains problèmes d'IA étaient réputés inaccessibles. Ils ont pourtant été résolus récemment grâce à l'**apprentissage statistique** qui permet d'analyser des données et d'en tirer de l'information utile même quand des modèles physiques n'existent pas ou sont inaccessibles, que ce soit dans des domaines où les données sont abondantes (réseaux sociaux, traitement du langage, reconnaissance d'images, jeux de Go, d'échec, ...) ou d'autres où les données sont rares (questions de médecine personnalisée, etc.) ;
- La **science des données** dépasse les frontières de l'apprentissage statistique en incluant la simulation, le stockage des données et la modélisation, y compris pour de très grands ensembles de données et/ou des algorithmes consommateurs de ressources, pour lesquels l'évolution du matériel et des logiciels informatiques est un élément clé;
- Le comportement des algorithmes eux-mêmes peut être analysé par des outils issus de la **combinatoire** ;
- Dans un monde connecté, les **codes correcteurs d'erreur** (turbo code) permettent de fiabiliser des transmissions dans des environnements bruités et/ou mobiles. Ils ont permis des avancées majeures en termes de débit et de fiabilité ;
- La **cryptographie** représente un élément clé en sciences de l'information, devant assurer un compromis entre confidentialité/inviolabilité et exploitation des données. Les fonctions cryptographiques doivent pouvoir s'exécuter de façon sûre sur une large palette d'équipements, de plus en plus embarqués ;
- Notre domaine scientifique a aussi permis des réalisations majeures autour de la **compression des données, de la sécurité et de la communication**. A titre d'exemple, les gains en compression sont de 50% tous les 5 ans. Même si cette discipline est ancienne, elle continue d'évoluer et conduit aujourd'hui à l'élaboration de standards technologiques majeurs ;
- Dans des systèmes critiques intégrant de plus en plus de logiciels, les méthodes de **preuves formelles** permettent de certifier le comportement d'un système. En garantissant des propriétés de façon formelle, la discipline a permis ainsi une dissémination dans un nombre toujours croissant d'applications ;
- Le **génie logiciel** a permis de rendre la partie logicielle flexible et réutilisable, dans un monde informatique de plus en plus complexe. Un logiciel repose maintenant sur un ensemble de modules, et ne peut plus être développé de façon monolithique ;
- L'**automatique numérique**, au sens de la convergence entre approche mathématique et outil numérique, a permis de faire le pont entre des méthodes très pratiques et des méthodes très formelles (conditions d'existence). Cette combinaison du formel et du numérique permet de résoudre des problèmes de plus en plus complexes, par exemple autour des systèmes embarqués connectés.

## 3.2. En lien avec les autres disciplines

Notre discipline a naturellement permis de grandes réalisations aux interfaces, en lien avec les autres disciplines. Telle une boutade, si nous tentons d'identifier les sciences qui ne sont pas impactées par les sciences de l'information, nous pourrions même répondre qu'il n'en existe aucune. En retour, ce travail aux interfaces vient également **enrichir nos problématiques scientifiques**, en nous aidant à identifier des verrous dans nos propres champs disciplinaires.

**Données scientifiques.** Dans de nombreuses sciences, le verrou est passé de l'acquisition des données à l'analyse des données, ce qui est un changement de paradigme. Dit autrement, le coût de l'analyse des données devient plus élevé que le coût de l'acquisition de ces mêmes données. Gérard Berry résume cela en faisant remarquer que nous sommes passés « *de la matière à l'information* ». Ce paradigme s'applique aussi bien en médecine qu'en biologie, ou en sciences humaines. Le traitement de ces données accumulées requiert des développements en sciences des données, souvent avec de nouvelles questions pour notre discipline. En biologie, par exemple, la nécessité de devoir gérer des quantités croissantes de données de génomes a conduit au développement de nouvelles structures d'indexation de données, telles que des index probabilistes ou des structures succinctes, qui permettent de passer à l'échelle.

**En médecine,** l'informatique a permis la réalisation d'avancées majeures, en particulier en imagerie médicale ou pour la conception de dispositifs implantables, l'élaboration de robots chirurgicaux ou de prothèses intelligentes ou plus généralement autour de la médecine personnalisée. De même, la radiologie bénéficie amplement des avancées en traitement d'image et en apprentissage automatique. Comme dans les autres sciences, le **formalisme** de nos disciplines diffuse également en médecine. Les sciences de l'information reposent très souvent sur une partie modélisation, pierre angulaire pour formaliser le contexte et permettre ainsi d'exécuter un algorithme efficace. Ainsi, en médecine, nous devons modéliser les processus (patient, intervention, contexte). Bien souvent, cette expertise métier est détenue par quelques personnes, qui ont engrangé durant de longues années un ensemble de cas permettant de forger des « règles ». Un formalisme permet notamment de réduire la variabilité.

**En sciences de l'univers,** les sciences de l'information ont joué un rôle majeur dans le succès de la mission Planck, en permettant le recueil et l'analyse des données. Cette mission a permis l'observation du rayonnement fossile de l'univers et constitue une découverte scientifique majeure pour aider à mieux comprendre les mécanismes qui ont présidé à l'apparition des galaxies. L'analyse des données n'a été possible que grâce à l'élaboration de nouvelles approches **de séparation de composantes** dans les signaux.

L'informatique permet maintenant d'attaquer des problèmes complexes via la résolution du **problème inverse**, c'est-à-dire déterminer les causes en partant d'observations. En effet, les observations ne suffisent souvent pas à caractériser entièrement le modèle. De nombreuses disciplines peuvent ainsi s'intéresser à des classes de problèmes plus larges, en s'appuyant sur notre discipline.

Dans le domaine de la **cybersécurité**, et de la science des données au sens large, le besoin d'anonymisation des données dans un nombre croissant d'applications (par ex. médical, défense, Cyber-Physical & Human Systems - CPHS, véhicule connecté) est

confronté à une problématique d'exploitation de ces données. Un compromis entre vie privée et utilisabilité doit être exploré et constitue un verrou scientifique en soi.

### 3.3. Impacts sociétaux (vie sociétale, vie économique)

Grâce à tous ces liens avec les applications, les succès majeurs de l'informatique ont eu un impact très important et souvent rapide dans la société. Les usagers s'y sont très vite habitués, et l'usage en a rapidement été banalisé.

Le déploiement croissant des smartphones a par exemple bouleversé de nombreux aspects de notre vie quotidienne. Nous avons changé notre manière de communiquer, de nous informer, de nous organiser. Nous nous sommes habitués à accéder à un volume croissant d'informations et de manière instantanée. Nous partageons, de manière contrôlée ou non, de plus en plus de données sur notre vie privée. Tout cela modifie nos façons de penser : le temps informatique et le temps humain ne sont plus sur la même échelle.

Les sciences du numérique ont également un impact massif dans la vie économique. Ainsi, le numérique prend une place de plus en plus prépondérante dans un nombre croissant de professions, menant à la « *digitalisation* » des métiers. L'informatisation des entreprises et des administrations est entrée dans nos quotidiens avec des conséquences importantes sur les organisations de ces structures. Ici, l'informatique n'est pas intervenue seule, mais a eu un effet de levier dans la vie courante (sécurité bancaire, analyse biométrique, codage, compression d'images, vidéos, internet, télécommunications, diffusion de contenus, agents conversationnels, etc.).

Une combinaison des nombreuses techniques informatiques (modélisation et simulation numérique, IA, sécurité et preuve des logiciels, etc.) a permis également d'améliorer la fiabilité et la sécurité dans les transports. Ainsi, de nouveaux systèmes d'assistance ou de pilotage automatique ont vu le jour récemment (trains, avions, voitures autonomes, drones, etc.). L'informatique conduit également à la transformation de plusieurs professions, comme celles de notaire ou de journaliste. Dans le domaine de la santé, nous pouvons citer l'essor de l'informatique à l'hôpital pour collecter, analyser et gérer les données de patients. L'exercice de la médecine commence ainsi à en être profondément modifié. Nous assistons même au développement de nouveaux secteurs économiques, tels celui concernant les objets connectés dans les bâtiments intelligents ou encore « *l'économie du partage* ». Les métiers liés aux progrès des sciences de l'information évoluent très vite, et une grande partie des futurs emplois ne sont pas encore connus.

## 4. Enjeux

### 4.1. Pour la discipline

L'évolution des usages dans un monde de plus en plus connecté implique d'avoir accès à des fonctionnalités riches à travers des réseaux variés et des interfaces multiples (du smartphone à l'ordinateur de bureau, de l'objet IoT à l'agent conversationnel et au robot). L'environnement et les systèmes à prendre en compte ne sont donc plus limités au paradigme utilisateur/interface/machine tel un ordinateur, son clavier et sa souris. Ils



deviennent beaucoup plus riches, mais aussi beaucoup plus complexes et dynamiques. L'informatique ubiquitaire, un thème de recherche déjà ancien, est en train de passer dans les usages avec par exemple les systèmes distribués et l'Internet des objets.

Les problèmes scientifiques derrière ces enjeux sont nombreux. Ils incluent par exemple les algorithmes en ligne (capables de traiter les données au fil de l'eau), la vérification, la certification, la transparence et l'explicabilité des algorithmes de traitement des données, le passage à l'échelle et la sécurité des données et des traitements. L'acceptabilité des futurs systèmes dépendra notamment de leur transparence et des niveaux d'explication qu'ils pourront fournir. Par exemple, les succès actuels de l'IA statistique se heurtent à sa capacité limitée d'explication. Aussi faut-il peut-être envisager des approches combinant l'IA symbolique avec des méthodes d'apprentissage profond, afin de répondre à ces enjeux de transparence et d'adaptabilité.

L'optimisation globale avec décision distribuée constitue un des grands défis de notre discipline. Un système doit pouvoir par exemple exploiter les capacités de calcul d'un réseau de capteurs, des méthodes efficaces pour le traitement du signal, l'analyse de données en grandes dimensions, les problèmes inverses, l'utilisation de la géométrie de l'information, etc.

#### 4.1.1. Confiance dans la technologie, responsabilité et preuve

A mesure que la complexité des objets numériques augmente, il est nécessaire d'améliorer la compréhension de ces objets par leurs utilisateurs. Ils doivent ainsi avoir les moyens de comprendre ce qu'ils utilisent et d'avoir une perception claire, étant informés de l'effet de leurs actions et du fonctionnement des systèmes.

D'un point de vue scientifique, nous retrouvons les enjeux liés au passage à l'échelle, à la complexité des systèmes hétérogènes manipulés et à la masse des données produites. Par exemple, le développement d'heuristiques ("boîtes noires") efficaces mais dont le fonctionnement n'est pas certifié représente un des défis majeurs de l'IA. Il lui faut fournir également l'explication de ses résultats ("explicabilité"). Pour accepter de déléguer des prises de responsabilité à de tels systèmes, il faut être en mesure de les comprendre et de les analyser. Si l'on conçoit l'automatisation comme une boîte noire, comment distinguer la responsabilité de l'utilisateur de celle du système ? La question devient de plus en plus complexe avec l'accroissement de l'autonomie des robots et des systèmes, par exemple lors de l'usage d'un véhicule autonome ou d'un robot médical.

La notion de preuve (informatique, mathématique, automatique...) est ainsi un enjeu majeur de l'exploitation des données et des connaissances et du développement des systèmes interactifs et autonomes. Que ce soit pour les développements de logiciels, de contrôleurs et plus généralement d'outils d'exploitation de données, la certification doit être au cœur des développements et devra de mieux en mieux tenir compte de la complexité des systèmes. Lorsque la complexité est telle que toute preuve formelle semble hors de portée ou que les systèmes sont par nature incertains, ce sont des techniques de tests statistiques qu'il faudra développer. Enfin, les systèmes d'interaction humain-machine doivent prendre en compte les capacités humaines et faire en sorte que l'humain ne soit pas soumis aux décisions ou au fonctionnement de la machine, mais que celle-ci au contraire augmente et amplifie ses capacités (de décision, d'action, ...).

#### 4.1.2. Transparence, sécurité et usages des données

L'un des grands enjeux à la fois sociétal et disciplinaire est lié à l'accumulation toujours croissante des données et à leur impact croissant dans tous les domaines de la société. Cette évolution nécessite que la pensée informatique s'accorde avec cette montée en volume et que l'on sache apporter des réponses à des problématiques telles que le passage à l'échelle, l'accès aux données et leur stockage, l'échantillonnage, l'exploration, la représentation de l'information, etc.

Les risques sont importants quant au pouvoir donné à des acteurs industriels critiques et à certaines nations qui ont la capacité de collecter des données et de les utiliser sans préoccupation des aspects liés à la vie privée. A cet égard, les enjeux se situent bien évidemment en cyber-sécurité (informatique des données, etc.) mais aussi dans la sensibilisation de la France et de l'Europe à cette thématique afin de pouvoir développer de nouvelles approches pour la protection des données et de la vie privée. Dans les domaines où l'utilisation des données entraîne des questions d'évaluation du risque (par exemple dans le domaine médical), il est important que les développements incluent les questions de fiabilité des données et des décisions.

Le phénomène récent des "infox" (ou "fake news" en anglais) nous rappelle que la vérité est aussi un droit. De nombreuses applications reposent sur un petit nombre de structures privées mettant en œuvre sans réelle transparence des moyens de calcul et des algorithmes utilisant des données de plus en plus importantes. L'analyse des données et décisions doit pouvoir détecter rapidement les *comportements anormaux*.

Bien que difficile dans certains cas (ex. informatique médicale et plus généralement dans l'apprentissage automatique), l'informatique permet la reproductibilité scientifique. Aussi faut-il encourager les démarches qui vont au-delà de la transparence en fournissant non seulement les informations principales (algorithmes généraux ou code) mais aussi tous les éléments nécessaires à la reproductibilité d'une évaluation expérimentale (jeux de données). C'est une garantie de sérieux permettant d'éviter des résultats erronés ou mal interprétés, et qui est porteuse d'informations et facilite les avancées. De même, il est particulièrement intéressant de développer des bases de données dites "de vérité" permettant d'évaluer et de comparer objectivement des résultats.

L'apparition de la "science ouverte" ("open science" en anglais) propose des bonnes pratiques pour savoir comment gérer le triptyque publication / code / données en vue d'une réutilisation optimale et d'une mise à disposition auprès de la communauté.

### 4.2. En lien avec les autres disciplines

Ces avancées nécessiteront des liens encore plus étroits avec les chercheurs d'autres disciplines, mais aussi des perspectives interdisciplinaires larges afin d'intégrer aux modèles et aux algorithmes les caractéristiques propres des données manipulées. La question des données ouvre de nombreuses questions quant à l'effacement des frontières disciplinaires. Dans le domaine de la santé, par exemple, les implications sont considérables, depuis le traitement des données biométriques et les méthodes de travail des professionnels de santé jusqu'aux questions de sécurité, de confidentialité et de souveraineté dans la gestion de ces



données. Des questions similaires vont se poser dans les domaines de l'énergie, du patrimoine culturel, etc.

Les domaines d'utilisation des nouvelles technologies représentent également un aspect important du problème. Petit à petit, les chercheurs sortent de leur zone de confort en proposant des approches novatrices dans d'autres disciplines. Les utilisations et implications des technologies dans les domaines artistiques et environnementaux ne sont que deux exemples où des révolutions sont en train de s'opérer, dépassant largement nos frontières disciplinaires.

Les évolutions futures au niveau des technologies, des données associées et de leur exploitation influent également sur les comportements humains, de manière consciente ou inconsciente. La modélisation de ces derniers est donc un enjeu majeur, à l'interface entre les SHS et les sciences et techniques. Nous pouvons ainsi citer les problèmes d'influence de la tarification de la fourniture d'énergie sur les profils de consommation, du découpage électoral qui pourrait engendrer des biais géographiques et/ou socio-économiques, nuisant à terme à la démocratie.

La régulation pour éviter ces risques est une nécessité ; un enjeu majeur est de voir comment cette régulation peut être effectuée, en partant du constat que la régulation législative est souvent trop lente, donc inefficace, par rapport aux évolutions technologiques. Les liens entre évolutions technologiques et questions juridiques (sciences du droit) vont être amenés à se renforcer.

Nos disciplines entraînent également des changements dans le profil des chercheurs des autres disciplines. Par exemple, en linguistique, certains jeunes linguistes ne sont plus forcément des littéraires. De même, nous aurons besoin, non seulement de spécialistes dans nos disciplines, mais également de spécialistes à la fois dans nos disciplines et une ou plusieurs autres disciplines. Citons par exemple l'informatique quantique ou les machines moléculaires qui bénéficieront certainement de convergences multidisciplinaires. Ces domaines ont été abordés dans le [précédent rapport de prospective du CSI-INS2I](#) et les enjeux scientifiques associés restent encore de mise même s'il est difficile d'anticiper quand convergeront sciences et technologies dans ce domaine. Il est cependant nécessaire de maintenir un équilibre entre pluridisciplinarité et travail au cœur de la discipline.

## 4.3. Pour la société

### 4.3.1. Enjeux économiques

Il est toujours complexe de prévoir a priori l'impact économique des avancées scientifiques. Comme nous l'avons mentionné précédemment, nous trouvons également l'informatique à l'origine de nouvelles professions ou de la transformation de professions existantes. De manière générale, nous pouvons considérer que l'évolution des sociétés modernes a été essentiellement caractérisée ces 30 dernières années par le développement des nouvelles technologies du numérique. Dans le même temps, nous assistons à deux phénomènes qui nécessitent une attention particulière :



- La concentration du marché des nouveaux services « informatiques » aboutissant à un petit nombre de fournisseurs qui produisent de facto des asymétries, coûteuses pour la société ;
- La mise en compétition du monde de la recherche par les services de R&D de grands acteurs (mentionnés auparavant), de nouveau de manière asymétrique.

Si nous ajoutons le problème que l'introduction massive de dispositifs de décision automatique n'est pas toujours faite avec une connaissance précise de leur impact, nous pouvons conclure que serait nécessaire une nouvelle réglementation de l'intégration entre humain et machines et de leur convergence au service de la société.

#### 4.3.2. Enjeux humains

Comme nous l'avons déjà dit, l'essor des sciences de l'information a bouleversé de nombreux aspects de notre vie quotidienne. La façon dont la technologie influence la société les comportements humains et les limites légales, personnelles et sécuritaires est au centre des préoccupations.

L'acceptation des systèmes numériques par l'humain représente un enjeu majeur. En effet, dans de nombreux domaines, les humains devront de plus en plus souvent interagir avec des systèmes de plus en plus difficiles à appréhender et les machines, les robots en particulier, devront interagir avec des humains. Il ne fait aucun doute que cela nécessitera de nombreuses adaptations et évolutions. Comment redéfinir la place de l'humain lorsqu'il est confronté à un système, une IA, s'il n'en comprend pas les décisions ? Que deviendra la notion d'expertise lorsque les systèmes d'apprentissage seront plus efficaces que les experts du domaine ? Comment faire en sorte que la migration des tâches ne conduise pas à une perte de compétence de l'humain mais au contraire au développement de nouvelles compétences (ex. des pilotes de ligne pilotent de moins en moins mais suivent des formations permanentes pour acquérir de nouvelles compétences) ? Cette question prend une dimension supplémentaire lorsque nous parlons des dispositifs implantables. Sans aller jusqu'au transhumanisme, de nombreuses questions éthiques se posent lorsque ces dispositifs embarquent de plus en plus d'algorithmes.

Les sciences de l'information ont déjà, et sont amenées à avoir un impact encore plus fort dans nos sociétés et sur nos vies. Les grands enjeux scientifiques sont ainsi totalement imbriqués avec les enjeux sociétaux, et les principaux risques pour le développement de nos disciplines sont liés à l'acceptation sociale, au respect de la vie privée et à l'utilisation « positive » des recherches. En effet, nos disciplines, comme toutes les disciplines, produisent des avancées à double tranchant. A titre d'exemple, la bioinformatique permet de mieux lutter contre les agents pathogènes et d'aller vers une médecine de précision et personnalisée. A l'inverse, elle produit de nombreuses données personnelles dont l'usage n'est pas toujours bien réglementé. Cependant, la protection de la vie privée ne doit pas se traduire par un frein pour la recherche (données médicales par exemple), surtout quand d'autres pays appliquent des règles moins strictes. Les enjeux de recherche consistent donc à la fois à garantir la sécurité des données et à permettre leur exploitation dans des conditions acceptables. Un enjeu majeur pour l'avenir de notre discipline est donc lié à ses interactions avec les SHS.

### 4.3.3. Enjeux de Formation

Toutes ces mutations doivent être accompagnées d'une sensibilisation générale de la société aux enjeux des sciences de l'information. Il est nécessaire de rendre les sciences numériques accessibles, de les démystifier et de prévenir les fantasmes liés à leurs usages, afin également d'éviter la fracture numérique. Le développement des sciences de l'information, leur diffusion et leur bon usage dans la société ne pourra se faire qu'avec une évolution en profondeur de notre système de formation. L'enjeu est double. Il s'agit à la fois de former l'ensemble de la société *a minima* à la pensée informatique et aux interactions avec les systèmes informatisés, mais aussi de faire face aux besoins critiques des métiers numériques.

Il faut œuvrer pour un vrai enseignement de l'informatique à l'école. Cet apprentissage peut démarrer dès l'école primaire, sous forme ludique. Il doit s'intensifier au collège et au lycée. Cela passera par un effort massif sur la formation des éducateurs, en particulier à travers le développement d'un enseignement disciplinaire propre, associé à un CAPES informatique. Une maîtrise de base de la discipline doit faire partie du bagage de tout scientifique au baccalauréat. L'université et les écoles d'ingénieurs, quant à elles, ont largement fait évoluer leur offre de formation depuis 20 ans. Cet effort doit être poursuivi, avec une meilleure intégration des sciences du numérique dans tous les cursus. Nous pourrions par exemple penser au développement de modules pluri et interdisciplinaires avec d'autres domaines scientifiques, comme la biologie, les mathématiques appliquées, la physique, la linguistique, les SHS. De plus, il faut favoriser la formation tout au long de la vie. Les modes d'apprentissage se sont élargis depuis plusieurs années, aidés en cela par les sciences de l'information, via des solutions d'accès simple et à coût faible, comme avec les cours en ligne ou les MOOC (Massive Open Online Course). Ces solutions d'apprentissage peuvent avoir un large impact sur la société de par leur diffusion de masse. L'apport des sciences de l'information dans ce domaine pourrait ainsi permettre de les développer (pour mieux s'adapter à l'apprenant par exemple).



## 5. Conclusion

Les sciences de l'information et leurs interactions ont pris une place nouvelle et centrale dans le paysage scientifique global et nous amènent à relever de nombreux défis technologiques, sociétaux et scientifiques. Un enjeu central sera de démontrer notre capacité à maîtriser notre destin dans le domaine du numérique, afin de ne pas être "colonisés" par des acteurs privés et mondialisés. L'Europe, et en particulier la France, doivent être en mesure de préserver une recherche et une éducation d'excellence dans le domaine du numérique pour garder leur indépendance et survivre. C'est bien sûr vrai pour des enjeux scientifiques, mais c'est également vrai pour des enjeux économiques, de souveraineté et de sécurité. Faire monter et/ou maintenir en compétence le tissu industriel national et européen, tout comme garantir notre sécurité en matière de numérique passe par le maintien d'une recherche française d'excellence dans le domaine des sciences de l'information. Un tel maintien en France dans notre domaine se fera dans un contexte d'extrême concurrence pour recruter les futurs talents et les retenir. En comparaison avec les GAFAM, notre avantage compétitif ne se fera pas sur les rémunérations. Même si les écarts devront être réduits, notre facteur de différenciation doit se focaliser d'une part sur le maintien d'un statut garantissant au chercheur un recrutement jeune et une liberté scientifique, dans un cadre éthique garanti, et d'autre part sur la mise à disposition de moyens publics significatifs afin d'alléger les contraintes actuellement très lourdes de recherche et de gestion de crédits. La demande extrême de "*multitâchisation*" des chercheurs représente un frein extrêmement important dans le maintien de l'excellence de notre recherche, et de son attractivité dans le monde sans-frontière de la recherche dans notre domaine.

Enfin, et pour conclure, à travers l'élaboration de ce rapport, les différents échanges qui ont eu lieu ont fait émerger un questionnement nouveau des participants sur la manière dont notre discipline évolue et autour de l'impact croissant de l'informatique dans nos vies quotidiennes et nos métiers. Cette prise de conscience "silencieuse" est en soit, elle-même, porteuse de changements à venir dans la manière de définir et d'aborder les enjeux scientifiques dans les sciences de l'information et leurs interactions. Ces changements restent difficiles à prédire, les directions qu'ils vont prendre produiront sans doute des ruptures dans la manière dont les chercheurs vont faire évoluer leurs thématiques scientifiques, changeant également nos manières de fonctionner et d'interagir avec les autres disciplines.



## Remerciements

Le conseil scientifique tient particulièrement à remercier les personnes suivantes pour leurs contributions au travail du conseil et à son organisation :

- Danuta Dufrat-Chabrière, *Assistante du Comité National*
- Invités permanents aux séances du Conseil Scientifique : Frédérique BASSINO et Hubert COMON-LUNDH, *Président-e-s de la section 6*; Michèle BASSEVILLE et Pierre-Olivier AMBLARD, *Président-e-s de la section 7*; Dimitri PEAUCELLE et Valerie BERTHE, *Représentants du Conseil scientifique du CNRS*; Eduardo ROCHA, Michael BLUM, Guillaume FERTIN, Dominique LAVENIER, *Représentants de la CID 51*; Jean-Pierre NADAL, *Représentants de la CID53*
- Invités des séminaires thématiques au cours du mandat : Jean-François GIBRAT, Eric FLEURY, Michel DE MATHELIN, Jean-Marc VÉZIEN, Vivien QUEMA, Brigitte PLATEAU, Isabelle COLLET, Anne PÉPIN, Gérard BERRY, Jean-Marc JÉZÉQUEL, Jean-Yves MARION, David MONNIAUX, Max DAUCHET, Raja CHATILA, Karën FORT, Herve CHNEIWEISS, Magali FITZGIBBON, Xavier LEROY, Marie-Anne SAINT-JALMES, Aurélien FRANCILLON, Andrzej DUDA, Merouane DEBBAH, Mariana NETTO, Frédéric VANDERHAEGEN, Franck MARS, Antoine CHAILLET, Remi RONFARD, Gérard ASSAYAG, Philippe GAUSSIER, Serge ABITEBOUL, Jérôme LANG, Benjamin NGUYEN, François PELLEGRINI, Franck LETHIMONNIER, Catherine NGUYEN, Carine GIOVANNANGELI, Etienne HIRSCH, Pierre-Olivier COURAUD

Par ailleurs, nous tenons à remercier vivement les personnes ayant participé aux ateliers de discussion préparatoires à l'élaboration de ce rapport de prospective :

- Gildas AVOINE, Meghyn BIENVENU, Laure BLANC-FERAUD, Patrick FLANDRIN, Olivier GASCUEL, Antoine GIRARD, Christine GUILLEMOT, Jean-Marc JEZEQUEL, Jérôme LANG, Claire MATHIEU, Laurence NIGAY, Joseph SIFAKIS, Sophie TARBOURIECH, Jocelyne TROCCAZ



# Annexes au rapport de prospective du Conseil Scientifique INS2I du CNRS

1. Synthèse sur *“la responsabilité sociale des algorithmes”*
2. Synthèse sur *“la Production et la valorisation des logiciels”*
  - Fiche d’autoévaluation « *Logiciels* »
3. Synthèse sur le sujet *“Sciences de l’information et Arts”*
4. Synthèse du CSI-INS2I sur *“Les Systèmes d’Information en Interaction avec, ou centrés sur, l’Humain (S2IH)”*
5. Synthèse sur *“l’Internet des Objets”*
6. Synthèse et propositions sur *“l’Éthique en sciences de l’information”*
7. Synthèse sur *“La Sécurité dans les sciences de l’information”*
8. Recommandations sur l’égalité Homme-Femmes en Sciences de l’Information
9. Synthèse du CSI-INS2I sur les *“Plateformes d’envergure en sciences de l’information”*
  - Proposition de « *Questionnaire Plateforme* »

# Synthèse et propositions sur la responsabilité sociale des algorithmes

Conseil Scientifique de l'institut CNRS INS2I

Fiche de synthèse, Septembre 2018

## 1 Motivations

De procédures formelles pour résoudre de problèmes complexes de “décision” existent depuis de siècles. Plusieurs de ces procédures sont utilisées par ailleurs aujourd’hui dans des “machines” qui gèrent de manière automatique ou semi-automatique de processus complexes : gestion de la production, gestion de réseaux de communication et/ou de distribution, gestion du trafic, allocation de fréquences en télécommunication mobile, conduite automatique (trains, métro etc.). L’utilisation de telles procédures (soit sous forme de logiciels ou simplement comme procédures d’allocation de ressources) n’avait soulevé que peu de débats de nature éthique ou encore sur leur impact potentiellement négatif sur les personnes et la société.<sup>1</sup>

Nous assistons aujourd’hui à une diffusion d’études, de discussions et d’articles de presse centrés sur l’impact social et économique de la diffusion “massive” de l’utilisation de telles procédures et des algorithmes associés. Si d’un côté nous assistons à une vision non scientifique de cette thématique nous ne pouvons pas non plus ignorer son existence. Pour cette raison le Conseil Scientifique de l’INS2I a promu l’organisation de quatre interventions le 25/09/2017 :

- Serge Abiteboul (ENS, INRIA) “Responsabilité sociale des algorithmes”,
- Jérôme Lang (LAMSADE, Dauphine) “Conception de mécanismes et responsabilité sociale des algorithmes”,

---

1. Notons cependant le débat en France dans les années 70 autour de la loi informatique et liberté de 1978 qui anticipait certaines conséquences négatives de l’utilisation généralisée des bases de données contenant des données personnelles.

- Benjamin Nguyen (LIFO, Orleans) “Responsabilité sociale des algorithmes : quelques enjeux”,
- François Pellegrini (LABRI, Bordeaux et membre de la CNIL) “Enjeux sociétaux des traitements informatiques”.

Le document que suit est essentiellement fondé sur les présentations de ces collègues et la discussion qui s’ensuivait.

## 2 Historique et positionnement

Le concept d’algorithme est ancien. Toute procédure qui peut être décrite dans un langage formel et qui permet de résoudre un “problème” à travers l’allocation de ressources à des tâches peut être définie comme un algorithme et l’histoire nous présente des exemples d’algorithmes depuis de siècles. Le caractère formel de la description de ce qu’un algorithme fait a permis d’imaginer (et de réaliser) de machines qui exécutent les algorithmes : de manière simplifiée nous sommes en train de décrire la naissance de la discipline Informatique comme le champ scientifique qui s’occupe de cette typologie de procédures. Ce qu’il faut retenir est :

1. l’existence des algorithmes avant la naissance de l’Informatique et
2. l’existence des logiciels qui permettent d’utiliser les algorithmes de manière massive et efficace.

Le concept de responsabilité a également été étudié depuis bien avant la naissance de l’Informatique. Toute action rend l’agent qui l’entreprend responsable des conséquences qu’elle entraîne. Il doit donc analyser les conséquences potentielles d’une action avant son exécution, de manière à prévenir, maîtriser, réglementer ou interdire. Toute action intentionnelle nécessite une décision : nous trouvons ainsi le lien entre la responsabilité d’une action et le processus de décision qui l’a causé. Aujourd’hui un tel processus peut être géré (ou est géré) par un algorithme.

Cependant le concept de responsabilité reste principalement associé aux personnes et groupes de personnes ainsi qu’à des organisations. Du point de vue juridique (et de manière intuitive) une responsabilité ne peut être associée qu’à une personne physique ou morale et non pas à une machine, ou encore à une procédure ou à des règles. Prenons l’exemple des systèmes de conduite automatique comme le métro VAL : la responsabilité en cas d’accident revient à la société de construction et/ou la société d’exploitation et non pas au VAL lui-même. De ce point de vue le terme “Responsabilité sociale des algorithmes” est un paradoxe :

les algorithmes ne peuvent pas être responsables des conséquences de leur utilisation.

Le fait est que malgré ce principe élémentaire nous assistons à une diffusion de préoccupations autour de l'utilisation des algorithmes dans la prise de décisions impactant notre vie quotidienne et dans des systèmes qui gèrent l'organisation de nos sociétés et de l'économie. Pourquoi ? Dans la suite nous énumérons quelques causes possibles :

1. L'ampleur de la diffusion de procédures automatiques (ou semi-automatiques) de décision.
2. Le recours à des procédures qui apprennent de manière autonome et sans qu'on puisse contrôler la manière par laquelle l'apprentissage intervient, entraînant un comportement non prévisible de l'algorithme.
3. La concentration de la production de logiciels entre les très grandes entreprises d'édition de logiciels avec une industrialisation de l'implémentation informatique d'algorithmes (en dehors des *open source*) qui les rend encore plus opaques.
4. L'utilisation de données "privées", "sensibles" ou assimilables à des communs (*commons*) sans un retour bien identifié aux fournisseurs de données et sans toujours une autorisation claire, ainsi que l'absence de connaissances sur l'étendue des utilisations.
5. Le constat que dans plusieurs cas les procédures automatiques de décision ont produit de conséquences biaisées, inattendues, parfois aberrantes avec des impacts à long terme potentiellement non discutés dans la société.
6. L'absence d'une limite claire sur le transfert d'autonomie décisionnelle jugé acceptable.
7. Le monopole *de facto* des grands collectionneurs de données (les GAFA), engendrant un sentiment d'absence d'alternatives et d'impuissance.

### 3 Thématiques

Essayons donc de structurer le problème et de comprendre les points qui peuvent poser des questions à la fois dans la communauté scientifique et dans la société en général. Nous le ferons à travers des interrogations selon cinq dimensions : l'autonomie décisionnelle, la confiance aux décisions automatiques, la vérification des systèmes de décision automatique, la régulation, l'asymétrie de la connaissance.

1. Le premier problème (probablement le plus fondamental) est l'autonomie décisionnelle grandissante qu'on transfère à des "machines", à des artefacts autonomes (AA) et la manière dont ce transfert est opéré. Rappelons cependant que ce transfert existe depuis long temps. Il faut comprendre ce qui le rend aujourd'hui un problème ou une menace potentielle.

1. Qui décide si un certain type de processus de décision peut être transféré à un Artéfact Autonome et rendu automatique ? À qui nous pouvons confier ce pouvoir de décision (dans le cas de décision publiques comme sur les modalités de l'Admission Post-Bac) ?
2. Supposons que il y a un accord sur l'automatisation d'un processus de décision (ou qu'une entreprise décide d'automatiser un processus de décision faisant partie de ses activités) ; qui décide la manière par laquelle cette automatisation sera rendue effective ?
3. L'interrogation précédente nous amène donc à nous interroger sur les propriétés que la solution et la procédure automatique doivent garantir. Qui décide les propriétés à respecter et comment concevoir des Artéfacts Autonomes qui les respectent ? Est-ce qu'on peut utiliser des outils de vérification de programmes pour démontrer les propriétés de la procédure ?
4. D'où viennent les données qui sont nécessaires pour faire marcher correctement la procédure de décision automatique ? Sont-ce des données privées ? Publiques ? Comment établir la propriété des données et les droits d'exploitation ? Comment garantir que la procédure n'introduit pas de biais ou discriminations ?
5. Qui est responsable pour une décision automatique prise par l'Artéfact Autonome ? Qui peut être tenu responsable en cas de litige devant la justice ?

La dernière demande soulève un problème d'ordre général : jusqu'à quel point sommes nous prêts à transférer de l'autonomie décisionnelle à des Artéfacts Autonomes ? Y a-t-il des processus de décision que nous ne souhaitons pas automatiser (même si c'est possible) ?

2. À quelles conditions sommes nous prêts à accepter une décision prise de manière automatique par un Artéfact Autonome ? Ceci soulève le problème de la confiance dans les algorithmes qui sont utilisés pour permettre à l'Artéfact Autonome de décider de manière autonome.

1. Étant donné un ensemble d'algorithmes utilisés par un Artéfact Autonome, sommes nous capable de construire une trace complète de l'activité des algorithmes ?

2. Si nous sommes capables de construire une telle trace, sommes nous capables de construire des explications (interprétables, simples et intuitives, utilisables) pour justifier les décisions prises à n'importe quelle partie prenante ?
3. Sommes nous capables de produire les "raisons ultimes" d'une décision qui a été prise de manière automatique ? Si c'est le cas pouvons-nous reproduire cette décision à partir des mêmes données en entrée ?
4. Si nous utilisons des algorithmes qui apprennent à chaque exécution (et se modifient en conséquence) alors nous ne pouvons pas garantir la reproduction du résultat : nous avons un comportement non déterministe de l'algorithme. Quelles explications/justifications/raisons nous accepterions dans ce cas ?
5. L'utilisation d'un Artéfact Autonome qui automatise certaines décisions peut avoir de conséquences à long terme que nous ne pouvons pas prévoir aujourd'hui (des 'inconnus aujourd'hui inconnus'). À quelle horizon temporel faut-il vérifier les impacts à long terme induites par l'utilisation de tels Artéfact Autonome ?

De manière générale les interrogations soulevées tournent toutes autour du problème de l'opposition possible à une décision prise de manière automatique, ou de sa défense dans le cas où une contrepartie considère que une telle décision soit opposable. Comment pouvons nous structurer une discussion argumentée autour de décisions prise de cette manière ? Au bout du compte une décision est "acceptable" si nous la retenons non opposable.

**3.** Clairement, une dimension de préoccupation dans la construction des Artéfacts Autonomes est le fait qu'il s'agit de logiciels complexes pour lesquels se pose le problème de la vérification formelle de leur comportement.

1. Supposons qu'on ait conçu un algorithme garantissant une certaine propriété (spécification) du résultat (voir le point 1 de la discussion ; par exemple un algorithme 'équitable'). Comment pouvons-nous garantir de manière formelle que le logiciel qui implémente un tel algorithme continue à respecter les bonnes propriétés de l'algorithme ? En d'autres termes : l'ingénierie des spécifications peut-elle garantir aujourd'hui le respect de cette typologie de demande ? Est-ce qu'on peut vérifier des Artéfacts Autonomes de la même manière qu'on vérifie aujourd'hui la spécification d'un programme contre son implémentation ?

2. Est ce que des logiciels qui garantissent de tels spécifications peuvent garantir également l'efficacité de l'algorithme ?
3. Est ce que nous pouvons garantir au même temps le respect des données privées et leur protection d'une part et de l'autre part la possibilité d'expliquer, tracer les décisions d'un Artéfact Autonome ? Où devons-nous placer le compromis nécessaire entre ces deux exigences ?
4. Si le recours à des logiciels libres peut garantir un contrôle plus efficace sur les caractéristiques et le comportement d'un Artéfact Autonome, comment intégrer l'utilisation des composantes cryptographiques et de sécurité en général qui nécessitent une certaine confidentialité du code ? Encore une fois comment trouver un compromis acceptable et quelles seront les conséquences sur la limite de conception d'un Artéfact Autonome ?

Plus généralement, le problème de la vulnérabilité et donc de la sécurité des Artéfact Autonome avec autonomie décisionnelle augmentée sera un sujet d'importance majeure dans l'immédiat (voir la discussion sur la sécurité du vote électronique).

**4.** Jusqu'à maintenant nous avons soulevé des questions qui concernent la conception d'algorithmes et de systèmes de décision automatique que nous avons appelé Artéfact Autonome. Vue l'impact potentiel de tels systèmes, une question qui se pose de manière naturelle est celle de la régulation de la conception, de la mise en œuvre et de l'utilisation de tels Artéfacts Autonomes.

1. Si nous ne pouvons pas prévoir l'impact d'un tel système (par la vérification de leur spécification), faut-il prévoir une phase de test (plus ou moins comme dans le cas de nouvelles molécules chimiques ou de nouveaux aliments) avant d'autoriser l'utilisation d'un Artéfact Autonome avec autonomie décisionnelle augmentée ?
2. Faut-il donc envisager la création d'une autorité indépendante de certification et d'autorisation ?
3. Étant donné l'impact potentiel sur la société dans certains contextes de l'utilisation de tels Artéfacts Autonomes, faut-il prévoir une déclaration explicite de la politique et de la raison pour laquelle l'utilisation d'un Artéfact Autonome est prévu ?
4. Étant donné l'importance économique de l'industrie informatique et de l'édition de logiciels, une telle régulation est-elle réaliste ? À quelles conditions ?

5. Étant donné le cycle de vie (souvent assez court) de tels Artéfacts Autonomes, est-il réaliste d'imaginer une régulation (qui a typiquement un cycle de vie bien plus long) ?

**5.** D'une manière générale, on peut constater une asymétrie massive de l'information : les GAFAs savent tout sur nous tandis que nous ne savons que peu sur eux, à commencer par les algorithmes qu'ils utilisent. L'exemple le plus frappant en est le *PageRank* de Google dont le fonctionnement est tenu secret. En économie il est bien connu que l'asymétrie informationnelle peut entraîner des défaillances du marché. L'importance de cette asymétrie et le contexte technologique posent ici des questions nouvelles.

1. Comment pouvons-nous juger l'explication d'une décision d'un algorithme qui nous est fournie si le fonctionnement de l'algorithme n'est pas connu ? Pouvons nous faire confiance à un tel service ?
2. Les conditions d'utilisation de multiples services requièrent actuellement l'accord des utilisateurs dans le cadre du nouveau Règlement général sur la protection des données (RGPD), et presque tous les utilisateurs acceptent en cliquant sur "J'ai lu" sans en avoir pris connaissance : comment éviter une telle situation ? Comment s'assurer d'une utilisation responsable de services de ce type ?
3. Si l'utilisation d'un Artéfact Autonome engendre de-facto la stipulation d'un contrat, comment s'assurer que les parties prenantes : "concepteur d'un Artéfact Autonome", "éditeur du logiciel", "utilisateur de l'Artéfact Autonome", "la société en général", soient conscients du contrat signé ?
4. Qui est l'autorité compétente en cas de litige dans le contexte de contrats de ce type ?

Nous laissons par ailleurs ouverts les dimensions éthiques de la manipulation de données et de la "morale" associable à de procédures de décision automatique : ces sujets demandent également une discussion et peut-être une régulation, cf. la discussion actuelle sur les 'machines morales' (*moral machines*). Nous partons effectivement de l'hypothèse que les raisons économiques qui poussent l'évolution et la création d'Artéfact Autonome dotés d'autonomie décisionnelle resteront impératives.

## 4 Ressources

La communauté Française a déjà plusieurs activités dans ce domaine. Dans la suite nous listons les initiatives dont nous avons connaissance.

- La CNIL a publié un rapport en novembre 2017 “Comment permettre à l’homme de garder la main ?”<sup>2</sup> suite à une série de débats publics dans le cadre de la mission éthique qui lui a été confiée par la loi pour une république numérique de 2016.
- Le GDR “Sécurité Informatique”<sup>3</sup> a un groupe de travail sur la protection de la vie privée dirigé par Benjamin Nguyen (LIFO, Université d’Orleans). Notons que ce sujet dépasse le cadre de ce GDR avec une dimension clairement interdisciplinaire.
- Le GDR “Intelligence Artificielle” a organisé en 2017 une Journée Explicabilité des systèmes d’IA<sup>4</sup> animée par Nicolas Maudet (LIP6, Université Pierre et Marie Curie); il est prévu de lancer un groupe de travail sur le sujet.
- Le GDR “Policy Analytics”<sup>5</sup> a organisé en 2017 un colloque sur la “Responsabilité Sociale des Algorithmes”<sup>6</sup>.
- À l’Université Paris Est Créteil existe un groupe de travail interdisciplinaire “ Algorithme et citoyenneté” (ALGOCIT)<sup>7</sup> animé par Pierre Valarcher.
- À l’ISCC (CNRS et Université Pierre et Marie Curie) Melanie Dulong de Rosnay et Francesca Musiani travaillent sur ces thématiques.<sup>8</sup>
- Deux ouvrages de Serge Abiteboul (“Le temps des Algorithmes”, avec Gilles Dowek et “Terra Data”, avec Valérie Peugeot) et l’ouvrage de Francesca Musiani (“Nains sans géants”).

Au niveau international signalons :

- Un projet Européen dirigé par Dino Pedreschi (Université de Pise) sur la collecte responsable de données.<sup>9</sup>

---

2. [https://www.cnil.fr/sites/default/files/atoms/files/cnil\\_rapport\\_garder\\_la\\_main\\_web.pdf](https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf)

3. <https://gdr-securite.irisa.fr>

4. [https://www.gdria.fr/journee\\_explicabilite/](https://www.gdria.fr/journee_explicabilite/)

5. <http://www.gdr3720.fr>

6. <http://www.lamsade.dauphine.fr/sra2017>

7. <http://www.algo-cit.fr/>

8. <http://www.iscc.cnrs.fr>

9. <https://sobigdata.eu>

- Un séminaire Dagstuhl sur la conception d’agents moraux.<sup>10</sup>
- La création de la communauté FAT (Fair, Accountable, Transparent Algorithms) essentiellement aux États-Unis.<sup>11</sup>
- La création de l’Institut 3A (Autonomy, Agency, Assurance) à l’Australian National University, dirigé par Geneviève Bell (ancienne vice-président de INTEL).<sup>12</sup>

## 5 Recommandations

Nous ne pouvons pas ignorer que les “algorithmes”, leur conception, leur analyse et leur utilisation représente un élément essentiel de la discipline Informatique et que leur étude du point de vue de leur impact dans la société et dans notre vie quotidienne doit être une préoccupation majeure pour tous les membres de notre communauté scientifique. La discussion de la section précédente montre que, bien que l’existence d’Artéfact Autonome dotés d’autonomie décisionnelle soit une réalité depuis très longtemps il est temps de passer à une phase de réflexion plus approfondie. De ce point de vue nous voyons deux types de recommandations : l’une en ce qui concerne les aspects scientifiques et l’autre les aspects sociétaux.

Nos recommandations s’adressent essentiellement au CNRS : l’INS2I de manière spécifique (vue que nous adressons avant tout la communauté Informatique), la Mission pour l’interdisciplinarité (vue la dimension fortement interdisciplinaire du sujet) et la direction du CNRS pour l’impact du sujet sur la société.

En ce qui concerne les *aspects scientifiques* il est probablement temps de susciter des actions de recherche orientées à construire des éléments de réponse aux demandes posées. Actions qui peuvent être prises :

- lancer un appel à PEPS (via le INS2I ou en coopération avec la MI) qui peuvent adresser une ou plusieurs de demandes posées dans ce document ;
- faire de ce sujet une année thématique du CNRS (INS2I et/ou MI) avec l’implication des GDR “Sécurité Informatique”, “IA”, “RO”, “MADICS”, “Policy Analytics” dans l’organisation de journées de discussion ; dans ce cadre (ou de manière indépendante) organiser un colloque national des chercheurs intéressés par la thématique ;

---

10. <http://www.dagstuhl.de/en/program/calendar/semhp/?semnr=16222>

11. <http://www.fatconference.org>

12. <https://3ainstitute.cecs.anu.edu.au/>

- proposer (direction du CNRS) le sujet à l'ANR et ou le Ministère (PIA AI) comme programme de recherche interdisciplinaire (et motiver la communauté scientifique à participer) ;
- proposer (direction du CNRS) également le sujet à l'attention de l'UE en ce qui concerne les priorités du 9ème PCRD ;
- construire (direction du CNRS/DRI) des liens avec des actions similaires au niveau international, comme l'Institut australien 3A et la communauté FAT déjà mentionnés plus haut, le DIMACS<sup>13</sup> et les financements NSF aux États-Unis, ou le projet Algorithm Watch<sup>14</sup> en Allemagne ; dans ce cadre proposer la création d'un réseau Européen (Action COST ?) sur la thématique.
- discuter (dans le cadre de l'alliance ALLISTENE) avec l'INRIA autour de la plateforme de test et d'audit des algorithmes pour créer une initiative nationale.

En ce qui concerne les *aspects sociétaux* nous proposons des actions dans deux sens : un vers le CNRS et sa communauté scientifique et l'autre vers la société en général. L'objectif de ces actions est d'attirer l'attention de la société en général ainsi que de composants spécifiques à l'importance du problème et de la nécessité d'investir de ressources et de l'intelligence pour s'occuper. Des actions que le INS2I, la MI et la direction du CNRS peuvent entreprendre incluent :

- un positionnement du CNRS sur l'ensemble des problèmes cités à travers un document succinct qui précise l'investissement du CNRS ;
- la publication d'un insert spécial dans le Journal du CNRS ;
- prendre l'initiative pour solliciter une discussion à l'Assemblée nationale sur la base d'un document CNRS/INRIA/Ministère ;
- se concerter avec d'autres organismes de recherche en Europe pour solliciter le Parlement Européen (créer un document commun ?) ;
- proposer (via la Direction de la Valorisation) la création d'une table ronde des entreprises qui produisent ou ont un recours massive à des Artéfacts Autonomes avec autonomie décisionnelle augmentée, en vue d'une réflexion sur l'auto-régulation ;
- proposer (via la Direction de la Valorisation) un brain-trust avec des organismes professionnels sur l'avenir de certaines professions et de certaines activités économiques qui sont les premières à être touchées par cette évolution.

---

13. <http://dimacs.rutgers.edu>

14. <https://algorithmwatch.org/en>

# Synthèse journée thématique Production et Valorisation des Logiciels

Conseil scientifique INS2I

19 Sept. 2016

## Résumé

Rédigée par le Groupe Thématique "Production et valorisation des logiciels" composé de : Anne-Catherine Letournel (animatrice), Christian Barillot, Isabelle Queinnec, Guy Gogniat, François Chaumette

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Les intervenants de la journée</b>	<b>2</b>
<b>3</b>	<b>Les trois formes d'un logiciel de laboratoire</b>	<b>3</b>
<b>4</b>	<b>La propriété intellectuelle et les attendus des tutelles</b>	<b>3</b>
<b>5</b>	<b>L'évaluation de la recherche</b>	<b>4</b>
<b>6</b>	<b>Conclusions</b>	<b>5</b>
	<b>Fiches pratiques</b>	<b>6</b>
<b>A</b>	<b>Comment se former aux outils d'hébergement de code ?</b>	<b>6</b>
<b>B</b>	<b>Fiche d'auto-évaluation du logiciel</b>	<b>8</b>
<b>C</b>	<b>Caractérisation simplifiée d'un logiciel</b>	<b>9</b>

# 1 Introduction

Lors du conseil du 9 mai 2016, la direction de l'INS2I a fait part d'une demande au CSI au sujet du transfert technologique, de la valorisation et de l'évaluation de la production logicielle. Il s'agit de faire un état des lieux de ce qui est fait dans les laboratoires de l'INS2I et ce qui serait à faire autour de la valorisation et du logiciel. En effet, la DIRE s'occupe de la valorisation plutôt focalisée sur les sciences du vivant et l'ingénierie, mais possède une expérience faible en valorisation logicielle. L'institut souhaite utiliser au mieux les moyens fournis par les écosystèmes locaux (SATT par exemple) ou par tout autre moyen extérieur aux unités, mais pouvant être complémentaire. Il demande au CSI des préconisations dans ce sens. Egalement, l'INS2I pense que les dépôts APP sont un bon indicateur de la production logicielle de ses laboratoires et souhaite mettre en place une politique incitative en la matière. Le faible nombre de dépôts APP ne reflète pas la richesse de la production logicielle et ne permet pas non plus de l'appréhender.

Il est intéressant de rappeler certains éléments de réflexion du mandat du précédent CSI sur la question de la recherche partenariale. En 2013, le CSI s'est penché sur les conditions de caractérisation de la recherche partenariale, comme nécessité économique, comme allers-retours constructifs entre théorie et applications, comme indicateur de l'adéquation du diplôme de docteur aux besoins de recrutement des entreprises. Un constat avait été fait sur les cultures différentes dans le monde industriel où l'on parle de 'produits' et dans le monde académique où l'on parle de 'problèmes'. Cela avait amené à proposer une lecture en deux axes pour la création d'une oeuvre logicielle par un laboratoire : le market-pull, en réponse à une demande exprimée par le milieu industriel ou le techno-push, pour valoriser des résultats dont les entreprises ignorent encore l'existence. Le market-pull faisant plutôt appel à des structures externes aux laboratoires mais associées au monde académique (comme les SATT) et le techno-push se basant lui, sur des moyens et méthodologies à mettre en oeuvre au sein des labos, sans alourdir les processus de production des logiciels.

Entre le logiciel réponse à une problématique scientifique et le logiciel produit sur étagère se situe la maturation. Les activités de maturation doivent donc être encouragées et soutenues dans les laboratoires pour inciter aux bons réflexes de qualité logicielle dès le démarrage des projets et pour aider les équipes à mettre en valeur leurs productions logicielles lorsque cela a un sens. Dans tous les cas, une forte motivation du chercheur est souvent indispensable à la réussite d'un transfert technologique, ce qui sous-entend pour le chercheur, d'une part la prise en compte d'un cadre législatif assez spécifique et souvent méconnu et d'autre part un investissement temporel important.

En conclusion, pour aller plus loin, la réflexion suggérerait de considérer comment les laboratoires traitaient ces questions aujourd'hui et avec quels moyens de soutien. Les initiatives des sections 6 et 7 affichant clairement la prise en compte de la production logicielle dans l'évaluation des dossiers des chercheurs sont un bon début, mais encore très insuffisant pour un chercheur, en comparaison du travail de publication scientifique.

## 2 Les intervenants de la journée

Afin de connaître comment sont traitées les questions de valorisation du logiciel dans les laboratoires de l'INS2I, le groupe de travail a organisé une enquête en une vingtaine de questions, envoyée à toutes les unités CNRS rattachées à l'INS2I, pour obtenir un premier aperçu de l'éventail des solutions mises en oeuvre. Les résultats ont été présentés en ouverture de la journée thématique [1].

Les liens avec les autres partenaires publics et acteurs régionaux de l'innovation sont à étudier pour ne pas être dans les laboratoires en opposition avec les partenaires, mais plutôt en complément. L'institut ne cherche pas à remplacer ce que font les SATT, la DIRE, BPI Banque Publique d'Investissement et OSEO, les pôles de compétitivité, les incubateurs etc ... Sur ce sujet, Marie-Anne St Jalmes, de la SATT Ouest Valorisation, est invitée pour présenter le rôle de la SATT dans le cycle complet de production et valorisation des logiciels [2].

L'évaluation de la production logicielle est au coeur de la question car elle semble mésestimée dans le système actuel. Si les chercheurs ne voient pas le résultat de leurs investissements dans le logiciel pris en compte, ils ne sont pas encouragés à y consacrer de l'énergie. Xavier Leroy, cher-

cheur Inria, membre de la commission d'évaluation Inria, est invité pour expliquer comment cette commission évalue la production logicielle lors des recrutements et promotion de ses chercheurs. Il présente également trois représentations possibles d'un logiciel : dispositif expérimental, autre forme de publication, objet de valorisation [3].

Enfin, pour le monde extérieur, le nombre de dépôts APP est un élément visible pour du patrimoine immatériel du CNRS. Malgré une réalité évidente de l'importance du transfert technologique dans la discipline des sciences de l'informations et de leurs interactions - le quart des startups du CNRS sont issues d'unités de l'INS2I d'après une enquête CNRS citée par W.Perruqueti- le dépôt APP est très peu pratiqué. L'institut souhaite mettre en place une politique incitative et demande l'avis du CSI. Magali Fitzgibbon est également invitée lors de cette journée thématique et son exposé [4] abordera cet outil à travers la notion plus générale d'actif du patrimoine immatériel que constitue le logiciel.

### 3 Les trois formes d'un logiciel de laboratoire

On propose la caractérisation simplifiée suivante pour tout logiciel de laboratoire :

- Lorsque le logiciel conserve une visibilité interne aux auteurs, c'est un prototype de test d'idées nouvelles (niveau 1, pas de choix de licence).
- Lorsque le logiciel est transféré gratuitement vers la communauté, la société ou l'enseignement, c'est une autre forme de publications que les publications scientifiques (niveau 2, plutôt une licence libre).
- Lorsqu'on parle de transfert technologique vers une application industrielle, c'est ce qu'on appelle communément "valorisation industrielle" (niveau 3, plutôt une licence propriétaire).

Cette classification n'est pas une échelle de maturation, le logiciel pouvant passer du niveau 1 au niveau 3 sans passer par le niveau 2, ou même exister partiellement avec une licence libre et partiellement avec une licence propriétaire.

Les solutions et les moyens de soutien à la production logicielle à privilégier sont très dépendants des trois catégories ci-dessus. Par exemple : accompagner et guider les ressources propres en ingénierie logicielle au niveau 1, fournir des ressources dédiées et pérennes (ingénieurs développement) pour constituer un soutien significatif de l'effort de développement au niveau 2, utiliser les dispositifs de maturation extérieurs aux laboratoires pour obtenir des ressources CDD au niveau 3.

Les plateformes matérielles portant du logiciel (type Equipex) ont déjà été évoquées lors d'une journée thématique dédiée en 2015 [5]. Elles sont un cas particulier de logiciel et sont concernées à ce titre par les préconisations évoquées dans la présente étude.

Les plateformes open-source collaboratives à visibilité mondiale (type logiciel de calcul SAGE ou COQ) sont un cas spécifique car le cadre juridique dépasse celui du CNRS. Les spécificités de valorisation des contributions à ce type de plateformes n'ont pas été évoquées lors de la journée.

**L'identification des trois catégories décrites ci-dessus pourrait constituer une première étape dans un processus de cartographie du patrimoine logiciel des unités de l'INS2I.**

### 4 La propriété intellectuelle et les attendus des tutelles

Les points d'intérêt suivants sont intrinsèques à l'objet 'logiciel' et sont incontournables pour mieux maîtriser un patrimoine logiciel :

- Le logiciel est un objet collaboratif par définition à cause des avantages et de la facilité de ré-utilisation du code.
- L'identification des parts respectives de contribution des différents auteurs ne fait pas partie implicitement du cycle de vie du logiciel.
- Dès qu'il est diffusé auprès d'un tiers, le logiciel a juridiquement besoin d'une licence. Une nouvelle fois, l'attribution d'une licence ne fait pas partie implicitement du cycle de vie du logiciel.
- Le brevet est rarement utilisé pour un logiciel, mais s'il l'est, il restreint la diffusion des technologies et ralentit le développement du logiciel.

La méconnaissance par les établissements employeurs des auteurs de leurs parts respectives est un frein à la maturation. Les solutions pour l'identification des auteurs passent par de l'outillage

adapté (forges de type Gitlab par exemple) dès le démarrage d'un projet ou de l'outillage adapté en post-traitement, pour identifier les composants ré-utilisés. **L'utilisation de tels outils doit être encouragée au moment de la production des logiciels.**

Pour aider les développeurs qui manquent de repères à s'initier aux bons outils, nous proposons en annexe A, une représentation simplifiée de l'écosystème des outils et forges d'hébergement de code. Des pistes sont indiquées pour aider à sélectionner les bons outils et démarrer avec l'utilisation d'un gestionnaire de versions logicielles.

Les questions de licences sont à traiter au cas par cas avec les auteurs, leurs employeurs et surtout, les acteurs locaux de l'innovation. Il n'y a pas de licence idéale ou par défaut qui puisse être préconisée dans l'absolu : licence propriétaire ou open-source ? Dans quels cas envisager une concession de licence ? Comment mettre en place une rétribution dans le cadre d'une collaboration avec un partenaire industriel ? La protection par brevet a-t-elle un sens pour mon projet ? **Avant sa diffusion, un logiciel doit forcément faire l'objet d'une réflexion sur la question de la licence et de sa protection.**

Autre constat : un découplage apparaît entre la gestion administrative et juridique du logiciel (par les SATT et les DR) et sa gestion plus scientifique (par l'institut et même les sections), qui conduit à une gestion morcelée et incomplète de la production et de la valorisation des logiciels. **Il est souhaitable d'encourager les interactions entre ces acteurs.**

## 5 L'évaluation de la recherche

- C'est le nerf de la guerre.
- il faut revaloriser la production logicielle.
- Il est difficile de prendre en compte tous les types de production logicielle.

Pour ses logiciels, Inria utilise un modèle d'évaluation multi-critères assez abouti basé sur l'audience, l'originalité, la maturité, l'évolution - maintenance et la distribution - licensing. Chacune de ces 5 catégories propose quatre à cinq degrés de réalisation parmi lesquels l'auteur choisit celui qui convient le mieux. Pour compléter, la contribution personnelle de l'auteur est caractérisée à la fois qualitativement (design-architecture / coding-debugging / maintenance-support) et quantitativement (pas impliqué / contributeur occasionnel / contributeur régulier) [6].

En annexe B, le modèle complet d'Inria est disponible. Il peut être repris tel quel dans le cadre de toute enquête ou la production logicielle est concernée. Il convient bien en particulier pour tout auteur déjà familier de la production et de la diffusion de logiciels.

La section 6 cite directement le référentiel Inria dans ses critères pour l'évaluation, le recrutement et la promotion des chercheurs et y ajoute des conditions précises notamment sur la disponibilité du logiciel à des fins de test [7].

La section 7 prend en compte la production de prototype et de logiciels à la fois au niveau de l'évaluation des contributions scientifiques mais aussi dans un critère spécifique de valorisation basé sur "l'impact technologique et/ou économique des activités de recherche sur contrat ou de consultation, dépôts de brevets, participation à des créations d'entreprises" [8]. Par ailleurs, la section 7 fournit des conseils pour la présentation de la production logicielle [9].

Les sections sont les relais naturels d'expression par les auteurs de la production logicielle des unités du CNRS et sont en première ligne pour favoriser cette expression au moment des dossiers de promotion notamment. Par ailleurs, s'adressant au public large et hétérogène de tous les candidats aux concours, elles sont aussi intéressées par cette information au moment du recrutement. Les sections 6 et 7 constatent malheureusement un faible taux de dossiers mettant clairement en valeur la production logicielle. Le modèle d'évaluation proposé par Inria, assez sophistiqué, peut décourager des auteurs occasionnels ou novices. C'est pourquoi le groupe de travail propose une grille alternative, plus épurée.

Suite aux réflexions générées par la journée thématique, et à l'analyse des grilles d'évaluations existantes, la proposition d'un modèle simple de qualification des logiciels, similaire à celui proposé par la section 7, spécifiant 3 niveaux de maturité, détaillés à la section 3 a l'avantage de couvrir la quasi-totalité de la production logicielle scientifique. Rapide à renseigner, il inciterait les auteurs à

citer davantage leurs logiciels quel que soit leur niveau de maturité et à les mettre en avant dans leur CV. **L'objectif de cette grille est d'être largement diffusée et utilisée dans tout formulaire institutionnel (INS2I, sections 6 ou 7, CID51 ou même CNRS tous instituts confondus etc ...), là où l'information est pertinente : crac, candidature concours, enquête, première étape dans une démarche de dépôt APP etc ...**

Ainsi, nous proposons, en annexe C, une grille très simple de caractérisation applicable à tout logiciel. Cette grille, plus sommaire que celle d'Inria, permet de cerner la catégorie de logiciel en donnant une première idée de sa maturité et de sa qualité.

## 6 Conclusions

Ce document présente un certain nombre de préconisations, reprises ci-dessous :

1. Cartographier le patrimoine logiciel des unités de l'INS2I :
  - En se rapprochant des DR pour consolider les données existantes de dépôts APP.
  - En incitant l'utilisation de référentiels de caractérisation des logiciels dans le cadre de procédures existantes : Crac, évaluations menées par les sections, dépôts APP ...
2. Encourager l'utilisation d'outils d'hébergement de code.
3. Inciter à la réflexion sur le choix des licences logicielles avec les acteurs du transfert technologique locaux
4. Développer les contacts entre les producteurs de code dans les laboratoires et les promoteurs d'innovation dans les structures locales (DR ou SATT...)

Quels dispositifs peuvent être activés, promus par l'INS2I ?

- Mise en oeuvre de missions doctorales pour aider aux dépôts APP dans les labos.
- Plans d'actions coordonnées sur un an ou deux avec communication ciblée auprès des labos : avant/en cours/après : cartographie des logiciels ou incitation à l'utilisation de forges ou éléments de réflexion sur le choix des licences...
- Création et animation d'un cercle interne à l'institut constitué de membres issus des laboratoires, en particulier de l'activité "appui/soutien logiciel à la recherche", lorsqu'elle existe, pour formaliser un socle commun.
- Nomination de correspondants valorisation, développeurs logiciels internes au laboratoire, chargés de relayer et de diffuser ce type d'actions sur le terrain, au plus près des acteurs.

# Fiches pratiques

## A Comment se former aux outils d'hébergement de code ?

Les deux tableaux 1 et 2 donnent quelques clés pour positionner la problématique de l'hébergement des codes sources, qui est la première question à se poser pour produire un code potentiellement valorisable.

Pour acquérir une formation dans ce domaine, quatre types de ressources sont possibles :

- Utiliser la documentation native du logiciel en ligne.
- Utiliser un tutorial plus pédagogique. OpenClassroom en propose une grande variété par exemple.
- S'adresser à la communauté locale des développeurs logiciels, s'il elle existe à proximité. Le réseau DevLog fédère des réseaux plus locaux qui proposent régulièrement des journées d'initiation à ces outils : [devlog.cnrs.fr](http://devlog.cnrs.fr).
- Faire appel à des sociétés de formation qui dispensent des prestations payantes sur catalogue.

Gestionnaire de version	CVS, un des plus anciens	svn	Team Foundation Server	git	mercurial
typologie	le stockage des dépôts se fait sur un seul serveur centralisé		Solution propriétaire Microsoft	architecture des serveurs de dépôt répartie. Clonage aisé des dépôts	
documentation officielle		<a href="http://svnbook.red-bean.com">svnbook.red-bean.com</a>	<a href="http://www.visualstudio.com/fr/tfs">www.visualstudio.com/fr/tfs</a>	<a href="http://git-scm.com/docs">git-scm.com/docs</a>	
Tutoriels open-classroom		<a href="http://gerez-vos-projets-a-l-aide-du-gestionnaire-de-versions-subversion">gerez-vos-projets-a-l-aide-du-gestionnaire-de-versions-subversion</a>		<a href="http://gerer-son-code-avec-git-et-github">gerer-son-code-avec-git-et-github</a>	

TABLE 1 – Table des principaux gestionnaires de version

Site web de la forge	<a href="http://gforge.inria.fr">gforge.inria.fr</a>	<a href="http://sourcesup.renater.fr">sourcesup.renater.fr</a>	<a href="http://sourceforge.net">sourceforge.net</a>	<a href="http://forge.git.de.laboratoire">forge git de laboratoire</a>	<a href="http://github.com">github.com</a>	<a href="http://bitbucket.org">bitbucket.org</a>
gestionnaire de la forge	Inria	GIP Renater : communauté académique française	Geeknet (US)	à installer soi-même et gérer chez soi	GitHub (US)	Atlassian (Australia)
Forge utilisée proposant aussi wikis, bug tracking, forums ...	Gforge	FusionForge	SourceForge	GitLab		
gestionnaires de version sous-jacents supportés par la forge	svn, git...	svn, git...	svn, git...	git only	git only	git, mercurial
particularités	seul un agent inria peut créer un dépôt	offre de service Renater	inscription et charte d'utilisation	à définir soi-même	inscription libre	
dépôts publics ou privés?	les deux sont possibles	les deux sont possibles	public, pour logiciels open-source, attention publiciels		gratuit si publics	

TABLE 2 – Table des principales forges logicielles

## Fiche d'autoévaluation du logiciel (modèle Inria)

Software Name :

Summary: 

A	SO	SM	EM	SDL	/	DA	CD	MS	TPM
---	----	----	----	-----	---	----	----	----	-----

### B.1 - Caractériser le logiciel

**Critères A (Audience)**

- A-1 Internal prototype.
- A-2 Used by people in the team or close to the team.
- A-3 Ambitious software, usable by people outside the team.
- A-4 Large audience, used by people outside the team.
- A-5 Wide audience, large user's community.

**Critères SO (Software originality)**

- SO-1 None.
- SO-2 Minor contributions to existing software, reusing known ideas.
- SO-3 Original software reusing known ideas and introducing new ideas.
- SO-4 Original software implementing a fair number of original ideas.

**Critères SM Critères SO (Software maturity)**

- SM-1 Demos work, loose documentation.
- SM-2 Basic usage works, terse documentation.
- SM-3 Well-developed software, good documentation, reasonable software engineering.
- SM-4 Major software project, strong software engineering.
- SM-5 High-assurance software, certified by an evaluation agency or formally verified.

**Critères EM (Evolution and Maintenance)**

- EM-1 No real future plans.
- EM-2 Basic maintenance to keep the software alive.
- EM-3 Good quality middle-term maintenance.
- EM-4 Well-defined and implemented plans for future maintenance and evolution.

**Critères SDL (Software distribution and licensing)**

- SDL-1 None.
- SDL-2 Privately distributed within the close community.
- SDL-3 Distributed to industrial partners in a contractual setting.
- SDL-4 Public source or binary distribution on the Web.
- SDL-5 External packaging and distribution, as part of a popular open source distribution or a commercially-distributed product.

### B.2- Caractériser la contribution de l'auteur

**Design and Architecture (DA), Specify if you are :**

- DA-1 Not involved
- DA-2 Nn occasional contributor a
- DA-3 Regular contributor
- DA-4 A main contributor

**Coding and Debugging (CD), Specify if you are :**

- CD-1 Not involved
- CD-2 An occasional contributor
- CD-3 A regular contributor
- CD-4 A main contributor

**Maintenance and Support (MS), Specify if you are :**

- MS-1 Not involved
- MS-2 An occasional contributor
- MS-3 A regular contributor
- MS-4 A main contributor

**Team/Project Management (TPM), Specify if you are :**

- TPM-1 Not involved
- TPM-2 An occasional contributor
- TPM-3 A regular contributor
- TPM-4 A main contributor

### B.3 Complements d'information

<b>Software's web site (if available):</b>		
<b>Objective of the software</b>	<i>Domain addressed and precise technical goals</i>	
	<i>Type of the software (autonomous, part of a library, part of a larger system)</i>	
	<i>Performance goals, functionalities/algorithms, dissemination, etc</i>	
<b>Users community : type (research, education, commercial/industrial, large open source community, etc.), size, how the community is animated, etc.</b>		
<b>Your assessment of the software's impact in your research community and your user community.</b>		
<b>Description of the state of the art, placement of the software w.r.t. the competition.</b>		
<b>Size of the software (lines of code), languages used, size of the development team, development effort, list of patents and main publications related to the software.</b>		
<b>Your contribution to the software (as a team or as an individual).</b>		
<b>Any meaningful additional information</b>		

## C Caractérisation simplifiée d'un logiciel

La fiche, unique pour un logiciel rassemble les informations suivantes :

- Intitulé du logiciel
- Problématique résolue
- Page web / documentation
- Licence
- Date du dernier dépôt APP
- Auteurs
  - Nom - Prénom
  - Institution
  - Participation et pourcentage : conception, développement...
- Date de création
- Historique des versions
- Contexte de réalisation : projet de recherche académique, contrat industriel, collaboration locale, nationale, internationale ...

## Références

- [1] AC. Letournel. un état des lieux, sept 2016.
- [2] MA. Saint Jalmes. Production et valorisation des logiciels, sept 2016.
- [3] X. Leroy. L'évaluation du logiciel à Inria, sept 2016.
- [4] M. Fitzgibbon. Atelier Thématique "Transfert et Valorisation Logicielle", sept 2016.
- [5] M. Beaudouin-Lafon. Plateformes d'envergures en Sciences de l'information, nov 2015.
- [6] Inria Evaluation Committee. Criteria for software self-assessment, 2015.
- [7] Section 6 du comité national du CNRS. Critères pour l'évaluation, le recrutement, la promotion des chercheurs, dec 2016.
- [8] Section 7 du comité national du CNRS. Critères d'évaluation de la section 07 proposés pour la mandature 2016 - 2021, dec 2016.
- [9] Section 7 du comité national du CNRS. Conseils de la Section 7 du CoNRS pour la présentation des productions logicielles, oct 2016.

## Fiche d'autoévaluation du logiciel (modèle Inria)

Software Name :

Summary: 

A	SO	SM	EM	SDL	/	DA	CD	MS	TPM
---	----	----	----	-----	---	----	----	----	-----

### B.1 - Caractériser le logiciel

**Critères A (Audience)**

- A-1 Internal prototype.
- A-2 Used by people in the team or close to the team.
- A-3 Ambitious software, usable by people outside the team.
- A-4 Large audience, used by people outside the team.
- A-5 Wide audience, large user's community.

**Critères SO (Software originality)**

- SO-1 None.
- SO-2 Minor contributions to existing software, reusing known ideas.
- SO-3 Original software reusing known ideas and introducing new ideas.
- SO-4 Original software implementing a fair number of original ideas.

**Critères SM Critères SO (Software maturity)**

- SM-1 Demos work, loose documentation.
- SM-2 Basic usage works, terse documentation.
- SM-3 Well-developed software, good documentation, reasonable software engineering.
- SM-4 Major software project, strong software engineering.
- SM-5 High-assurance software, certified by an evaluation agency or formally verified.

**Critères EM (Evolution and Maintenance)**

- EM-1 No real future plans.
- EM-2 Basic maintenance to keep the software alive.
- EM-3 Good quality middle-term maintenance.
- EM-4 Well-defined and implemented plans for future maintenance and evolution.

**Critères SDL (Software distribution and licensing)**

- SDL-1 None.
- SDL-2 Privately distributed within the close community.
- SDL-3 Distributed to industrial partners in a contractual setting.
- SDL-4 Public source or binary distribution on the Web.
- SDL-5 External packaging and distribution, as part of a popular open source distribution or a commercially-distributed product.

### B.2- Caractériser la contribution de l'auteur

**Design and Architecture (DA), Specify if you are :**

- DA-1 Not involved
- DA-2 Nn occasional contributor a
- DA-3 Regular contributor
- DA-4 A main contributor

**Coding and Debugging (CD), Specify if you are :**

- CD-1 Not involved
- CD-2 An occasional contributor
- CD-3 A regular contributor
- CD-4 A main contributor

**Maintenance and Support (MS), Specify if you are :**

- MS-1 Not involved
- MS-2 An occasional contributor
- MS-3 A regular contributor
- MS-4 A main contributor

**Team/Project Management (TPM), Specify if you are :**

- TPM-1 Not involved
- TPM-2 An occasional contributor
- TPM-3 A regular contributor
- TPM-4 A main contributor

### B.3 Complements d'information

<b>Software's web site (if available):</b>		
<b>Objective of the software</b>	<i>Domain addressed and precise technical goals</i>	
	<i>Type of the software (autonomous, part of a library, part of a larger system)</i>	
	<i>Performance goals, functionalities/algorithms, dissemination, etc</i>	
<b>Users community : type (research, education, commercial/industrial, large open source community, etc.), size, how the community is animated, etc.</b>		
<b>Your assessment of the software's impact in your research community and your user community.</b>		
<b>Description of the state of the art, placement of the software w.r.t. the competition.</b>		
<b>Size of the software (lines of code), languages used, size of the development team, development effort, list of patents and main publications related to the software.</b>		
<b>Your contribution to the software (as a team or as an individual).</b>		
<b>Any meaningful additional information</b>		

# Sciences de l'Information et Arts

## Document de synthèse

### Conseil scientifique de l'INS2I

**Groupe de travail:** Marie-Paule Cani, Inbar Fijalkow, Michel Beaudouin-Lafon

Avec la participation de: Gérard Assayag, Rémi Ronfard, Philippe Gaussier

## 1. Introduction

Alors que le numérique se répand dans toutes les sphères de notre société, son interaction avec le monde de l'art ouvre de nouvelles applications passionnantes. Nous présentons ci-dessous un panorama des collaborations entre les laboratoires de l'INS2I et le monde de l'art au sens large, regroupées selon la nature de nos contributions scientifiques. Pour mieux identifier les applications en rupture, les verrous scientifiques et les nouveaux défis posés à notre discipline, le CSI a organisé une demi-journée thématique lors de sa réunion du 15 Mai 2017, en invitant trois experts issus respectivement des domaines du "visual computing" (informatique graphique et vision par ordinateur) pour Rémi Ronfard (LJK, UMR 5224, Grenoble), de l'informatique musicale pour Gérard Assayag (STMS - UMR 9912, Paris), et de la robotique pour Philippe Gaussier (ETIS, UMR 8051, Cergy-Pontoise). Ce document de synthèse reprend et intègre ces différentes contributions, permettant de déboucher sur une identification des enjeux, des risques et d'un ensemble de préconisations concrètes.

## 2. Panorama des collaborations ART-STIC

Il y a de nombreuses façons pour l'informatique de collaborer avec le monde de l'art, pris au sens large. Un premier travail a été d'identifier les secteurs avec lesquels nous interagissons et la nature de nos contributions. Dans cette partie, nous nous appuyons sur une classification en quatre points pour présenter un panorama des collaborations existantes, selon qu'il s'agisse de travailler sur des **contenus préexistants** ou de contribuer à la **création de nouveaux contenus**, et, dans chacune de ces catégories, selon qu'il s'agisse de systèmes utilisés par les professionnels de l'art seuls, ou en direct, par ou devant un public.

Les partenaires des collaborations évoquées ci-dessous vont des chercheurs en histoire de l'art, des musées et d'autres organismes de médiation scientifique aux artistes et créateurs eux-mêmes dans les domaines des beaux-arts, du cinéma, de la musique ou du spectacle vivant (danse, opéra...).

### 2.1 Travail sur des contenus préexistants

**Secteur des partenaires du monde de l'art :** Chercheurs en histoire de l'art, Conservateurs (physico-chimistes), Musées, Organismes de médiation culturelle.

**Nature des contributions des laboratoires INS2I :** Conservation, aide à la restauration et médiation pour l'art et le patrimoine culturel : Il s'agit d'aider à l'analyse et à la conservation d'œuvres

existantes (numérisation, restauration, reconstruction, indexation,...), et/ou d'aider à leur présentation au public (musées virtuels, interactifs ou augmentés).

**Structuration :** Ce domaine est largement structuré en région parisienne par le Labex *PATRIMA* dont les partenaires sont la fondation des sciences du patrimoine, l'Université Cergy-Pontoise, l'Université de Versailles Saint Quentin, des laboratoires du ministère de la culture (C2RMF, CRC, IPANEMA) et des instituts patrimoniaux (Louvre, Quai Branly, Château de Versailles, BNF, ...). Le réseau Ile-de-France (DIM) « matériaux anciens et patrimoniaux » vient d'être lancé en mars 2017 pour renforcer cette dynamique (contact : Dan Vodislav d'ETIS pour les sciences informatiques). De nombreux autres projets ont lieu en France, comme par exemple l'ANR eRoma (LIRIS Lyon, LJK Grenoble, Musée Fourvière Lyon, U. Paris Sorbonne).

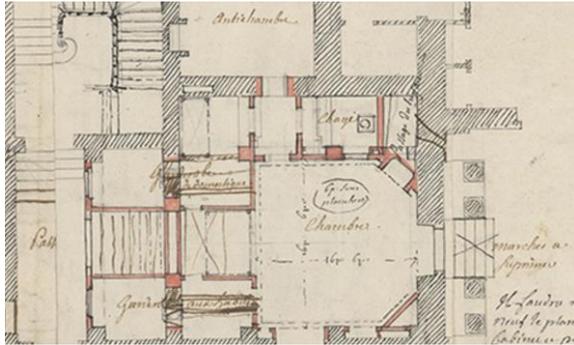
**Moyens matériels:** Ils sont nombreux. A titre d'exemples, nous pouvons citer la plateforme de numérisation *ExRealis* d'ICUBE (Strasbourg), l'équipe-plateforme de réalité virtuelle *Immersion* (IRISA, Rennes), ou l'équipEx PATRIMEX pour l'imagerie laser et les bases de données.

## **A. Conservation, restauration, indexation des œuvres**

Les recherches en informatique portent sur la numérisation 2D et 3D non invasive (tomographie, rayons X, ultrasons, scanners 3D etc), le traitement d'images [CEPS17], de vidéos, la reconstruction 3D, l'indexation dans des bases de données ou des ontologies, et l'analyse de données massives (avec un volet apprentissage statistique), l'amélioration des méthodes nécessaires pour mener à bien ces projets « en taille réelle » conduisent régulièrement à des publications, par exemple dans le domaine du traitement de données 3D issues de scanner [HDC17].

### **Quelques exemples:**

- Utilisation de l'imagerie acoustique pour comprendre l'altération des marbres du jardin du château de Versailles (SATIE) ;
- Numérisation d'objets ethnographiques (coupes rituelles, masques, statuettes Dogon du Mali) à des fins éducatives (ICUBE) ;
- Numérisation et/ou restauration virtuelle de statues de la cathédrale de Strasbourg (ICUBE) à Paris et au musée Fourvières de Lyon (projets ANR Art3D, Eros3D, et actuellement eRoma, LIRIS Lyon).
- Reconstruction 3D à partir de plans et de mesures (projet VERSPERA, ETIS), ou par vision par ordinateur (projet POMPEI, DI ENS Paris).
- Analyse et classification par apprentissage statistique de documents écrits médiévaux (GREYC, LITIS, ETIS, LABRI), décryptage des lettres de Marie-Antoinette (projet REX, ETIS).
- Création d'une base de données patrimoniale annotée et ouverte (ETIS, EquipEx PATRIMEX) ; recherche par le contenu visuel dans des bases de données patrimoniales (ANR Culture3DCloud, ETIS).
- Logiciel Lignes de Temps (IRI / Centre Pompidou) pour l'analyse cinématographique.



Projet Vespera : ETIS



Projet Pompéi, DI ENS

## B. Médiation culturelle

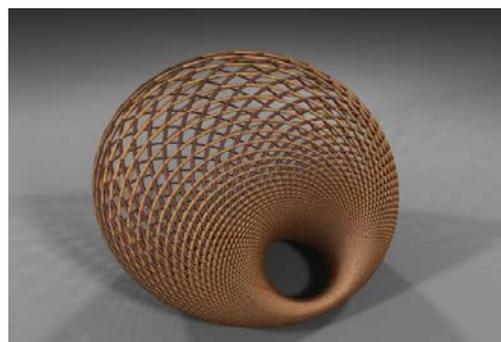
Nos contributions scientifiques portent sur l'amélioration de techniques de réalité virtuelle ou augmentée pour la création et la navigation dans des musées virtuels, et sur la création de nouveaux contenus interactifs ou animés, venant compléter les œuvres brutes disponibles dans les musées traditionnels. Les publications portent sur les nouvelles techniques de création de contenus, d'interaction et d'immersion, par exemple musicale [SNG14].

### Quelques exemples:

- Apprentissage du sens esthétique par le robot Berenson (ETIS et musée du Quai Branly) [VG14].
- Animation 3D de statues et bas-reliefs (ANR eRoma, LIRIS, LJK, Université Paris-Sorbonne, Musée Gallo-Romain de Lyon Fourvière).
- Mathématiques artistiques, génération d'images et d'impressions 3D (CRISTAL, Lille).
- Visite en réalité virtuelle de l'histoire de la musique (JMS 2016, équipe Hybrid, Irisa et équipe-plateforme Immersia, Rennes, équipe "Arts : pratiques et poétiques" de l'Université de Rennes 2).



Visite en RV de l'histoire de la musique, Immersia, Irisa



Cyclide de Dupin, CRISTAL

## 2.2 Création de nouveaux contenus artistiques

**Secteurs des partenaires du monde de l'art :** Beaux-arts (dessin, peinture, sculpture...), Musique, Cinéma, Spectacle vivant (danse, opéra...).

**Nature des contributions des laboratoires d'INS2I :** Nos laboratoires contribuent à la création de nouvelles œuvres soit via la conception de systèmes d'aide à la création pour les médias traditionnels, soit en co-concevant avec les artistes de nouveaux média destinés à servir de support à de nouveaux types d'œuvres, comme par exemple des spectacles mêlant des éléments réel et virtuels.

**Structuration du secteur et moyens :** Le financement direct de certains laboratoires par le ministère de la culture, comme l'IRCAM à Paris ou ACROE-ICA à Grenoble, favorise la structuration de ce secteur, en particulier grâce au financement d'artistes en résidence. Une initiative d'Inria, le groupe « Inriart » a tenté de structurer ce domaine de 2012 à 2016, mais n'a pas été reconduite. La Diagonale Paris-Saclay finance des projets arts-science et organise le festival annuel Curiositas. De nombreuses autres initiatives art-sciences sont confrontées à l'absence de moyens ou au manque de cohérence dans le temps entre les moyens obtenus par les parties informatiques et SHS/Art des projets.

### C. Aide à la création de contenus artistiques traditionnels

Les systèmes d'aide à la création de nouvelles œuvres pour les médias traditionnels peuvent relever de recherches en informatique musicale, en informatique graphique (dessin, sculpture) [SCC13, SCV16, CDT16, MVH17], en animation par ordinateur [BRB14], ou en traitement d'images et de vidéo en ce qui concerne la création théâtrale [RES15, KGR17]. Ici, le dispositif numérique mis en place peut changer les pratiques des professionnels, mais ne fait pas partie de l'œuvre lors de sa présentation.

#### Quelques exemples

- Aide à la conception cartographique, à partir de sources d'inspiration artistiques (peintures, mouvements artistiques, cartes anciennes). (ANR MapStyle, LJK, IGN/, LaSTIG, IRIT/)
- Systèmes de sculpture virtuelle (LIRIS et LJK)
- Outils d'aide à la composition musicale (IRCAM)
- Aide à l'animation de personnages virtuels expressifs (LJK)
- Aide à la cinématographie 3D via un système de prévisualisation de l'effet de profondeur des caméras 3D (LJK),
- Captation et indexation de répétitions de théâtre pour l'aide à la mise en scène (projets Scenoptic et ANR Spectacles-en-ligne, LJK, théâtre des Celestins à Lyon, LIRIS, CERILAC)



Spectacles en ligne (LJK, CERILAC, LIRIS)



Sculpture virtuelle (LIRIS et LJK)

## D. Création de nouveaux média pour l'expression artistique

Le dispositif numérique fait ici partie intégrante de l'œuvre, ouvrant la place à des œuvres interactives, interagissant en temps-réel avec les artistes et/ou avec le public. Les recherches portent sur la captation de mouvements collectifs ou individuels, la reconnaissance de son ou du mouvement en temps-réel, sur l'interaction (IHM), et sur la création ou la sélection de contenus, à la volée, que ce soient des sons, des images 2D ou 3D animées ou des mouvements de robots souples. Les publications peuvent être soit dans des revues techniques [PC15], soit également dans des revues d'art-science comme *Leonardo* (MIP Press) [BGC17].

### Quelques exemples :

- Systèmes de co-crédation musicale artistique et numérique par improvisation croisée (IRCAM)
- Installation artistique participative « la mer est mon miroir » mettant en place un système de réalité virtuelle contrôlé en temps-réel par l'imagerie cérébrale, pour produire un océan virtuel reflétant les émotions du participant (Lab-STICC et SPECULAIRE).
- Installations interactives de réalité augmentée pour les spectacles vivants, ANR INGREDIBLE (Lab-Stic Brest), Gardien du temple (LIMSI, LadHyX, Laboratoire Victor Vérité), Vera Iconica (LIMSI et Laboratoire Victor Vérité). D'autres projets sont en cours de montage au LIMSI et à l'ISIR sur l'intégration d'acteurs virtuels et de visages parlants dans différents types de spectacles.
- Danse, vidéo et réalité augmentée interactives sont explorés dans les projets SKIN, Radical Choregraphie RCO, et Ergonomics (UPSUD, LRI). Les recherches portent sur l'interaction basée sur l'analyse de mouvement. Robotique interactive dans les spectacles de Blanca Li, « robot » en 2015.
- Le projet Opera ID de Cristal combine réalité virtuelle et musique, avec l'interaction avec un personnage virtuel chantant (recherche en captation de mouvement et en informatique graphique). Projet similaire au Lab-STICC de Brest dans le cadre des percussions corporelles.
- Un média pour la création et l'exploration en réalité virtuelle de dessins 3D a été créé au CRISTAL avec une artiste plasticienne. Les recherches portent sur l'interaction réel-virtuel.
- D'autres média de type dessin ou sculpture permettent la création d'œuvres d'art numériques animées, comme celles réalisées au XLIM (Poitiers), dans le projet « matière de l'image » [ARL17].
- D'autres œuvres d'art animées prennent la forme de robots souples, grâce aux recherches en simulation, contrôle fabrication 3D et interaction au laboratoire CRISTAL.



La mer est mon miroir, lab-Sticc, et RCO, LRI.

Dessin 3D et robots souples, CRISTAL

### 3. Diffusion des recherches

La plupart des collaborations art/informatique citées conduisent à des contributions scientifiques de bonne facture dans les domaines de recherche des scientifiques partenaires. La publication se fait dans les revues et conférences majeures de ces domaines, par exemple CHI en IHM ou SIGGRAPH en informatique graphique.

En complément, un certain nombre de conférences ou de numéros spéciaux de journaux, sont plus particulièrement dédiés à la présentation de collaboration Art/sciences. Ils sont plus propices aux publications pour lesquelles l'accent est mis sur une nouvelle application ou un média innovant utilisant l'état de l'art du domaine scientifique. Par exemple, dans le domaine du traitement du signal et des images, on peut citer :

- le numéro special « signal processing for art investigation » de IEEE Signal Processing Magazine en 2015.
- La session spéciale « traitement du signal et des images pour l'art et le patrimoine » au colloque national GRETSI 2015
- La journée « Traitement du signal et des images pour l'art et le patrimoine » du GDR ISIS en Mai 2016
- La special issue on « image processing for cultural heritage », journal of electronic imaging 2017,

Dans les domaines de l'informatique graphique, de la visualisation et de la réalité virtuelle (GDR IGRV), les revues et conférences les plus orientées art/sciences sont :

- Le colloque annuel “Expressive” (ACM SIGGRAPH - Eurographics) regroupant les conférences NPAR (rendu et animation non photo réalistes), SBIM (sketch-based interfaces and modeling), et CAe (Computational Aesthetics)
- La conférence « Cultural Héritage” et le journal de l'ACM du même nom, ainsi que le « workshop on Graphics and Cultural Heritage » d'Eurographics.
- La revue Art/science « Leonardo » dont les publications sont présentées sous forme de « Art papers » lors de la conférence d'informatique graphique « Siggraph », permet la publications d'articles orientés vers la contribution artistique.

Notons que dans le domaine de l'informatique musicale, il n'y a pas lieu de faire la distinction entre conférences ou journaux majeurs du domaine et ceux spécialisées sur les applications artistiques. On peut citer :

- SMC Sound and Music Computing,
- JMM, The Journal of Mathematics and Music
- CMC, Computer Music Conference
- MCM, Mathematics and Computation in Music Conference

Voir la bibliographie à la fin de ce rapport pour quelques exemples de publications.

## 4. Avancées majeures et sujets en rupture

Lors de la séance du conseil INS2I du 15 Mai 2017, trois intervenants ont présenté des exemples d'avancées récentes, ainsi que leur vision des grands défis et des sujets en rupture : Rémi Ronfard (LJK, informatique et arts visuels: cinéma, théâtre), Gérard Assayag (IRCAM, informatique et musique), et Philippe Gaussier (ETIS, IA, robotique et modélisation du sens esthétique). Nous synthétisons ci-dessous ces réflexions, qui mettent en évidence des enjeux transversaux pour notre discipline.

### 4.1 Arts visuels

Les points fondamentaux sont la numérisation de la chaîne de production cinématographique ainsi que la convergence entre cinéma et spectacle vivant, avec l'émergence de nouveaux types de média pour lesquels il s'agit d'orchestrer les interactions entre acteurs réel et numériques :

- Le cinéma d'animation 3D est déjà un nouveau média art-science créé depuis 30 ans ; pour le cinéma traditionnel, les effets spéciaux à base numérique ont maintenant très largement pris le pas sur les méthodes classiques. Les recherches en informatique graphique et en édition d'images numériques sont immédiatement prises en compte et intégrées dans ces deux types de chaînes de production;
- Quel que soit le type de film, la salle de tournage est maintenant une salle informatique, ouvrant la porte à de nombreuses avancées dans les domaines de l'aide au cadrage et au montage : des outils en cours de développement sont dédiés à faciliter le contrôle des nouvelles caméras 3D en simulant la sensation de profondeur offerte au spectateur ; d'autres automatisent le recadrage et proposent des montage semi-automatiques ce qui permettra à court terme d'automatiser la création d'un film à partir d'un spectacle vivant.
- Les progrès récents en apprentissage statistique pour la synthèse de visages parlants permettront à court terme de doubler de manière invisible les acteurs, de changer leur expression faciale, voire de les remplacer par d'autres.
- On assiste à l'émergence de dramaturgies et de spectacles vivants numérique reposant sur la synchronisation entre acteurs réels et virtuels (exemple : projection vidéo sur le masque de mannequins). L'enjeu est d'orchestrer cette synchronisation, tout en laissant une certaine marge de liberté aux acteurs. Pour d'autres média à mi-chemin entre film et jeu vidéo, comme par exemple les films en réalité virtuelle, il s'agit de laisser le spectateur explorer tout en le guidant pour qu'il puisse découvrir la suite de l'intrigue. Ces nouveaux média s'appuient sur une recherche conjointe en capture et analyse temps-réel de mouvements, compréhension des actions humaines visuelles motrices et linguistiques, détection d'événements, interaction, synthèse d'images temps réel et direction d'acteurs virtuels animés de manière procédurale.
- Enfin, la création même de ces nouvelles formes d'art visuel demande le développement d'outils d'aide au concepteur, généralisant ceux qui facilitent le passage du scénario au storyboard, du storyboard à l'animation et du 3D au montage : il s'agira de remplacer le scénario par de véritables partitions scéniques, et de proposer de nouveaux média pour concevoir, pré-visualiser et finir de mettre au point ces nouveaux types de narration.

Ces sujets en rupture, qui soulèvent des enjeux artistiques et sociétaux, demandent une grande synergie des recherches dans les domaines du traitement du signal et des images, de la vision par ordinateur et de l'apprentissage, de l'IA (analyse et synthèse de comportements et des expressions faciales), de l'informatique graphique, de la réalité virtuelle et augmentée et de l'interaction homme-machine.

## 4.2 Informatique et musique

Si ce secteur est plus avancé que les arts visuels pour la mise en place de collaborations durables art-science, il connaît le même développement rapide vers plus d'interactivité, de jeu et d'improvisation conjointe entre musiciens réels et virtuels. Ici, le besoin d'aide à la création et d'exécution de nouveaux types de partitions ont donné lieu à des recherches qui irradient toutes les strates des sciences numériques, des modèles mathématiques et symboliques à l'apprentissage statistique et de la logique aux langages de programmation dédiés pour le temps-réel :

- Les représentations du son vont du bas niveau (signal) aux représentations symboliques et sémantiques. Faire communiquer ou hybrider ces différents types de représentations devient en enjeu majeur pour la création musicale. De même, les représentations du temps en musique, vont, comme en informatique, du temps discret (pulsation) au temps continu (signal). Ainsi, une création musicale peut être vue comme la définition d'une trajectoire (temporelle) dans une matrice de « matériaux » sans signification temporelle. De nombreuses recherches en informatique musicale portent sur les systèmes d'aide à la création, en lien avec les domaines de la sémiotique, de l'esthétique, de la linguistique (dont sont issues les notions de structures profondes et de surface), de cognition (représentations symboliques et sub-symboliques).
- Permettre à un système numérique de jouer une partition prédéfinie avec des musiciens réels impose un effort de recherche important sur la boucle analyse-synthèse : il s'agit d'analyser le signal, pour se resynchroniser et adapter la tonalité si nécessaire en fonction du jeu réel.
- Plus ambitieux, des systèmes émergents d'open-music s'attaquent au problème de la co-création d'une œuvre. Ici, il n'y a pas de partition à suivre pour le système informatique, mais éventuellement un apprentissage préalable à partir d'exemples - comme par exemple l'apprentissage des standards du jazz. Pendant la séance de co-création, il s'agit d'analyser le signal sonore produit par le musicien réel et parfois d'interpréter ses gestes, d'identifier les structures de haut niveau pour interagir en temps-réel via la production de nouvelles structures symboliques spatiales et temporelles – tenant compte des aspects harmonie et mélodie, elles-mêmes donnant lieu à un son numérique. La programmation de tels systèmes passe par des modèles d'apprentissages, par exemple Bayésiens, et par de l'informatique réactive, synchrone ou asynchrone. L'un des enjeux est de modifier la mémoire du système artificiel pour tenir compte des données apprises, mais aussi de ce qui vient ou est en train d'être joué.
- La création musicale peut devenir massive et distribuée grâce aux objets intelligents détenus par les spectateurs comme leurs téléphones portables.

### 4.3 L'IA et la médiation artistique

Des recherches récentes en robotique cognitive se sont attaquées à la création d'une « esthétique artificielle », comme dans l'exemple du robot français *Bergson*, testé en interaction avec des visiteurs au Musée du quai Branly. Si un robot « visiteur de musée » ne semble pas avoir d'utilité économique directe, il constitue un outil pour redistribuer l'attention des visiteurs, les questionner sur notre propre appréciation des œuvres et représente une installation artistique en soi, interactive et très attractive pour le public, [KGV13] [VG14].

Lorsqu'elle se focalise sur le cerveau, l'intelligence artificielle tend à oublier le social. Cette dimension a cependant une grande importance dans l'émergence du sens esthétique. Ainsi, le robot *Berenson* a été doté des capacités d'apprendre et de développer progressivement des compétences en jugement esthétique en observant le visage des visiteurs du musée. Le robot est muni de mécanismes permettant de suivre les regards, de reconnaître et d'imiter les expressions du visage, ce qui lui permet à la fois d'apprendre et de communiquer.

---

## 5. Synthèse et Préconisations

Le monde de l'art constitue un **terrain d'expérimentation extrêmement riche** pour les recherches en informatique. Ces collaborations revêtent également un fort enjeu sociétal, à la fois en ce qui concerne la conservation et la présentation au public du patrimoine artistique que lorsqu'il s'agit de favoriser l'émergence de nouvelles formes d'art. Les exemples que nous avons décrits permettent de faire émerger un certain nombre de **défis majeurs pour notre discipline** dans des domaines variés, de l'analyse de données multimodales massive à la création graphique 3D et la réalité virtuelle ou augmentée, et de la capture et reconnaissance des signaux humains, aux langages réactifs et à l'IA.

### Quatre manières de collaborer

Comme c'est le cas pour l'ensemble de nos interactions avec les sciences humaines, il existe quatre manières pour des chercheurs en informatique d'interagir avec le monde de l'art :

- 1) **l'ingénierie au service de l'art** : nous agissons dans ce cas comme des "techniciens", par exemple en permettant l'utilisation de nos plateformes de numérisation et en offrant nos compétences en reconstruction géométrique pour numériser des statues ;
- 2) **l'art au service de l'ingénierie** : l'art peut servir de justification pour des recherches en informatique, permettant de motiver dans une publication et d'illustrer nos travaux sur de nouvelles méthodes d'analyse de données, de nouveaux mécanismes d'interaction ou d'immersion, des algorithmes de sculpture numérique ou des méthodes d'apprentissage originales;
- 3) **la convergence d'enjeux artistiques et scientifiques** : certains projets permettent de faire avancer les recherches dans les deux domaines à la fois, comme ceux qui combinent histoire de l'art et animation graphique, ou certains projets de médiation permettant de défendre un point de vue sur les sciences et notre société.
- 4) **la fertilisation croisée entre art et sciences**. Les avancées dans nos domaines peuvent conduire à l'émergence de nouvelles formes d'art, mises en place conjointement avec des artistes ; et réciproquement, le questionnement de l'artiste peut permettre de reformuler

certains de nos problèmes "classiques" tout en élargissant le cadre (par exemple, les impressionnistes et la perception visuelle nous questionnent : avons-nous besoin des contours pour percevoir un objet ?).

## Risques

- **Etre considérés comme de prestataires au service de l'art :** Si les projets de type 1) peuvent servir de démonstrateurs d'un savoir-faire, et éventuellement nourrir un premier contact débouchant ensuite sur des collaborations plus équilibrées, ils ne sont pas à encourager sur le long terme, en particulier s'ils demandent un investissement en ressources humaines de notre part sans conduire à de nouvelles recherches en informatique : il y aurait le risque à terme que nos laboratoires de recherche soient utilisés comme des prestataires de service à bas coût. Une suggestion est d'orienter les chercheurs impliqués régulièrement dans ce type de partenariat vers des cellules de valorisation et transfert.
- **Faire des recherches isolées, sans que le monde artistique n'en bénéficie réellement:** Les projets de type 2) sont extrêmement nombreux, car ils peuvent être couverts par les financements classiques des laboratoires et sont facilement valorisables sous forme de publications. Cependant, malgré la bonne volonté des chercheurs, si une application artistique sert d'alibi sans contact préalable avec des artistes, la collaboration voit rarement le jour une fois la publication parue. Les raisons sont le manque de temps pour démarcher des partenaires d'autres domaines alors que ce n'est plus une priorité, et de financement pour transférer les résultats des recherches sous la forme de systèmes exploitables par le monde de l'art.
- **Ne pas réussir à obtenir de financement synchronisé :** si les projets de type 3) et 4) sont les plus fructueux pour les deux communautés, leur principal frein est la difficulté à obtenir des financements synchronisés dans le temps pour les partenaires artistiques et informatiques. Par exemple, un artiste ayant obtenu une résidence d'un an ne pourra pas travailler de manière fructueuse si – outre les chercheurs responsables – un ingénieur n'est pas financé pour réaliser les développements nécessaires. La difficulté est encore accrue lorsque la collaboration demande le suivi dans le temps d'une installation pour assurer son fonctionnement et analyser son impact sur le public. Comme souvent, le principal écueil n'est pas tant le financement du matériel nécessaire que celui des intervenants.
- **L'aspect 'hors cadre' et en particulier, la difficulté de validation des recherches :** Il n'est pas toujours facile de prédire quelle science "sérieuse" peut ressortir d'une collaboration art-informatique, ni de faire déboucher ces recherches sur des publications. Lorsque l'avancée scientifique consiste à développer un média totalement nouveau pour l'expression artistique, comme par exemple pour un projet de type 4, comment valider ces recherches ? Plusieurs des collaborations citées sont confrontées à cet écueil.

## Préconisations concrètes

- **Soutenir les collaborations art science :** Nous tenons à rappeler l'importance de la création artistique pour nos sociétés. Encourager ce type de collaborations entre dans notre mission en

tant qu'institut de recherche au même titre qu'en ce qui concerne les missions de valorisation et transfert, qui irriguent d'autres secteurs. Si collaborer avec le monde de l'art tout en sortant des sentiers battus présente de véritables risques scientifiques, il y a aussi potentiellement de grands gains. Les artistes repoussent les limites de la technologie, notamment informatique, et se l'approprient de façon inattendue, ce qui fournit des sources intéressantes de problèmes et de questions de recherche. Le CNRS se doit d'encourager les chercheurs à favoriser les recherches scientifiques à risque par rapport, par exemple, à la prestation de service.

- **Tenir compte de leur transversalité en informatique :** Il serait très profitable d'offrir des lieux et des occasions de rencontre transversaux aux sous-disciplines informatiques (comme par exemple une journée « arts et informatique »), de manière à favoriser la fertilisation croisée. Cela permettrait par exemple aux avancées récentes en musique, comme les partitions interactives et leurs nouveaux langages, les systèmes d'aide à la création, d'irradier les recherches de nature similaire dans le domaine des arts visuels, qui s'attaquent aux partitions scéniques et aux systèmes d'aide à la mise en scène. Cela permettrait aussi de travailler conjointement sur de nouveaux défis comme la créativité computationnelle, et la co-créativité, qui reposent sur le fait de capter et comprendre les signaux humains, et de combiner connaissances à priori, apprentissage et transfert de style, en tenant compte de la structure temporelle multi-échelle d'une œuvre et la mise en œuvre multimodale de ces principes.
- **Mettre en place des appels à projet interdisciplinaires :** La mission interdisciplinaire du CNRS peut jouer un rôle déterminant pour permettre des financements synchronisés des deux communautés, en ouvrant des défis orientés art et sciences de l'information sur deux ou trois ans. Les financements interdisciplinaires classiques devront cependant être complétés en termes de personnel ou synchronisés avec d'autres sources pour permettre, par exemple, le suivi et l'analyse d'une installation dans un musée grâce à l'embauche d'un ingénieur polyvalent, ou la présence d'un artiste en résidence pendant le temps du projet.

## 8. Références

*(cette liste donne quelques exemples de publications faisant suite à des projets Art/Informatique. Elle n'est pas destinée à être exhaustive, mais à témoigner de la diversité des contributions scientifiques amenées par les collaborations de l'informatique avec le monde de l'art)*

[AKSJ15] P. Abry, A.G. Klein, W.A. Sethares, and C.R. Johnson, Jr., special issue on « signal processing for art investigation », *IEEE Signal Processing Magazine*, 2015.

[ARL17] E. Andres, L. Richaume, G. Largeteau-Skapin (2017). Digital Surface of Revolution with Hand-Drawn Generatrix. *Journal of Mathematical Imaging and Vision*. doi:10.1007/s10851-017-0708-6 2017.

[BRB14] Adela Barbulescu, Rémi Ronfard, Gérard Bailly, Georges Gagneré, and Hüseyin Cakmak (2014). Beyond basic emotions: expressive virtual actors with social attitudes. *Motion in Games*, 2014.

[BGSA15] Bigo, L., Ghisi, D., Spicher, A., Andreatta, M. "Representation of Musical Structures and Processes in Simplicial Chord Spaces", *Computer Music Journal*, 39(3):11-27, 2015 [hal-01263299]

[BGC15] Panagiotis-Alexandros Bokaris, Michele Gouiffès, Véronique Caye, Jean-Marc Chomaz, and Christian Jacquemin (2015). Gardien du Temple: concealment and revelation. An interactive installation involving poetry, performance and spatial augmented reality. *Leonardo*, 2015.

[BNCB17] Dimitri Bouche, Jérôme Nika, Alex Chechile, Jean Bresson. Computer-aided Composition of Musical Processes. *Journal of New Music Research*, Taylor & Francis (Routledge), 2017, 46 (1).

[CEPS17] A. Chetouani, R. Erdmann, D. Picard, F. Stanco, special issue on « image processing for cultural heritage », *journal of electronic imaging* 2017.

[CDT16] Christophe S., Dumenieu B., Turbet J., Hoarau C., Mellado N., Ory J., Loi H., Masse A., Arbelot B., Vergne R., Brédif M., Hurtut T., Thollot J., Vanderhaeghe D. (2016). Map Style Formalization: Rendering Techniques Extension for Cartography. *Expressive 2016*, May 2016, Lisbonne, Portugal.

[HDC17] A. Hamdi-Cherif, J. Digne, R. Chaine (2017). Super-resolution of Point Set Surfaces using Local Similarities. *Computer Graphics Forum*, 2017.

[KGR17] Moneish Kumar, Vineet Gandhi, Rémi Ronfard, Michael Gleicher (2017). Zooming On All Actors: Automatic Focus+Context Split Screen Video Generation. *Computer Graphics Forum*, Wiley, April 2017.

[KGV13] Karaouzene, A., Gaussier, P., & Vidal, D., A robot to study the development of artwork appreciation through social interactions. *Third IEEE ICDL-Epirob Conference on Development and Learning and Epigenetic Robotics*, 2013.

---

[MVH17] Mellado N., Vanderhaeghe D., Hoarau C., Christophe S., Brédif M., Barthe L. (2017) Constrained Palette-Space Exploration, *ACM Transaction on Graphics* (Siggraph 2017).

[NCA17] Jérôme Nika, Marc Chemillier, Gérard Assayag. ImproteK: introducing scenarios into human-computer music improvisation. *ACM Computers in Entertainment*, 2017, . <hal-01380163>

[PC15] Parenthoen M., Catala , L. (2015). The sea is your mirror. *ACM Motion in Games conference*. 2015.

[RES15] Rémi Ronfard, Benoit Encelle, Nicolas Sauret, Pierre-Antoine Champin, Thomas Steiner, Vineet Gandhi, Cyrille Migniot, Florent Thiery (2015). Capturing and Indexing Rehearsals: The Design and Usage of a Digital Archive of Performing Arts. *Digital Heritage*, Sep 2015, Grenade, Spain. (Best paper award).

[SNG14] Laurent S. R. Simon, Florian Nouviale, Ronan Gagne, Valérie Gouranton (2014) Sonic interaction with a virtual orchestra of factory machinery. *Sonic Interaction for Virtual Environments*, Mar 2014, Minneapolis, United States. pp.1-6, 2014

[SCC13] Lucian Stanculescu, Raphaëlle Chaine, Marie-Paule Cani & Karan Singh (2013). Sculpting multi-dimensional nested structures. *Computers and Graphics*, vol. 37, n°6, pp. 753-763.

[SCV16] Artur P. Sampaio, Raphaëll Chaine, Creto A. Vidal & Joaquim B. Cavalcante-Neto (2016). Temporally Coherent Sculpture of Composite Objects. *Computers and Graphics*, vol. 57 Issue C, pp. 118-127.

[VG14] D Vidal, P Gaussier, Un robot comme personne. Ontologies comparée et expérimentale au musée du quai Branly, Terrain. *Anthropologie & sciences humaines*, 152-165

## Synthèse et propositions du nouveau champ d'investigation sur les Systèmes d'Information en Interaction avec, ou centrés sur, l'Humain (S2IH)

Cette fiche de synthèse fait suite au [Séminaire Cyber-Physical&Human-Systems \(CPHS\) organisé en séance du CSI INS2I, le 27 février 2017 de 14:30 à 16:30](#). Les thèmes suivants ont été présentés : « Introduction sur les principaux défis CPHS », par Mariana NETTO (IFSTTAR) ; « Effets indésirables associés à la délégation de fonctions dans les CPHS: les enjeux de l'analyse des risques » par Frédéric VANDERHAEGEN (CNRS-LAMIH) ; « Les facteurs humains dans les véhicules autonomes : questions actuelles et perspectives » par Franck MARS (CNRS-LS2N) ; et « Contrôle d'oscillations cérébrales en boucle fermée et perspectives » par Antoine CHAILLET (CNRS-L2S).

### SYNTHESE

Les interventions et coopérations humaines avec des systèmes hautement automatisés, complexes et en réseau, et la conception de technologies centrées sur les humains, sont devenues des défis d'une importance majeure pour répondre aux exigences de compréhension et de progrès, voir par exemple [1,3]. **Des solutions nouvelles et fondamentales sont nécessaires pour permettre aux humains d'interagir avec ces systèmes, et vice-versa...** Les S2IH émergents seront coordonnés, distribués et connectés, et doivent être robustes et adaptables à des environnements perturbés. **Les Systèmes d'Information en Interaction avec, ou centrés sur, l'Humain (S2IH) de demain devront dépasser, de très loin, les systèmes d'aujourd'hui en termes de capacité, de stabilité, d'adaptabilité, de résilience, de fiabilité, de sécurité, de facilité d'utilisation et de coût. Les S2IH de demain seront aussi innovants, avec la création de nouveaux outils pour l'amélioration de l'humain aussi bien pour ses capacités physiques (amélioration des personnes à mobilité réduite par exemple) que mentales voire intellectuelles...** Ces progrès ne pourront pas être accomplis sans la **collaboration multidisciplinaire** qui doit inclure **des spécialistes de la cognition, de l'ergonomie, des sciences sociales et humaines, de l'informatique et de l'automatique**, ainsi que beaucoup d'autres experts. Sans aucun doute, **les aspects sur l'éthique et sécurité devront accompagner ces nouvelles recherches.**

À l'heure actuelle, on peut répertorier les niveaux de coopération S2IH comme suit:

#### 1. Symbiose Homme-Machine/Matériel

- Les prothèses intelligentes (autonomes / communicantes)
- La neurostimulation
- Les exosquelettes
- Les implants biomédicaux
- ...

#### 2. Les humains comme opérateurs de systèmes d'ingénierie complexes

- L'aéronautique, l'interaction Homme-machine dans les avions
- L'automobile
- Les systèmes de production
- La chirurgie robotique
- Les trains semi-automatisés
- Commande à distance d'équipes robotisées, par exemple lors d'interventions en milieu hostile
- ...

#### 3. Les humains comme agents dans les systèmes multi-agents

- Le transport routier intelligent
- La gestion du trafic aérien
- L'usine du futur
- La robotique d'assistance
- Les réseaux de systèmes d'énergie
- La mobilité urbaine
- ...

#### 4. Les humains comme éléments dans les systèmes contrôlés

- Le contrôle de confort dans les maisons
- Les villes intelligentes
- La robotique de sauvetage
- Les dispositifs d'assistance
- Les infrastructures intelligentes
- Les bâtiments connectés
- Les équipements s'auto-adaptant au handicap (permanent ou temporaire)
- ...

#### 5. Les humains comme éléments dans les systèmes d'information et de données

- Le profilage des utilisateurs
- La médecine personnalisée
- La surveillance et la détection
- La détection de fraudes
- ...

#### 6. Les humains face aux interactions homme-machine (IHM)

- Activités humaines assistées par ordinateurs ou par robots
- Ordinateurs ou robots assistés par les humains
- Systèmes innovants de captage d'information pour le travail coopératif (e.g., interface ordinateur-cerveau, communication par la pensée)
- Amélioration des connaissances  
(e.g., recherche *Big Data* ou *Knowledge Discovery and Data mining*)
- Risques de dissonances d'usages d'IHM
- ...

**Ci-dessous sont résumées quelques problématiques illustrant certains des champs précédents d'investigation :**

**Véhicule autonome :** La problématique du véhicule autonome vu comme un problème systémique Homme-véhicule-environnement nécessite une approche pluridisciplinaire à l'articulation entre différents champs des STIC et des sciences du comportement humain, en particulier l'ergonomie cognitive. Le véhicule autonome est une continuité technologique avec l'intégration de briques techniques qui arrivent à maturation, mais c'est aussi une véritable rupture en termes d'usage, car l'Homme n'est plus au cœur de la boucle de contrôle du système. Deux grandes questions se posent: celle de la surveillance et de la supervision du conducteur, qui nécessite la fusion de plusieurs éléments de diagnostic (postures, état perceptivo-moteur, cognitif et émotionnel, vigilance et endormissement, etc.), et celle de la gestion des interactions avec le véhicule (retour d'information au conducteur sur l'état du système, identification des intentions d'action au niveau tactique en anticipation de l'action sur les commandes, pertinence et répartition du partage de contrôle des commandes lors des transitions). Un verrou transversal à ces problématiques est la modélisation du conducteur par des modèles globaux qui représentent l'ensemble des processus dynamiques impliqués dans la conduite, par exemple, des modèles hiérarchiques, adaptatifs, génériques ou multi-échelle. Il est également impératif de prendre en compte i) les interactions véhicule-environnement et notamment les questions relatives aux marges de sécurité (cohérente avec celles de l'humain à bord et à l'extérieur), ii) l'adaptation à une situation qui peut amener à violer les règles prescrites (code de la route) et iii) l'étude de la responsabilité en cas d'accident impliquant un véhicule à conduite déléguée. Des problématiques liées au niveau d'acceptation et de confiance de l'humain envers le système seront également à prendre en considération.

**Robotique médicale :** (prothèses intelligentes, implants biomédicaux, neurostimulation, exosquelettes...) : Plusieurs défis relèvent à la fois du traitement du signal, de la robotique et de l'automatique comme le monitoring et l'évaluation en temps-réel du mouvement et de l'activité musculaire par EMG (électromyographie) et IMU (Inertial measurement Unit), l'utilisation de réseaux de capteurs auto-calibrables, filaires ou non, le support individualisé en fonction de la situation par stimulation électrique fonctionnelle (FES). Par ailleurs, l'intégration du cloud servira à superviser les patients, améliorer le système par apprentissage via des données venant d'autres utilisateurs et des données de localisation (terrains, montées, escaliers, ...).

**Stimulation cérébrale en boucle fermée :** Les quinze dernières années ont vu des progrès significatifs dans notre compréhension du cerveau grâce à une collaboration intensive entre experts en neurosciences et en STIC. Des méthodes issues des systèmes dynamiques et de la théorie de la commande sont devenues nécessaires pour certains verrous scientifiques en neurosciences, du fait de l'omniprésence des processus dynamiques impliqués aux différentes échelles du fonctionnement cérébral et de leurs interactions. Ces interactions donnent lieu à des boucles fermées complexes impliquant des systèmes aux dynamiques souvent non-linéaires et à des échelles spatiales ou temporelles diverses. La complexité qui en résulte impose l'utilisation de méthodes nouvelles d'identification paramétrique, de réduction de modèles, et d'analyse formelle du comportement des boucles fermées. Si le bouclage est omniprésent dans le fonctionnement naturel du cerveau, des boucles artificielles peuvent être ajoutées à des fins de traitement thérapeutique. Les moyens technologiques permettant d'influer sur l'activité neuronale peuvent désormais prendre de nombreuses formes, et incluent les stimulations électriques (stimulation cérébrale profonde), magnétique (stimulation magnétique transcrânienne), chimique (médicamenteuse), et optique (optogénétique). La boucle est alors fermée lorsque ces signaux de stimulation dépendent de mesures en temps réel de l'activité du sujet, telle que l'activité musculaire ou l'activité électrique dans certaines zones cérébrales. Les avantages cliniques attendus d'une stimulation cérébrale en boucle fermée sont nombreux, notamment en termes d'efficacité, de robustesse, d'adaptabilité et de parcimonie. De telles approches sont en cours de développement, notamment pour la maladie de Parkinson, mais la plupart des approches évaluées expérimentalement se cantonnent pour le moment à des stratégies simplistes du point de vue de l'automatique, telles que la stimulation « à la demande » qui n'est autre qu'une stimulation en boucle ouverte déclenchée dès qu'une activité pathologique est détectée. La possibilité d'influencer, voire de piloter, la réponse de structures neuronales ciblées en temps réel ouvre la voie à de nombreuses avancées thérapeutiques, notamment pour la maladie de Parkinson, l'épilepsie, la dépression chronique, l'obésité ou la maladie d'Alzheimer. Cela soulève des questions fondamentales pour la théorie de la commande, notamment en termes de commandabilité, d'observabilité, de stabilisation robuste en boucle fermée, et de commande optimale.

**Répartition des tâches (délégation) entre S2IH et opérateurs humains :** Quelques exemples pratiques montrent que parfois ce type de délégation peut générer des situations conflictuelles indésirables voire dangereuses :

- afin d'éviter une collision, un radar anticollision peut stopper un véhicule dans une zone dangereuse alors que le conducteur de celui-ci pourrait préférer éviter l'obstacle sans s'arrêter,
- la corrélation entre les éléments perçus par un conducteur et les informations données par un système d'aide à la navigation peut générer des erreurs d'orientation lorsque l'interprétation de la signalisation ou de la cartographie portent à confusion,
- le conditionnement d'un système d'apprentissage automatique peut générer des erreurs d'usage lorsque la configuration d'un objet dont l'interprétation est liée aux situations précédentes produit une nouvelle signification dans une situation similaire. L'usage d'un régulateur automatique de vitesse par un conducteur peut générer des affordances (i.e. usage d'objets conçus initialement pour réaliser des fonctions prédéfinies afin d'en créer de nouvelles), des contradictions (i.e., existence pour un même décideur d'actions opposées pour gérer deux situations différentes) ou des ingérences (existence d'actions opposées de différents décideurs). Dans les trois cas, face à des situations d'urgence par exemple, des risques potentiels peuvent être identifiés dus à une augmentation des temps de réponse ou de l'hypogilance.

D'une façon générale, une dissonance est un conflit entre connaissances ou éléments de connaissances, pour un individu donné, pour un groupe d'individus ou entre systèmes cyber-physiques et systèmes humains. Elle concerne donc les comportements, les attitudes, les croyances, les préférences, les intentions, les décisions, les actions, etc. L'occurrence d'une dissonance peut affecter des facteurs humains comme l'attention, la vigilance, le stress, la charge de travail, le confort, etc. Ainsi, la conception de systèmes cyber-physiques nécessite de prendre en compte les usages potentiels de ces systèmes par des opérateurs humains afin d'identifier les dissonances et les risques associés.

**Contrôle partagé :** Le contrôle partagé est un domaine d'ingénierie prometteur qui brise les frontières de l'Automatique actuelle, où les humains interagissent avec des robots ou des véhicules partiellement automatisés. Les principaux défis actuels de la robotique et de l'automatisation concernent des environnements moins structurés et plus imprévisibles, dans lesquels les humains opèrent naturellement, et surtout lorsqu'ils ont besoin d'interagir avec d'autres humains. En conséquence, de nombreux systèmes semi-automatisés doivent aujourd'hui être supervisés par des opérateurs humains. L'interaction Homme-machine doit être complètement repensée en raison de l'augmentation de plus en plus rapide des niveaux d'automatisation de nombreux systèmes. Contrairement au paradigme du contrôle de supervision - où le contrôle est échangé entre l'humain et la machine - le paradigme du contrôle partagé suppose implicitement l'interaction entre deux ou plusieurs agents indépendants qui accomplissent ensemble une tâche pour atteindre un but commun. Cela implique que la conception du système de contrôle partagé n'est pas nécessairement centrée sur l'humain, mais est coopérative entre l'humain et le système. Dans les systèmes de contrôle partagé, tous les agents doivent être conscients des capacités, des faiblesses, de l'autorité des autres agents, qu'ils soient des machines ou des humains. Par conséquent, la communication réciproque des frontières opérationnelles de chaque agent, qu'il soit humain ou machine, est essentielle. Un autre exemple du contrôle partagé est celui résultant d'une forte interaction haptique entre l'humain et le système, e.g. la chirurgie à l'aide des robots. Le contrôle partagé est important aussi pour alléger la charge ou les efforts de l'humain. La conception de systèmes de contrôle pour les prothèses, e.g. les orthèses, est encore un autre exemple. Bien d'autres nouvelles technologies feront aussi appel au contrôle partagé.

**Détection de fraudes :** La détection de fraudes vise à détecter des comportements inhabituels d'utilisateurs, en les isolant par rapport aux comportements considérés comme normaux, plus nombreux. Les transactions bancaires frauduleuses peuvent par exemple être détectées, de même que des comportements anormaux sur des sites d'achats, etc. Néanmoins, le simple raisonnement d'associer anormal à frauduleux est un raccourci d'une réalité plus complexe, où des caractéristiques des opérateurs humains doivent être associés aux comportements afin d'en tirer des conclusions qui ont une certaine probabilité de correspondre à la réalité. Cependant les fraudeurs, ne sont pas naïfs et peuvent très bien chercher à influencer les informations les concernant afin de passer à travers les systèmes de détection, ce qui entraîne la nécessité d'avoir d'une part une détection performante, mais également de pouvoir conforter la détection par une analyse causale.

**Profilage et communautés :** Le profilage d'utilisateurs est, à titre d'exemple, un enjeu majeur dans le ciblage commercial et publicitaire. Peu d'actions commerciales échappent à la tendance de vouloir identifier précisément les habitudes et les intentions d'achats, à travers l'identification de communautés d'utilisateurs, de communautés d'intentions, ou d'une combinaison des deux. L'humain est de façon évidente un acteur de ce processus, qu'il ne contrôle cependant pas. Il existe cependant un risque important de réduction de la diversité, tant le profilage est guidé par des algorithmes uniques ou convergents, ainsi qu'un risque de remplacer certaines intentions humaines par des décisions algorithmiques. Il est donc important de comprendre et de contrôler le partage du processus de décision entre l'humain et l'algorithme dans les systèmes d'analyse de données.

## A. PROPOSITIONS

D'une façon générale, les technologies de pointe, couplées à l'humain, sous différentes formes et à différents degrés, donnent et donneront accès, à la fois, à des modèles dynamiques raffinés, et à des techniques de mesures et d'actionnement beaucoup plus précises et réactives, dans des environnements eux-mêmes dynamiques et incertains. De nouvelles thématiques scientifiques et des perspectives radicalement nouvelles s'ouvrent pour la communauté de l'INS2I, avec également la nécessité de nouvelles compétences en ingénierie.

**Dans ce contexte, il est par exemple suggéré des axes de recherche fondamentaux et études portant sur:**

- Les problématiques de capacité, à la fois de l'humain et des systèmes autonomes ou coopératifs, face à des événements imprévus ou sans précédents (éventuellement conduisant à un arrêt d'urgence car le système technique est incapable de déterminer une alternative fiable sans l'aide de l'opérateur humain et de sa créativité).
- Les études sur la communication des données, des connaissances, des préférences, des humains aux systèmes numériques et vice versa.
- les risques liés au fonctionnement de systèmes autonomes ou interagissant avec des humains (sécurité, cybersécurité, impacts environnementaux,...).
- La compréhension des différents cas d'usage que requièrent l'autonomie ou l'évolution des capacités de l'humains : augmentation (« l'humain augmenté ») ou compensation (déficience, handicap).
- les futurs capteurs et systèmes de traitement des S2IH effectuant l'acquisition d'informations multiples et multi-sources (caméras, micros, radars, oculomètres,...) avec des interprétations de situation sans ambiguïté dans un contexte opérationnel dynamique, variable et incertain.
- La compréhension et la régulation de l'interaction entre les processus décisionnels humains et le contrôle algorithmique dans les systèmes d'analyse de données liées aux personnes.
- L'indispensable modélisation de l'humain dans toute sa diversité (fortement demandée par les industriels, notamment pour les véhicules autonomes) et commande de ces nouveaux modèles.
- A l'interface entre neurosciences et automatique, pour une compréhension plus fine des dynamiques cérébrales et leur pilotage en boucle fermée
- ...

**Dans chacun des axes de recherche étudié, les questions éthiques, sociales et juridiques devront être prises en compte, avec une analyse de l'attitude des utilisateurs potentiels tenant compte à la fois de mesures objectives et subjectives, par exemple :**

- Détecter la fatigue et le niveau de stress des opérateurs humains
- Imposer la sécurité de l'environnement
- Assurer la protection de la vie privée et des données
- Faire en sorte que le niveau d'intelligence développé par la machine reste compréhensible par les opérateurs humains
- ...

---

[1] F. Lamnabhi-Lagarrigue et al., Systems & Control for the future of humanity, research agenda: Current and future roles, impact and grand challenges, Annual Reviews in Control (2017), 43, 1-69.

[2] Special Section on Cyber-Physical and Human Systems (M. Netto et S. Spurgeon, Eds), Annual Reviews in Control (2017), 44, 249–251.

[3] Cyber Physical & Human Systems:

CPHS 2016 <http://www.cphs2016.org>

CPHS 2018 <http://www.cphs2018.org>

Après avoir contacté leur responsable, les GDRs relevant principalement de l'INS2I concernés par la thématique S2IH sont à l'heure actuelle les suivants :

GDR717 Modélisation, analyse et conduite des systèmes dynamique(MACS)  
Isabelle QUEINNEC [isabelle.queinnec@laas.fr](mailto:isabelle.queinnec@laas.fr)

**Tous les thèmes**

GDR725 Réseaux et Systèmes Distribués (RSD)  
André-Luc BEYLOT [beylot@enseeiht.fr](mailto:beylot@enseeiht.fr)

**Thèmes 3 et 4**

GDR2018 Aspects formels et algorithmiques de l'Intelligence Artificielle (IA)  
Sebastien KONIECZNY [porteur@gdria.fr](mailto:porteur@gdria.fr)

**Tous les thèmes**

GDR3000 Informatique Géométrique et Graphique, Réalité Virtuelle et Visualisation (IG-RV)  
Dominique BECHMANN [bechmann@unistra.fr](mailto:bechmann@unistra.fr)

**Tous les thèmes**

GDR3072 Robotique (Robotique)  
Philippe FRAISSE [philippe.fraisse@lirmm.fr](mailto:philippe.fraisse@lirmm.fr)

**Thèmes 1, 2 et 3**

GDR2033 « Le Toucher : Analyse, Connaissance, simulaTion » TACT  
Betty SEMAIL [betty.semail@polytech-lille.fr](mailto:betty.semail@polytech-lille.fr)

**Thèmes 1, 5 et 6**

**Il est envisagé d'organiser un Colloque National sur les thématiques S2IH en 2020.**

# Séminaire thématique "*l'Internet des Objets*"

## Document de synthèse

Conseil Scientifique de l'INS2I

**Groupe de travail** : Véronique Cortier, Inbar Fijalkov, Philippe Lamarre, Fabrice Théoleyre (animateurs).

**Avec la participation de** : Mérouane Debbah (Huawei/Supélec), Andrzej Duda (LIG), Aurélien Francillon (EURECOM), Jean-Marie Gorce (CITI / INSA Lyon), l'axe rescom du GDR RSD (Eric Fleury, Thomas Noel), le pré-GDR sécurité (Marc-Olivier Killijian, Gildas Avoine), Sihem Amer-Yahia (LIG), le GDR MADICS (Christine Collet)

**Executive summary** : L'Internet connaît actuellement une révolution, en intégrant de petits objets intelligents, capables de réagir de façon autonome, donnant naissance à un monde cyber-physique (CHPS). De par le domaine d'application, l'Internet des Objets (IoT) requiert une approche pluridisciplinaire, allant de la conception des objets matériels à la conception d'applications intelligentes autonomes tout en intégrant les propriétés de sécurité, d'efficacité énergétique et de passage à l'échelle. Pour tenter d'identifier les verrous clé de ce domaine d'application, le CSI a organisé le 6 Décembre 2016 un séminaire de réflexion autour de cette thématique, en invitant plusieurs experts autour de quelques grandes thématiques identifiées comme clés pour l'INS2I dans l'IoT.

## Table des matières

<b>1</b>	<b>Remarques préliminaires</b>	<b>3</b>
<b>2</b>	<b>Introduction</b>	<b>3</b>
<b>3</b>	<b>Transmission radio</b>	<b>5</b>
3.1	Éléments de prospective . . . . .	5
3.2	Structuration nationale . . . . .	5
<b>4</b>	<b>La mise en réseau</b>	<b>6</b>
4.1	Éléments de prospective . . . . .	7
4.2	Structuration nationale . . . . .	7
<b>5</b>	<b>La sécurité pour des objets</b>	<b>9</b>
5.1	Éléments de prospective . . . . .	9
5.2	Structuration nationale . . . . .	9
<b>6</b>	<b>La gestion et le traitement des données</b>	<b>11</b>
6.1	Éléments de prospective . . . . .	11
6.2	Structuration nationale . . . . .	12
<b>7</b>	<b>Autres Domaines</b>	<b>13</b>
7.1	Ethique . . . . .	14
7.2	Capteurs matériels . . . . .	14
7.3	Interagir avec l'IoT . . . . .	14
<b>8</b>	<b>Propositions du CSI</b>	<b>15</b>
<b>9</b>	<b>Récapitulatif des préconisations</b>	<b>16</b>

## 1 Remarques préliminaires

Nous insistons sur le fait que cet texte sur la recherche dans le domaine de l'Internet des Objets **ne se veut pas comme exhaustif**. Il a pour but de donner des **exemples** d'équipes, thèmes, et verrous scientifiques associés au domaine de l'Internet des Objets. En particulier, certains thèmes et équipes, ne sont peut-être pas mentionnés dans ce document bien qu'ils travaillent sur ou touchent à la problématique de l'Internet des Objets.

**Nous encourageons ceux que nous avons oubliés ou même dont nous avons mal décrit les travaux (erreur, incomplétude, etc.) d'utiliser le formulaire à l'adresse <https://csins2i.irisa.fr/seminaire-thematique-internet-des-objets/> pour ajouter des précisions.**

Tout document de prospective scientifique du domaine est par exemple bienvenu. De même, nous encourageons tous les lecteurs à consulter les commentaires issus de ce formulaire à la même adresse.

## 2 Introduction

Goldman Sachs a pronostiqué une croissance du nombre d'objets connectés à Internet sur (2014- 2017) de 60%, avec au total 28 milliards d'objets connectés<sup>1</sup>. Cisco prédit que le marché de l'Internet des Objets (IoT) générera un chiffre d'affaire de 14,4 milliards de dollars en 2022<sup>2</sup>.

Cependant, la communauté est assez divisée sur la définition exacte à donner à l'Internet des Objets. Un petit équipement, plus ou moins intelligent, se connecte de façon continue ou intermittente à Internet. Un capteur va mesurer une grandeur physique, tandis qu'un actionneur va au contraire déclencher une action physique. De façon générale, il s'agit donc de l'extension de l'Internet, permettant de rendre les mondes virtuels et physiques plus perméables (Cyber Physical Systems).

Les objets peuvent présenter une diversité de caractéristiques très importante :

1. Une TV connectée représente un objet qui va récupérer des informations d'Internet afin de les afficher pour l'utilisateur.
2. Un objet fonctionnant avec une pile va mesurer une grandeur physique (température, pression), qu'il envoie à un entrepôt de données dans Internet, où l'information sera traitée. De même, de tels objets sont particulièrement pertinents pour la maintenance préventive, permettant de prévenir l'opérateur d'une panne (chaudière, vanne).
3. Un tag RFID sert à identifier un objet en transit, permettant de le localiser dans la chaîne de logistique. Cependant, il ne s'agit dans ce cas là que d'un objet passif, ne pouvant traiter de l'information.
4. Une voiture connectée permet de récupérer en temps-réel les conditions de circulation, interagit avec son environnement proche pour prévenir les accidents, utilise des données de capteurs embarqués pour faciliter la conduite.

Les actions récentes tendant à vouloir offrir une connectivité spécifique pour les objets. Ainsi, la 5G développe un ensemble de technologies offrant une connectivité sans-fil à faible débit pour des objets potentiellement mobiles, à faible consommation énergétique.

---

1. <http://www.goldmansachs.com/our-thinking/pages/iot-infographic.html>.

2. <http://www.forbes.com/sites/louiscolombus/2015/12/27/roundup-of-internet-of-things-forecasts-and-market-estimates-2015/>

Les industriels montrent un intérêt croissant dans l'Internet des Objets. Ainsi, l'IoT Valley<sup>3</sup> a vu le jour en 2011 à Labège pour fédérer les industriels et académiques travaillant autour de cette thématique. Un deuxième site est en cours d'aménagement pour faire face à l'engouement industriel, de nombreux acteurs venant rejoindre le projet.

**Proposition 1:** *Le CNRS devrait étudier l'opportunité de participer activement dans l'IoT Valley. Il existe une opportunité intéressante pour favoriser à travers cette structure les collaborations entre académiques et industriels.*

La communauté se focalise depuis quelques années sur les défis suivants :

**Énergie** : beaucoup des objets connectés ne sont pas reliés à une source d'alimentation continue. Eventuellement, des dispositifs de récupération d'énergie sont disponibles (solaire, cinétique, etc.) mais ne couvrent qu'une partie des besoins. Une grande partie de l'énergie étant consommée par l'interface radio, il est nécessaire de concevoir des solutions économes.

**Mobilité** : de nombreuses objets mobiles commencent à se connecter à Internet (ex : géolocalisation de véhicules, e-sport). Ils doivent pouvoir envoyer et recevoir des informations d'Internet tout en restant éteints le plus souvent.

**Densité** : les usages connaissant une croissance très forte, nous attendons des densités extrêmement importantes. Il est donc nécessaire que les nouveaux réseaux prennent en charge plusieurs milliers d'équipements par centaine de mètres carrés. Bien que chaque équipement envoie peu d'information individuellement, le volume agrégé crée des défis importants.

Nous avons choisi ici de nous focaliser dans ce document sur les activités scientifiques se développant sur l'Internet des Objets plutôt que sur les usages, regroupant un très vaste ensemble de domaines (smart grids électriques, climatologie, etc.). En effet, l'IoT est dans ce cas là considéré comme un *outil* et non comme un objet d'étude à part entière. Ce document est notamment inspiré des présentations réalisées au CSI du 5 décembre 2016 par Mériouane Debbah, Andrzej Duda et Aurélien Francillon<sup>4</sup>. Par ailleurs, le groupe de travail a choisi de se focaliser sur les quatre aspects suivants :

**Traitement du signal** : l'IoT repose sur des transmission radio, bas débit, à basse consommation pour des réseaux très denses.

**Réseau** : la contrainte énergétique est un paradigme nouveau dans Internet, requérant à la fois des protocoles frugaux et le support d'une connexion bidirectionnelle intermittente ;

**Sécurité** : les équipements embarqués présentent des capacités limitées, sans IHM pour leur configuration. Il est donc important de sécuriser l'IoT afin d'en pouvoir étendre les usages sans risque sur leur utilisation quotidienne dans des tâches de plus en plus essentielles ;

**Données** : l'IoT génère un très grand volume de données, qui doivent être traitées et filtrées au plus près des *producteurs*.

Enfin, pour chacun de ces domaines, nous avons essayé d'identifier des thèmes en rupture dans le domaine, spécifiques à l'Internet des objets. Nous avons également tenté de dresser un panorama des outils et de la structuration nationale de la communauté française sur ces thématiques. Nous sommes restés focalisés dans ce document sur la recherche académique.

---

3. <http://www.iot-valley.fr/fra>

4. L'ensemble des présentations est disponible sur le site du CSI — <https://csins2i.irisa.fr/seminaire-thematique-internet-des-objets/>

## 3 Transmission radio

Dans cette section, nous nous focaliserons sur les aspects bas-niveau de l'Internet des Objets, ou comment envoyer (et recevoir) avec des transmissions radio des informations générées par de petits objets autonomes en énergie.

### 3.1 Éléments de prospective

**Codage de canal** : les objets envoient des paquets contenant très peu d'information (quelque fois, une seule mesure, par exemple pour de la télé-relève). Il est donc très complexe dans ce cadre de faire un codage de canal efficace avec des codes courts ;

**Bande étroite avec signaux recouvrants** : la densité d'accès est très importante, avec des milliers de sources potentiellement en contention. Les nouvelles solutions essaient de gérer un accès en bande étroite, avec des signaux partiellement recouvrant : il s'agit de pouvoir décoder les différents signaux correctement : une retransmission coûte à la fois de la bande passante et de l'énergie. De nouvelles techniques d'accès non orthogonales restent donc à proposer ;

**Accès asynchrone** : ces très nombreux capteurs ne peuvent être synchronisés en temps pour accéder à la transmission, il faut chercher de nouveaux mécanismes d'accès multiple ;

**Géolocalisation** : la plupart des services innovants reposent sur une connaissance précise de l'objet émettant les données (tracking de flotte, etc.). Il s'agit donc d'utiliser la technologie radio de transmission pour effectuer également de la géolocalisation précise en 3D (hauteur et position géographique), en indoor ou outdoor (d'une cave à un parc) ;

**Bidirectionnalité et Full-duplex** : de plus en plus de noeuds combinant capteurs et actionneurs sont intégrés. Ainsi, une communication bidirectionnelle doit être établie, poussant les mesures, et récupérant les commandes à exécuter. Puisque l'équipement embarqué doit *s'endormir* (i.e. couper son interface radio), une transmission full-duplex serait particulièrement recommandée, afin de pouvoir simultanément émettre et recevoir et donc réduire les instants de réveil.

**Efficacité énergétique** : le problème énergétique est encore actuellement une barrière : à la fois concernant la consommation en pic, et en terme de longévité. Les techniques de machine learning sont prometteuses pour apprendre quand émettre, et comment.

**Latence faible** : certaines applications telles les voitures connectées demandent des délais d'accès au canal très faibles, requérant des techniques non OFDM pour la transmission.

### 3.2 Structuration nationale

#### 3.2.1 Plateformes

Nous avons identifié les plateformes suivantes pouvant servir d'outil de recherche à la communauté radio :

**FIT Cortex-Lab** (Lyon)<sup>5</sup> : la plateforme est la première à intégrer le software defined radio (SDR), la radio cognitive (CR) et les réseaux auto-optimisant (SON). La plateforme est ouverte en libre accès à la communauté ;

---

5. <http://www.cortexlab.fr/>

**OpenAirInterface** (Sophia-Antipolis)<sup>6</sup> : Eurecom a initié la création de cette plateforme, ayant pour but de virtualiser les plateformes de télécommunications 5G, incluant les aspects IoT.

### 3.2.2 GDR

Par ailleurs, les aspects *transmission* de l'IoT sont notamment traités par les GDR suivant :

**ISIS** : l'Internet des Objets est couvert principalement par l'axe D (Télécommunications : compression, protection, transmission), mais également les axes A (Méthodes et modèles en traitement du signal et de l'image) et C (Algorithme-architecture en traitement du signal et des images) ;

**Ondes** : devrait couvrir les nouvelles fréquences mmWaves.

### 3.2.3 Laboratoires & Équipes

Sans pouvoir être exhaustifs, nous avons identifié les laboratoires suivants travaillant dans le domaine :

**CITI** : l'équipe Socrate possède une expertise forte dans les communications radio et la radio logicielle pour l'IoT. Elle héberge Cortex-Lab, équipex pour les expérimentations radio dans l'IoT ;

**ETIS** : chaire IoT avec OrangeLabs sur la dynamique temporelle (théorie des jeux), codage pour paquets courts, one-bit massive MIMO ;

**LEAT** : les équipes CMA et MCSOC au LEAT s'intéressent au développement d'antennes, aux protocoles basse consommation et à l'interface matérielle/logicielle pour les objets communicants.

**IRIT** : certains chercheurs de l'équipe réseaux ont étudiés les problématiques de l'IoT, notamment les aspects de codage de canal ;

**LIG** : l'équipe Polaris s'intéresse notamment aux problèmes d'optimisation des communications radio.

**LSS** : le groupe LANEAS possède une compétence forte en radio, MIMO, et techniques d'accès pour l'IoT ;

**Télécom Lille** l'équipe *signal processing* étudie les problématiques liées à la modélisation du canal et de la couche physique ;

**Télécom ParisTech** : l'ERC starting grant de Michèle Wigger s'intéresse notamment à la 5G et à l'optimisation des communications d'objets radio coopérant entre eux ;

## 4 La mise en réseau

Dans ce sous-thème, nous nous focalisons aux aspects réseaux de l'Internet des Objets : quels protocoles et algorithmes sont nécessaires pour qu'une collection d'objets plus ou moins intelligents puissent échanger des informations via une interface radio tout en optimisant la consommation d'énergie ?

---

6. <http://www.openairinterface.org/>

## 4.1 Éléments de prospective

Nous avons identifié les points suivants, représentant des défis de rupture par rapport aux réseaux classiques, et l'Internet en général :

**Accès massif** : les scénarios actuels envisagent des densités de plus en plus élevées. Les technologies longues portée telles que LoRa ou Sigfox envisagent même de couvrir des zones de plusieurs kilomètres carré. Les techniques algorithmiques traditionnelles permettaient plutôt de gérer l'accès pour un nombre faible de flots envoyant beaucoup de données.

Le paradigme dans l'IoT change : il s'agit de gérer l'accès de très petits flots (quelques paquets par heure ou par jour) en très grand nombre (quelques milliers).

**Auto-configuration** : un équipement connecté à l'IoT peut avoir des capacités très limitées, et ne possède souvent pas d'IHM. Une configuration manuelle n'est donc pas envisageable, du fait même de la scalabilité des solutions ;

**Auto-adaptation** : les scénarios d'utilisation ont tendance actuellement à se multiplier, et de nouveaux usages sont chaque jour trouvés. Il est donc inconcevable de fixer a priori les conditions d'utilisation.

La propriété de versatilité est donc primordial : il s'agit de mesurer l'environnement, en estimer les caractéristiques, et basculer sur un algorithme / protocole permettant d'avoir des performances optimales dans la situation considérée.

**Garantie de performances** : de plus en plus d'applications demandent à ce que des contraintes applicatives soient respectées. Le service définit par exemple le délai et le taux de perte de bout en bout acceptables, et le réseau doit respecter ces contraintes. En particulier, le réseau doit pouvoir supporter une qualité de service différenciée, ne consommant de l'énergie que pour que les contraintes soient respectées ;

**Garantie de fonctionnement** : certaines applications critiques doivent pouvoir garantir qu'elles fonctionnent correctement quelle que soit la situation (véhiculaire, réseaux industriels de contrôle-commande). Les techniques de type vérification formelle sont dans ce cadre nécessaires pour vérifier le bon fonctionnement des protocoles mis en place.

## 4.2 Structuration nationale

### 4.2.1 Plateformes

Nous avons identifié les plateformes suivantes pouvant servir d'outil à la communauté réseau :

**FIT IoT-Lab**<sup>7</sup> (Grenoble, Lille, Paris, Rennes, Strasbourg) : 2 700 capteurs sont répartis sur 6 sites géographiques. Un libre accès est donné aux chercheurs, afin d'évaluer *in vivo* les performances des algorithmes et protocoles qu'ils proposent. La plateforme intègre à la fois des noeuds statiques dans une large variété de topologies, et des noeuds mobiles, dont la trajectoire est contrôlable à distance. IoT-Lab est labellisé IR ;

**DOMUS**<sup>8</sup> (LIG) : un appartement équipé permet de tester des hypothèses sur "*l'écosystème numérique et humain*". Bien que le focus soit donné aux usages, sur le comportement humain-numérique, des aspects réseau peuvent être introduits ;

**IS**<sup>9</sup> (LORIA) : la plateforme considère de nouvelles relations entre un système informatique, son utilisateur humain et son environnement. Un des axes de recherche de la plateforme concerne les "réseaux de capteurs".

#### 4.2.2 GDR

Les GDR suivants comprennent dans leur spectre de thèmes explorés la problématique de la mise en réseau pour l'Internet des Objets :

**MACS** : l'axe *Systèmes de commande et interactions* est particulièrement concerné, regardant les aspects contrôle/commande dans l'Internet des Objets ;

**RSD** : le pôle *Rescom* étudie les aspects réseaux, tandis que le pôle *Systèmes Distribués* se focalise plus sur les aspects intergiels, dont l'Internet des Objets représente un des domaines applicatifs.

Une réunion inter-GDR a été organisée en 2016 entre les GDR RSD et SoC-SiP sur la thématique de l'IoT.

**Proposition 2:** *Le CSI propose aux GDR d'organiser régulièrement des journées thématiques inter-GDR sur le problème de l'IoT, comme l'ont fait RSD & SoC-SiP en 2016.*

#### 4.2.3 Laboratoires & Équipes

Nous avons notamment identifié les laboratoires suivants traitant cet aspect des problèmes :

**CEA-LETI** s'intéresse aux petits objets communicants, avec une partie de ses chercheurs s'intéressant aux protocoles et à la conception d'objets efficaces, auto-alimentés (energy harvesting) ;

**CITI** : l'équipe Agora étudie les réseaux urbains, incluant les aspects IoT ;

**Eurecom** : David Gesbert porte une ERC advanced grant qui vise à mieux utiliser les ressources "calcul et mémoire" de tout objet connecté,

**ICube** : l'équipe réseau pour la partie sans-fil étudie le problème de l'IoT dans son cadre protocolaire et algorithmique. Elle héberge notamment une partie de la plateforme équipex IoT-Lab.

**INRIA Lille** : l'équipe FUN étudie les aspects IoT, et héberge une partie de la plateforme équipex IoT-Lab. Elle possède une compétence forte notamment en RFID, et en réseaux mobiles pour l'IoT ;

**INRIA Paris** : l'équipe EVA s'intéresse aux réseaux sans-fil non organisés, tels que les réseaux de capteurs ;

**INRIA Saclay** : l'équipe INFINE regarde les nouveaux paradigmes de communication, s'intéressant notamment aux problématiques liées au content-centric networking (ou comment utiliser les données comme paradigme de communication) ;

**IRISA** : l'équipe OCIF (Réseaux et Architectures pour les Technologies Innovantes de Communication) s'intéresse aux problématiques réseau dans le cadre des smart grids et du transport intelligent, touchant notamment à l'IoT.

L'équipe Dionysos s'intéresse également à l'IoT, avec un focus sur les aspects évaluation de performances ;

**LIG** : l'équipe Drakkar explore les problématiques des réseaux sans-fil, touchant notamment à l'Internet des Objets ;

**LIMOS** : le thème Réseaux de Capteurs de l'axe Systèmes d'Information et de Communication se focalise sur l'IoT ;

**LORIA** : l'équipe Madynes s'intéresse à la configuration, la surveillance et la sécurité du nouvel Internet, s'intéressant en partie à l'IoT ;

**SAMOVAR** : l'équipe R3S a travaillé sur la conception de protocoles et algorithmes efficaces en énergie.

## 5 La sécurité pour des objets

La sécurité des objets connectés peut se tourner en partie vers des technologies existantes. Certains problèmes sont relativement simples et peuvent bénéficier de solutions existantes, notamment lorsque les objets connectés sont en fait comparables à des ordinateurs (smartphones, tablettes, ...). D'autres problèmes sont plus spécifiques et demandent des solutions adaptées à des objets qui disposent de peu de moyens de calculs et de peu de ressources en énergie. Aux défis purement technologiques s'ajoutent plusieurs difficultés plus pratiques :

- Les constructeurs d'objets connectés ne sont pas spécialistes de la sécurité. Les personnes qui déploient des objets connectés sont encore moins spécialistes et n'activent pas toujours les protections (a minima, changement des mots de passe, génération d'une clé publique, etc.)
- Les techniques utilisées sont en général couvertes par le secret industriel. Il est difficile voire impossible pour le monde académique d'auditer les solutions et de les faire progresser.

### 5.1 Éléments de prospective

**cryptographie à basse consommation / capacité** : les équipements intégrés à l'IoT possèdent des capacités très limitées. A l'inverse, les équipements mis en oeuvre pour décrypter ont des ressources importantes (ex : cluster de calcul). Comment assurer un (dé)chiffrement facile sur de petits équipements, tout en rendant le décryptage compliqué ?

**zéroconf** : un grand nombre d'objets doivent pouvoir être configurés sans intervention humaine.

**analyse de vulnérabilité** : chaque objet est déployé en très grand nombre, rendant une attaque *rentable* (i.e. une vulnérabilité détectée peut compromettre un grand nombre d'objets). Comment détecter ces vulnérabilités et circonscrire leur impact ?

**frontière avec les SHS** : comment créer de la confiance dans un environnement distribué ? Il s'agit d'intégrer des usagers qui doivent accepter d'être surveillés numériquement, tout en gardant le contrôle sur les données générées et leur usage ;

**confiance** : des objets doivent pouvoir créer une confiance entre eux. Ainsi, un consommateur de données doit pouvoir avoir confiance dans un producteur. Cette confiance non binaire doit se traduire dans les algorithmes de traitement des données.

### 5.2 Structuration nationale

#### 5.2.1 Plateformes

Aucune plateforme n'est réellement spécifique aux aspects sécurité. Les plateformes utilisées pour valider les protocoles peuvent par exemple très bien être utilisées également pour

évaluer la sécurité (ex : sécurisation d'un protocole, implémentation d'algorithmes cryptographiques, etc.) Ainsi, la section 4.2.1 reste ici pertinente pour le domaine de la sécurité.

### 5.2.2 GDR

Au niveau de la communauté, le pré-GDR Sécurité a été créé en 2016 (<http://gdr-securite.irisa.fr/>). Un sous-groupe animé par Carlos Aguilar Melchor (<http://gdr-securite.irisa.fr/gt-sri.html>) se focalise notamment sur les aspects Sécurité des réseaux et des infrastructures, dont l'IoT. D'autres sous-groupe tels que *vie privée*, traitent des questions également présentent au sein des objets connectés.

### 5.2.3 Laboratoires & Équipes

**CITI** : l'équipe Privatics s'intéresse à la vie privée dans l'Internet en général, incluant les aspects IoT en partie ;

**CEA-LIST** s'intéresse à la sécurité, la gestion et la mise à jour de réseaux de capteurs sans-fil étendus.

**CRISTAL** l'équipe 2XS se focalise sur la conception de logiciels fiables ;

**ENS Paris** : l'équipe ISG possède une expertise forte en attaques (génération automatique d'exploits), botnets, etc. L'équipe s'est lancé depuis peu de temps dans le domaine de l'IoT ;

**GREYC** : l'équipe Monétique & biométrie s'intéresse à la sécurisation des transactions électroniques, rentrant ainsi dans le domaine de l'IoT (RFID) ;

**IRISA** : le LHS s'intéresse en particulier à la sécurité des cartes à puces, domaine pouvant être intégré dans le périmètre de l'IoT ;

L'équipe EMSEC cible les questions de recherche liées à la sécurité des systèmes informatiques et électroniques embarqué (RFID, cartes à puce, FPGA).

L'équipe CIDRE étudie les problèmes de détection d'intrusion, de préservation de la vie privée, et de la confiance ;

**Heudyasic** : l'équipe RO s'intéresse à la fois aux aspects infrastructure réseau et à la sécurité, intégrant notamment les aspects confiance ;

**INRIA Paris** : l'équipe Secret s'intéresse aux aspects cryptographie, notamment pour créer des fonctions sur du matériel à *bas-coût*.

**LAAS** : l'équipe TSF s'intéresse à la sécurisation d'objets intelligents (véhicules, TV, etc.)

**LABRI** : l'équipe MUSE (Mobilité, Ubiquité, Sécurité) s'intéresse à la sécurisation d'une flotte de drones, pouvant être considéré comme un IoT mobile ;

**Labsticc** : l'équipe SFIIS s'intéresse à la sécurisation, tant les aspects cloud (IaaS cloud) que communications radio (e.g. Wireless HART)

**LIRIS** : l'équipe DRIM s'intéresse au problème de la confiance et de la préservation de la vie privée ;

**LMV** : l'équipe CRYPTO s'intéresse à la cryptographie sur matériel embarqué.

**LORIA** : l'équipe Madynes s'intéresse aux problèmes de la sécurité des protocoles de l'IoT (déni de service, homme du milieu, etc.) L'équipe possède également une expertise forte en monitoring de réseaux distribués, dont l'IoT ;

L'équipe CARAMBA regarde les aspects cryptographiques, Marine Minier ayant regardé les aspects sécurité protocolaire pour l'IoT ;

**SAMOVAR** : l'équipe R3S s'intéresse à la sécurité dans l'IoT, étudiant des solutions adaptées à des terminaux et environnements contraints, ainsi que la sécurité des données personnelles ;

**XLIM** : l'équipe Cryptis regarde la sécurité et la cryptographie, appliquée à l'IoT par certains de ses membres.

**Proposition 3:** *Le CSI encourage l'émergence de collaborations focalisées sur l'Internet des Objets, au sein du pré-GDR sécurité. Ce domaine recouvre plusieurs des axes du pré-GDR. Ces interactions pourraient être favorisées par exemple par des journées thématiques, la création d'un axe, etc.*

## 6 La gestion et le traitement des données

L'Internet des Objets génère un large volume de données qui doivent être traitées, éventuellement en temps-réel afin de pouvoir interagir avec l'environnement. La problématique recoupe partiellement les problèmes du big data, avec des contraintes supplémentaires.

### 6.1 Éléments de prospective

**Vélocité** : dans l'Internet des Objets, les données sont produites en continu, et potentiellement en très grand nombre. Étant donné un problème, il est nécessaire de déterminer rapidement les données à considérer et de les traiter pour fournir un résultat fiable en un temps acceptable. En particulier, pour certaines applications comme par exemple le contrôle commande dans l'Internet des Objets Industriels (IIoT), les temps de réponse sont bornés.

**Véracité** : au-delà de la diversité et de l'hétérogénéité, les données produites par l'Internet des Objets sont souvent sujettes aux imprécisions. Les causes sont nombreuses : qualité des matériels, conditions d'usage, impact de l'environnement, etc. Dans ces conditions, obtenir des résultats de qualité (si possible qualifiable ou quantifiable) à un coût (ressources, temps) raisonnable, tout en conservant la possibilité d'adresser à la fois le volume et la vélocité, nécessite des méthodes proposant un équilibre entre efficacité et exactitude.

**Composante temporelle** : dans l'Internet des Objets les données sont produites par des objets différents avec des fréquences indépendantes. Elles sont acquises de manière asynchrones et ne sont pas temporellement alignées. Cela complique l'obtention de résultats de qualité à partir de la composition de données issues de sources différentes.

**Composante spatiale** : comme pour la dimension temporelle, les données (mesures) produites par les capteurs sont également très hétérogènes au niveau spatial (granularités différentes, relations spatiales complexes) et souvent difficile à intégrer avec une interprétation précise.

**Adaptabilité** : Le volume, la vélocité et la volatilité des données et des sources de l'IoT rendent indispensable l'adaptabilité des solutions génériques. Les questions liés à adaptabilité sont actuellement très majoritairement abordées au niveau des traitements. Cependant,

d'autres éléments ont un impact non négligeable sur cette dimension. Par exemple, la collecte des données a un effet évident sur la complétude et le volume des données à traiter. Les caractéristiques de l'IoT peuvent permettre/nécessiter de développer des approches plus globales de l'adaptabilité.

## 6.2 Structuration nationale

Ce thème semble peu couvert en France, le big data pour l'IoT étant visiblement peu étudié dans ses spécificités dans nos laboratoires. Néanmoins la France accueille des conférences du domaine comme ICLR 2017 (1000 participants dont les Gafa <http://www.iclr.cc>).

Une action interdisciplinaire ouverte par la MI CNRS MASTODONS en 2012, et poursuivie dans l'action EADM du GDR MADICS <http://sabiiod.org/EADM>, fédère une trentaine de chercheurs en traitement des masses de données environnementales ((bio)acoustique, qualité air, luminosité,...).

### 6.2.1 Plateformes

**SMIoT**<sup>10</sup> (PROTEE & LIS Toulon) : Smart Microsystems & Internet of Things, innove depuis 2016 en acquisition et mise en réseaux, traitements embarqués grande autonomie, des masses de données multimodales synchrones et véloces. Les systèmes SMIoT acquis par le CNRS, Parc Nationaux et des ONG concernent le suivi de la faune sous-marine et terrestre joint à la pollution environnementale (son, lumière, chimie), i.e. DAQ avancé 5 voies 2mHz Fe intégrant compas, GPS, transmission radio, GSM, Wifi, accéléromètre, thermo, hygromètre et luxmètre longue autonomie et haute dynamique pour le suivi nocturne. Plusieurs To sont ainsi traités par semaine.

**Proposition 4:** *Le CSI préconise de réfléchir au problème de la reproductibilité des expérimentations. Des scénarios et jeux de données pourraient par exemple être définis afin de servir d'étalon aux solutions, et d'identifier plus précisément les cas d'usage et solutions adaptées. Ce problème de reproductibilité se retrouve dans de nombreux domaines de l'IoT, touchant à la science expérimentale.*

### 6.2.2 GDR

Les GDR suivants abordent de façon partielle la gestion de données dans l'IoT :

**GPL** : le GDR Génie Logiciel et Programmation étudie les aspects logiciels de l'IoT, notamment au travers des axes *Génie Logiciel pour les systèmes cyber-physiques*.

**MADICS** s'intéresse aux nouvelles méthodes et outils pour la gestion, l'exploitation et la valorisation des données scientifiques. Les données considérées sont produites en grande quantité (dont la taille est difficilement appréhendable), à un rythme intense, avec des structures variées ou hétérogènes, avec des sémantiques variables et présentant de fait une qualité et précision très peu contrôlables. Les caractéristiques de ces données scientifiques soulèvent de nombreux défis concernant leur traitement qui peut se représenter comme un continuum de tâches non indépendantes, adaptatives et performantes allant de la collecte des données à leur nettoyage, leur intégration, leur analyse et l'interprétation de ces analyses

par des experts. Les Actions mises en place dans MaDICS se focalisent sur une ou plusieurs de ces tâches et adressent un type de données scientifiques. Les réflexions en cours dans ces Actions sur les nouveaux outils algorithmiques et mathématiques (e.g. pour les très grands graphes, l'analyse de flux de données, la reproductibilité via des workflows, etc.) peuvent bénéficier amplement aux données de l'IoT.

### 6.2.3 Laboratoires & Équipes

Nous avons notamment identifié les laboratoires suivants traitant cet aspect des problèmes :

**I3S** : l'équipe SPARKS du laboratoire I3S développe des formalismes de composition logicielle pour le déploiement de comportements intelligents dans les réseaux de capteurs hétérogènes et la synthèse des masses de données collectées dans la hiérarchie du réseau. La contractualisation entre objets de l'IoT est également abordée.

**INRIA Paris** : l'équipe MiMove s'intéresse au middleware pour des systèmes distribués.

**LIG** : l'équipe SLIDE (ScaLable Information Discovery and Exploitation) développe des algorithmes pour analyser de larges volumes de données.

L'équipe HADAS s'intéresse à la gestion de masses de données (stockage, indexation à la volée, composition de flux) et aux systèmes de requête adaptatifs (optimisation par apprentissage, langage de requêtes hybrides, opérateurs sur des données en continu).

**LIRIS** : l'équipe BD (<https://liris.cnrs.fr/bd/>) porte des travaux sur l'intégration de données, les données capteurs dans le cadre des bâtiments et de la ville intelligente ainsi que sur le calcul distribué de requêtes continues et la détection de corrélations dans les flux de données.

**LIS Toulon** : DYNI (Information Dynamics <http://sabiiod.org/dyni>) innove des algorithmes en capture et traitement de masse de données acoustiques (coll. NYU, Cornell Univ...). Elle a co-créé la plateforme SMIoT (suite à JASON, Pôle Information Numérique Prévention UTLN), et centre ses recherches en traitement du signal et apprentissage pour la compression (ondelette, DeepLearning) embarquée et en réseau de masses acoustiques véloces (ultrasoniques). Les applications couvrent le suivi de la faune (terrestre et sous-marine) joint aux pollutions audio-lumineuses, atmosphériques etc.

**UVSQ / INRIA** : l'équipe SMIS (<https://project.inria.fr/smis/>) s'intéresse au problème de traitement de gros volumes de données, en utilisant des objets contraints en termes de capacité.

**LIPADE** : l'équipe DiNo (<http://dino.mi.parisdescartes.fr/>) s'intéresse aux problèmes d'indexation et d'analyse de séries temporelles à large échelle avec des applications diverses dans la maintenance prédictive et l'astronomie.

**LIP6** : l'équipe BD (<http://www-bd.lip6.fr/>) s'intéresse aux problèmes d'intégration de données pour l'interrogation et l'analyse sémantique de données IoT.

## 7 Autres Domaines

L'IoT ne se restreint pas aux 4 domaines sur lesquels nous avons souhaité nous focaliser principalement dans ce document. Nous introduisons donc ci-dessous certains domaines à la frontière pouvant intéresser l'INS2I. Nous rappelons que ce document n'a pas pour but d'être exhaustif, un référencement complet du domaine n'étant pas son objectif.

## 7.1 Ethique

Les discussions lors du séminaire IoT du Conseil Scientifique de l'INS2I ont permis de mettre en exergue des problématiques liées à l'éthique des usages. Nous avons notamment identifié :

**Vie Privée** : les capteurs collectent un volume très important de données, pouvant être exploitées pour en déduire des informations privées (ex : présence de personnes dans un logement pour les compteurs d'eau, identification de comportements habituels). Assurer le service demandé tout en garantissant la vie privée des usagers reste un problème ouvert ;

**Santé** : de plus en plus de capteurs bio-médicaux voient le jour. S'inscrivent donc les problématiques liées à la bio-éthique.

**Automatisation des décisions** : l'IoT peut fonctionner de façon autonome, sans humain intervenant dans la boucle de décision. Il est donc nécessaire d'introduire une formalisation des règles, incluant les critères éthiques si besoin est.

**Proposition 5:** *Le CSI pense qu'une action avec l'INSHS serait pertinente pour développer les aspects inter-disciplinaires liés à l'éthique dans l'IoT et ses usages. Elle pourrait prendre la forme d'une année thématique, ou d'un appel PEPS.*

Les GDR suivants peuvent s'intéresser à cette thématique :

**NoST** s'intéresse aux normes sciences & Techniques. Ce GDR pourrait aider à encadrer juridiquement l'usage de l'IoT.

**Economie & Sociologie** pourrait regarder les problématiques liées à la monétisation des données.

## 7.2 Capteurs matériels

Naturellement, l'Internet des Objets comprend également les dispositifs matériels le composant, chargés de générer ou traiter les données. Les premières applications intégraient des capteurs *basiques* de température, humidité, luminosité haute dynamique, acoustique infra, ultra ou sonique...

Ce thème est à l'interface au CNRS entre INS2I et INSIS. Les GDR suivant sont concernés :

**Soc-SIP** : les trois axes du GDR trouvent des applications dans le domaine de l'IoT (SoC, CPHS, systèmes fiables, etc.).

**MADICS** : l'axe EADM du GDR couvre la capture par capteurs haute dynamique et basse consommation dans le domaine de l'IoT.

Nous avons identifié les défis suivants :

## 7.3 Interagir avec l'IoT

L'IoT requiert de reconcevoir l'interface avec ces machines : aucun écran n'est disponible, les périphériques tels que clavier ou la souris n'existent plus. Ainsi, la sécurité ou la mise en réseau reposent sur une auto-configuration. Cependant, il est également nécessaire de reconcevoir la façon d'interagir avec cette masse d'équipements connectés.

Nous pouvons notamment identifier les défis suivants :

**Reconnaissance vocale** : le domaine a fait des progrès spectaculaires, et devrait continuer à se développer (ex : bâtiments intelligents) ;

**Human Interface Device** : des équipements spécifiques commencent à être étudiés, telles que les lentilles connectées, permettant la réalité augmentée.

**Réalité augmentée** : il s'agit d'ajouter l'information numérique aux images captées.

**Eco-monitoring** : il s'agit de proposer au citizen scientist de placer des objets IoT dans son environnement pour observer la biodiversité qui l'environne.

Les GDR suivants ont été identifiés à l'interface avec l'IoT :

**I3** : le thème 5 "Interaction et coopération" pourrait s'appliquer à l'IoT.

**IG-VR** s'intéresse aux problématiques de la réalité augmentée ;

## 8 Propositions du CSI

Le caractère transverse de l'IoT rend difficile l'identification de problématiques de recherche propres à l'IoT en particulier dans les domaines de la transmission, de la sécurité ou des données. Faudrait-il de grands projets pluridisciplinaires allant des capteurs aux données ?

**Proposition 6:** *Le CSI a identifié un besoin d'augmenter le dialogue et les échanges entre académiques & industriels. Il préconise des réunions mixtes organisées par les GDR considérés.*

Quel que soit le domaine, l'Europe, dans le cadre des ESFRI, demande à ce que les données collectées soient librement diffusées, à des fins de vérification, d'étude, d'exploitation. Dans le cadre de l'IoT, il s'agit donc tant des données collectées sur les plateformes que sur des déploiements réels. L'opendata va plus loin que la reproductibilité scientifique.

**Proposition 7:** *Le CSI demande d'insister sur le problème de l'opendata dans le cadre d'expérimentations pour l'IoT. A chaque publication / projet devrait être adossé a minima un lieu public ou récupérer les données, préalablement anonymisées si la mise à disposition des données pose des questions de vie privée.*

L'Internet des Objets représente un domaine pluridisciplinaire, dont les défis scientifiques sont tirés des contraintes propres au domaine d'application. Le thème demande donc des compétences pratiques afin de pouvoir développer une recherche expérimentale.

## 9 Récapitulatif des préconisations

**Proposition 1:** *Le CNRS devrait étudier l'opportunité de participer activement dans l'IoT Valley. Il existe une opportunité intéressante pour favoriser à travers cette structure les collaborations entre académiques et industriels.*

**Proposition 2:** *Le CSI propose aux GDR d'organiser régulièrement des journées thématiques inter-GDR sur le problème de l'IoT, comme l'ont fait RSD & SoC-SiP en 2016.*

**Proposition 3:** *Le CSI encourage l'émergence de collaborations focalisées sur l'Internet des Objets, au sein du pre-GDR sécurité. Ce domaine recouvre plusieurs des axes du pré-GDR. Ces interactions pourraient être favorisées par exemple par des journées thématiques, la création d'un axe, etc.*

**Proposition 4:** *Le CSI préconise de réfléchir au problème de la reproductibilité des expérimentations. Des scénarios et jeux de données pourraient par exemple être définis afin de servir d'étalon aux solutions, et d'identifier plus précisément les cas d'usage et solutions adaptées. Ce problème de reproductibilité se retrouve dans de nombreux domaines de l'IoT, touchant à la science expérimentale.*

**Proposition 5:** *Le CSI pense qu'une action avec l'INSHS serait pertinente pour développer les aspects inter-disciplinaires liés à l'éthique dans l'IoT et ses usages. Elle pourrait prendre la forme d'une année thématique, ou d'un appel PEPS.*

**Proposition 6:** *Le CSI a identifié un besoin d'augmenter le dialogue et les échanges entre académiques & industriels. Il préconise des réunions mixtes organisées par les GDR considérés.*

**Proposition 7:** *Le CSI demande d'insister sur le problème de l'opendata dans le cadre d'expérimentations pour l'IoT. A chaque publication / projet devrait être adossé a minima un lieu public ou récupérer les données, préalablement anonymisées si la mise à disposition des données pose des questions de vie privée.*

# Synthèse et propositions sur l'éthique en sciences de l'information

Conseil Scientifique de l'institut CNRS INS2I

L'**éthique** d'un domaine scientifique invite à réfléchir aux impacts des travaux de recherche de ce domaine sur la société et à anticiper les problèmes qu'ils peuvent susciter. Elle se distingue de l'**intégrité scientifique**, qui traite des méthodologies et des bonnes pratiques de l'activité de recherche en général. Les deux ont bien sûr des liens.

Dans certaines disciplines, en particulier en biologie et santé, la dimension éthique des recherches est déjà prise en compte depuis longtemps et fait l'objet de lois. Ce n'est pas (encore) le cas en sciences de l'information. Différentes instances ont pourtant été mises en place ces dernières années, avec des objectifs et des périmètres sensiblement différents :

- La CERNA<sup>1</sup>, le comité d'éthique de l'alliance Allistène, existe depuis 2012. Elle produit des rapports et mène des actions de sensibilisation.
- Le Comets<sup>2</sup> est le comité d'éthique du CNRS. Créé en 1994, son périmètre est plus large. Il a émis des avis qui concernent les sciences de l'information (sur les enjeux éthiques du partage des données scientifiques, notamment).
- Le COERLE<sup>3</sup>, lancé par l'INRIA en 2012, est plus axé sur les enjeux légaux ou éthiques des recherches, et se préoccupe de l'intégrité scientifique de ses chercheurs. Il peut interdire ou sanctionner certaines pratiques (pour des raisons de sécurité, par exemple).
- La CNIL ([www.cnil.fr](http://www.cnil.fr)) est également chargée des questions sociétales et éthiques du numérique depuis la loi numérique de 2016.

Des instances internationales comme l'IEEE sont aussi en train de se positionner sur ce thème<sup>4</sup>. Par ailleurs, des workshops sont régulièrement proposés aux chercheurs en marge de certaines grandes conférences du domaine<sup>5</sup> et des revues scientifiques ont consacré des numéros spéciaux à l'éthique de leur domaine<sup>6</sup>.

Ces instances, et d'autres issues d'initiatives diverses, ont donné lieu à des rapports (cf. celui sur « Éthique de la recherche en robotique » produit par la CERNA) ou des chartes (cf. celle sur « [Éthique et big data](#)<sup>7</sup> » soutenue notamment par l'ATALA et l'AFCP et adoptée par Cap Digital). Compte-tenu des progrès et de l'intégration grandissante des technologies de l'information dans les modes de vie, l'INS2I reconnaît qu'il devient de plus en plus nécessaire de prendre en compte la dimension éthique des recherches relevant de son périmètre. Cette prise en compte est mise en avant par exemple dans :

- des ateliers, des conférences, des écoles d'été... spécifiquement dédiés à ces questions ;
- beaucoup d'appels à projets (notamment européens), qui incluent un volet « éthique » ;
- les contrats doctoraux, pour lesquels la CERNA préconise d'intégrer une information sur l'intégrité scientifique et l'éthique<sup>8</sup>.

<sup>1</sup> <http://cerna-ethics-allistene.org/La-CERNA/>

<sup>2</sup> <http://www.cnrs.fr/comets/>

<sup>3</sup> <http://www.inria.fr/institut/organisation/instances/coerle/presentation-du-coerle>

<sup>4</sup> Ethics Advisory Group de l'European Data Protection Supervisor : <https://secure.edps.europa.eu/EDPSWEB/edps/EDPS/Ethics>

<sup>5</sup> cf. <http://ii.tudelft.nl/edia2016/> par exemple

<sup>6</sup> cf. par exemple <https://www.atala.org/-57-2-TAL-ET-ETHIQUE->

<sup>7</sup> <http://wiki.ethique-big-data.org/>

<sup>8</sup> [http://cerna-ethics-allistene.org/digitalAssets/48/48627\\_Proposition\\_formation\\_ethique\\_CERNA.pdf](http://cerna-ethics-allistene.org/digitalAssets/48/48627_Proposition_formation_ethique_CERNA.pdf)

# Propositions

Le CS de l'INS2I souhaite accompagner et promouvoir cette prise de conscience, afin que l'éthique devienne une composante naturelle de la recherche scientifique dans les sciences de l'information. Le CS de l'INS2I propose pour cela diverses mesures qui doivent être vues comme une invitation aux chercheurs à anticiper l'évolution et l'impact de leur domaine, et à les aider à faire entrer plus intimement la réflexion éthique dans le quotidien de leur travail.

## Mesures relevant de la communication :

- Communication à destination des chercheurs : continuer à faire connaître et soutenir les initiatives déjà existantes : liens vers des sites pertinents sur le site Web du CNRS, diffusion d'informations sur cette thématique via le service communication : cette préoccupation est déjà présente dans les publications de l'institut et du CNRS en général, l'effort doit être poursuivi (rubrique pérenne, mise à disposition de documents et vidéos, pages avec liens de référence...). Outre les liens cités ci-dessus, on peut signaler un blog sur l'Éthique et le TAL, ouvert aux débats<sup>9</sup>.
- Communication à destination du grand public : les actions de communication du CNRS à destination du grand public (fête de la science, forum du CNRS...) pourraient aussi systématiquement intégrer une dimension "éthique".

## Mesures relevant de la formation :

Continuer à soutenir ou organiser des "écoles d'été" ou d'autres types de formations sur cette thématique, qui doivent s'adresser à tous, jeunes chercheurs et chercheurs confirmés (formation auprès des DU par exemple). Des sessions de sensibilisation courtes (1 ou 2 jours) pourraient ainsi être proposées au niveau des délégations, notamment à destination des membres des unités appelés à jouer le rôle de "correspondant éthique" (cf. proposition suivante).

## Mesures destinées à diffuser la prise en compte de l'éthique dans la démarche scientifique, à tous les niveaux :

- Un "correspondant éthique" pourrait être identifié dans chaque unité. Ce dernier devrait suivre une formation et aurait pour mission de relayer la réflexion sur l'intégrité scientifique et l'éthique à l'intérieur de son unité. Chaque institut pourrait aussi nommer en son sein un tel correspondant.
- Plutôt que de créer un GDR qui serait spécifiquement dédié à l'éthique (des sciences de l'information ou plus large), il serait souhaitable que chacun des GDR existant intègre une dimension éthique : un "correspondant" chargé de relayer ces questions pourrait être nommé dans chacun des GDR existant.
- Demander systématiquement la présence d'une rubrique "éthique" dans les réponses aux appels à projets issus de l'institut, pour inciter les chercheurs à anticiper les conséquences de leurs recherches. La présence d'une réelle réflexion dans le cadre de cette rubrique ("workpackages" avec financements associés, car cette prise en compte peut être coûteuse dans certains cas) doit jouer en sa faveur lors de son évaluation.

---

<sup>9</sup> <http://www.ethique-et-tal.org/>

# Séminaire thématique « Sécurité en sciences de l'information »

Conseil Scientifique d'Institut de l'INS2I

Fiche de synthèse – journée du 29 Février 2016

Rédigée par le groupe de travail « Sécurité en sciences de l'information » : Guy Gogniat (animateur), Véronique Cortier, Julien Gossa, Isabelle Queinnec

Contributions, concernant les éléments de prospective, du GDR ISIS (avec le concours de William Puech, Cédric Demonceaux et Mari Kobayashi), du GDR SoC-SiP (avec le concours des équipes du Lab-STICC, du LCIS, du LIRMM, du Laboratoire Hubert Curien, du TIMA, de l'IRISA, du LTCI et du département Systèmes et Architectures Sécurisées), du GDR MaDISC (avec le concours de Benjamin Nguyen) et du GDR MACS (avec le concours d'Isabelle Queinnec).

---

## Executive summary

Le domaine de la sécurité en sciences de l'information est un domaine pluridisciplinaire, depuis la physique jusqu'au traitement de l'information. Cette dimension pluridisciplinaire nécessite une organisation de la recherche concertée afin de pouvoir apporter des réponses complémentaires aux enjeux de sécurité. Les questions juridiques et sociologiques sont également essentielles et doivent être prises en compte. La question de l'acceptabilité des solutions technologiques par les utilisateurs ne doit pas être oubliée. Pour appréhender ces différents points, le CSI a organisé le 29 février 2016 une journée de réflexion autour de ce thème. Suite à cette journée, le CSI recommande 8 propositions à l'INS2I :

Proposition 1 : Suivre les actions et le développement du pôle d'excellence cyber (PEC) afin de voir comment ce dernier contribue aux évolutions de ce domaine.

Proposition 2 : Promouvoir les actions visant à rapprocher les acteurs académiques et industriels autour du développement de prototypes afin d'encourager les transferts technologiques.

Proposition 3 : Recenser les plateformes d'évaluation présentes dans les laboratoires de l'INS2I et veiller, lorsque cela est pertinent, à leur mutualisation afin de permettre aux chercheurs et industriels d'avoir accès à ce type d'infrastructure.

Proposition 4 : Encourager davantage les actions du type écoles thématiques et rencontres entre les doctorants et les entreprises.

Proposition 5 : La mise en place du pré-GDR « Sécurité » par l'INS2I est une action marquante qui doit être soutenue, promue et étendue.

Proposition 6 : Mettre en place une programmation pluriannuelle d'appel à projet PEPS dans le domaine de la « sécurité » et faire une cartographie des domaines couverts par les propositions afin d'identifier les éventuels points de faiblesse.

Proposition 7 : Mettre en place un défi dans le domaine de la « sécurité » afin de promouvoir les collaborations interdisciplinaires.

Proposition 8 : Développer un document visant à faire la promotion des chercheurs de l'INS2I dans le domaine de la « sécurité ».

Le CSI propose également des éléments de perspectives basés sur des échanges avec plusieurs GDR traitant de la sécurité en sciences de l'information. Ainsi, les enjeux liés à la protection de la vie privée et des données multimédia y sont présentés. Les questions scientifiques liées à la vidéosurveillance et à la biométrie y sont également abordées. La protection de la couche physique dans le contexte des télécommunications et les défis liés à la sécurité des composants logiciels et matériels sont également présentés.

---

## 1. Introduction

Depuis près de 20 ans, on assiste à un bouleversement considérable des Sciences et Technologies de l'Information et de la Communication (STIC) avec des éléments marquants comme l'accroissement massif des systèmes embarqués et des objets communicants, la domination de l'Internet, la dématérialisation des systèmes/infrastructures de calcul et de communication, la multiplication des *smart devices*, la montée en puissance des systèmes de production de biens manufacturés et de services qui sont de plus en plus connectés<sup>1</sup>. Ces modifications profondes des technologies et de leurs usages ont un impact majeur sur nos sociétés qui sont chaque jour davantage consommatrices de technologie, d'information et de communication. Au-delà des progrès et du confort que ces bouleversements procurent, ils soulèvent de nombreuses questions, notamment dans le domaine de la sécurité en sciences de l'information. En effet, il devient nécessaire d'introduire la sécurité comme pivot du développement de nos technologies et de nos échanges d'information en adressant notamment les questions suivantes : Comment concevoir des systèmes numériques de confiance ? Comment garantir la confiance de leur implantation et de leur contrôle ? Comment anticiper les risques et les vulnérabilités ? Comment définir des modèles et des politiques de sécurité ? Comment avoir l'assurance de la protection des biens et des services ? Ou encore comment garantir la sécurité par conception ?

Ces enjeux touchent tous les domaines de la société du fait du caractère ubiquitaire des technologies de l'information et de la communication. On peut par exemple citer les domaines du transport (automobile, aéronautique, ferroviaire...), de l'énergie (gaz, pétrole, nucléaire...), de la santé (biotechnologies...), des médias (images, vidéo, musique...), de l'industrie (SCADA...), de la distribution de biens (eaux...), des objets connectés (téléphonie, internet...), de la finance (banques...). Un des enjeux est que pour plusieurs de ces domaines il n'existe pas de culture de sécurité et que les menaces évoluent très vite avec des conséquences potentiellement dramatiques du point de vue humain et financier. Il y a donc une forme d'urgence à permettre à l'ensemble des acteurs économiques et industriels de monter en compétence dans le domaine de la sécurité en sciences de l'information afin de leur permettre de répondre aux attentes et aux exigences des marchés et des utilisateurs.

On perçoit bien la complexité sous-jacente en termes de sécurité du fait de l'hétérogénéité des systèmes, des besoins et des utilisateurs. Les enjeux en sécurité sont donc très larges et les réponses à apporter nombreuses et caractérisées par une très forte interdisciplinarité. Parmi ces enjeux scientifiques, on peut citer : la cryptographie, les méthodes formelles et les propriétés de sécurité, les services de sécurité, la détection d'intrusion, la détection d'anomalie, la sécurité du matériel, la sécurité des systèmes, la sécurité des réseaux, la sécurité des données et des systèmes de sauvegardes, la sécurité des applications et des logiciels, la sécurité des systèmes d'exploitation, la sécurité applicative, le développement sécurisé, la virtualisation sécurisée, l'authentification, la traçabilité et la journalisation, la sécurité et la protection de la vie privée des humains et de la société..

L'INS2I est au cœur de ces préoccupations et rassemble de nombreuses équipes de recherche qui font référence aux niveaux national et international. Aussi, afin de mieux appréhender la masse critique et les expertises présentes en France, le CSI de l'INS2I s'était déjà emparé de cette problématique à travers l'organisation d'un séminaire thématique « Sécurité », le 9 Décembre 2013. Ce dernier, animé par Hubert Comon-Lundh (ENS Cachan), avait accueilli les intervenants suivants : Loïc Duflot (Agence Nationale de la Sécurité des Systèmes d'Information), David Pointcheval (CNRS, INRIA, ENS Paris) et Véronique Cortier (LORIA, CNRS, Nancy). Les conclusions du CSI de l'INS2I étaient les suivantes : domaine de recherche en expansion, nécessité d'une aide au mouvement thématique en soutenant en priorité des projets exploratoires, nécessité d'un GDR « Sécurité », nécessité de faire évoluer les formations initiales, et d'adresser la question difficile du transfert industriel dans ce domaine. Comme précisé dans la suite de cette fiche de synthèse, certaines de ces recommandations ont été retenues et mises en place depuis 2015 par l'INS2I.

---

<sup>1</sup> Agence Nationale pour la Recherche Sciences et Technologies de l'Information et de la Communication

Le thème de la sécurité en sciences de l'information soulevant de nombreux enjeux et les évolutions dans ce domaine étant rapides, le CSI de l'INS2I a souhaité presque 2 ans plus tard, le 29 février 2016, s'emparer de nouveau de cette question en organisant une nouvelle journée d'auditions et de réflexion autour de ce thème. Cette journée n'a pas abordé toutes les problématiques, tant le périmètre du domaine de la sécurité en sciences de l'information est large, elle a néanmoins essayé d'éclairer une initiative intéressante au niveau national à travers la mise en place du Pôle d'Excellence Cyber (PEC) et d'adresser des problématiques spécifiques dans le domaine de la virologie et de la preuve formelle. Les présentations et les réflexions ont donc porté sur ces questions à travers les interventions des invités suivants :

Jean-Marc Jézéquel, IRISA « Panorama des recherches dans le domaine de la cybersécurité notamment dans le cadre du Pôle d'Excellence Cyber »

Jean-Yves Marion, LORIA « Enjeux et avancées dans le domaine de la virologie »

David Monniaux, VERIMAG « Enjeux et avancées dans le domaine de la vérification formelle - application à la sécurité »

## 2. Bilan des réflexions et propositions d'actions

Le domaine de la sécurité en sciences de l'information est par essence un domaine pluridisciplinaire, depuis la physique jusqu'au traitement de l'information. En effet, la sécurité d'un système se mesure à travers son maillon le plus faible. Il est donc essentiel d'appréhender la question de la sécurité à travers la mise en place d'une chaîne de confiance (sécurité de bout en bout). Cette dimension pluridisciplinaire nécessite une organisation de la recherche concertée afin de pouvoir apporter des réponses complémentaires aux enjeux de sécurité. Les questions juridiques et sociologiques sont également essentielles et doivent être prises en compte. La question de l'acceptabilité des solutions technologiques par les utilisateurs doit également être appréhendée. La question de la formation représente également un enjeu et cela depuis les études secondaires afin de permettre aux utilisateurs d'être davantage éclairés sur ces questions jusqu'aux études supérieures afin de former les futurs experts du domaine.

**Proposition 1 : Le pôle d'excellence cyber (PEC), qui adresse les dimensions recherche, formation, transfert technologique, représente une initiative intéressante qui s'inscrit dans cette démarche. Il rassemble de nombreux acteurs académiques et industriels. Le CSI recommande à l'INS2I de suivre ses actions et son développement afin de voir comment ce dernier contribue aux évolutions de ce domaine.**

La complexité des systèmes est telle qu'il est important de disposer d'une masse critique suffisante afin de pouvoir explorer en profondeur les réponses scientifiques et technologiques aux enjeux de sécurité. Cette masse critique est indispensable afin de passer du concept théorique à une validation expérimentale sur des systèmes représentatifs de systèmes réels. En effet, il est essentiel de tester l'ensemble de la chaîne de protection depuis l'utilisateur jusqu'au système et cela au sein d'environnement confinés (notion de bac à sable) afin de contrôler les expérimentations. De telles approches de validation nécessitent des ressources humaines expertes qui permettent de maintenir et de développer les plateformes d'évaluation. La question du soutien en ressources humaines est un point important pour aboutir à des prototypes plus avancés et probablement plus attractifs pour les industriels. Le développement de démonstrateurs représentatifs des systèmes réels est donc un enjeu. Certains industriels français sont des acteurs économiques majeurs du domaine, aussi il est important d'établir de façon plus systématique des passerelles entre les chercheurs et les industriels. En effet, la question du transfert industriel reste complexe dans ce domaine et ce point doit être approfondi afin de créer une plus forte synergie entre recherches exploratoires et pré-compétitives.

**Proposition 2 : Le CSI recommande à l'INS2I de promouvoir des actions visant à rapprocher les acteurs académiques et industriels autour du développement de prototypes afin d'encourager les transferts technologiques.**

**Proposition 3 : Le CSI recommande à l'INS2I de recenser les plateformes d'évaluation présentes dans ses laboratoires et veiller, lorsque cela est pertinent, à leur mutualisation afin de permettre aux chercheurs et industriels d'avoir accès à ce type d'infrastructure.**

Les initiatives récentes telles que l'école d'été « Cyber in Bretagne<sup>2</sup> » organisée conjointement par le pré-GDR « Sécurité » (présenté ci-après) et le Pôle d'Excellence Cyber (PEC) en juillet 2016 ainsi que la rencontres entreprises doctorants sécurité (REDOCS<sup>3</sup>) qui se tiendra en octobre 2016 où des doctorants des domaines de la sécurité informatique s'associeront sous forme de groupes afin de répondre à des problèmes posés par des entreprises sont à saluer et doivent être davantage encouragés.

**Proposition 4 : Le CSI recommande à l'INS2I d'encourager davantage d'actions du type écoles thématiques et rencontres entre les doctorants et les entreprises.**

La communauté adressant les questions de sécurité en sciences de l'information est relativement large et aujourd'hui présente dans différents GDR ce qui ne facilite pas toujours leurs interactions. On peut notamment citer les GDR suivants :

- GDR Informatique-Mathématique (groupe de travail Codage et Cryptographie) ;
- GDR Information, Signal, Image et ViSion, (thème Télécommunications : compression, protection, transmission, axe compression et protection) ;
- GDR System on Chip – System in Package (axe Confiance Matérielle, groupe thématique sécurité numérique) ;
- GDR Réseaux et Systèmes Distribués ;
- GDR Génie de la Programmation et du Logiciel (groupe Formalismes et Outils pour la Vérification et la Validation) ;
- GDR Modélisation Analyse et Conduite de Systèmes dynamiques ;
- GDR Masses de Données, Informations et Connaissances en Sciences.

La mise en place du pré-GDR « Sécurité<sup>4</sup> » par l'INS2I est donc une initiative très importante qui devrait permettre à terme de décloisonner la communauté « sécurité » française. En effet, comme indiqué dans la feuille de route du pré-GDR « Sécurité », ce dernier doit permettre de rassembler une communauté scientifique unifiée autour de tous les aspects de la sécurité informatique, tels que la cryptologie et le codage, la vérification, les questions de vie privée, l'étude des vulnérabilités et des mécanismes de protection, la sécurité matérielle, etc. Ce GDR doit permettre de mieux connaître les acteurs académiques et industriels et rendre plus visible leurs complémentarités afin d'aboutir à une cartographie des expertises au sein de la communauté française. Il doit aussi être un outil de veille et de promotion afin de rendre plus visible à l'international l'expertise française dans ce domaine.

**Proposition 5 : La mise en place du pré-GDR « Sécurité » par l'INS2I est une action marquante qui doit être soutenue, connue et étendue.**

Afin de favoriser les interactions entre les chercheurs de l'INS2I, la mise en place d'appel à projet sur le thème de la « sécurité » correspond à un outil d'animation et d'accompagnement scientifique important. L'appel PEPS INS2I 2016 « Sécurité Informatique et des Systèmes Cyber-physiques » est une initiative qui va dans ce sens. Il serait également pertinent d'évaluer les domaines couverts par les propositions reçues afin de voir si certains domaines spécifiques, qui sont jugés stratégiques, n'en sont pas absents. Une programmation pluriannuelle de projets sur ce thème devrait être encouragée.

<sup>2</sup> <https://project.inria.fr/cyberinbretagne/fr/>

<sup>3</sup> <http://confiance-numerique.clermont-universite.fr/redocs2016/>

<sup>4</sup> <http://gdr-securite.irisa.fr/>

**Proposition 6 : Le CSI recommande à l'INS2I de mettre en place une programmation pluriannuelle d'appel à projet PEPS dans le domaine de la « sécurité » et de faire une cartographie des domaines couverts par les propositions afin d'identifier les éventuels points de faiblesse.**

La sécurité est une approche fortement interdisciplinaire, aussi il est essentiel d'associer la mission interdisciplinarité du CNRS aux actions en cours à travers par exemple la mise en place d'un défi dans ce domaine afin de mettre l'humain et la société au cœur des actions de recherche et de fédérer les différentes expertises au-delà des sciences de l'information.

**Proposition 7 : Le CSI recommande à l'INS2I de mettre en place un défi dans le domaine de la « sécurité » afin de promouvoir les collaborations interdisciplinaires.**

Afin de renforcer la visibilité des chercheurs dans le domaine de la « sécurité » il serait intéressant de recenser les chercheurs présents dans les laboratoires de l'INS2I ayant obtenu un ERC ou étant membre de l'IUF.

**Proposition 8 : Le CSI recommande à l'INS2I de développer un document visant à faire la promotion de ses chercheurs dans le domaine de la « sécurité ».**

### 3. Éléments de prospective

Sans avoir la prétention d'être exhaustive, cette dernière partie de la fiche de synthèse met en évidence plusieurs thématiques de recherche présentant des verrous scientifiques à lever.

La protection de la vie privée (PVP) est un thème central à de nombreuses discussions sur la société de nos jours, comme avec la récente loi sur le renseignement en France, les négociations sur le TAFTA (partenariat transatlantique de commerce et d'investissement), le projet de loi pour une République numérique, le règlement européen sur les données personnelles, etc. En effet, dans le monde connecté où nous vivons désormais, des données personnelles toujours plus nombreuses sont générées, souvent à l'insu des personnes concernées. Le traitement de ces données massives conduit à un profilage important des individus.

De nombreux problèmes éthiques, philosophiques et bien entendu juridiques se posent sur le concept de la protection de la vie privée. Il est donc nécessaire d'étudier (i) comment ces problèmes se traduisent en termes de problèmes théoriques fondamentaux, (ii) s'ils peuvent donner lieu à des réalisations informatiques (logiciels de type « Privacy Enhancing Technology » ou PET), (iii) les interactions avec les autres disciplines comme le droit et la sociologie sur les problèmes de PVP. Ce travail doit être mené avec les GDR des disciplines de droit et de sociologie (par exemple le GDR Normes Sciences et Techniques).

Ces problèmes incluent par exemple le contrôle de l'individu sur ces données, à la fois dans leur partage, leur exploitation, ou leur vente, son anonymat, au cours des nombreux traitements que subissent ses données, ou lors de son interaction avec des services numériques, la transparence dans le traitement, l'acquisition ou le transfert des données, la responsabilité (ou « accountability »), ou encore l'équité de la relation entre l'individu et les fournisseurs de services.

En ce qui concerne les traitements, il s'agit bien souvent de techniques d'analyse de données. On parle alors de « Privacy-Preserving Publishing » ou de « Privacy-Preserving Data Mining » (PPDM) lorsqu'on souhaite étudier la prise en compte de la PVP dans ce domaine.

Il est important de renforcer les actions de recherche sur la protection de la vie privée au sein des laboratoires de recherche de l'INS2I en partenariat avec les disciplines de droit et de sociologie.

Les pistes explorées par les équipes travaillant sur la PVP en France sont assez diverses. Elles incluent des approches cryptographiques ou des approches formelles (logique, graphes, preuve) pour proposer des protocoles ou des architectures permettant de garantir des propriétés de PVP, des techniques

d'interrogation de données protégeant la VP, des techniques de calcul sécurisées (cryptographie homomorphe, réseau de confiance ou matériel sécurisé comme présenté dans la suite).

La protection des données multimédia ; notamment comment sécuriser des données visuelles pendant leur transmission, leur archivage et leur visualisation est également une problématique importante. En effet, d'après CISCO, les données visuelles (images, vidéos, objets et scènes 3D) représenteront 80% du trafic Internet mondial en 2019. Avec le tout numérique, il devient de plus en plus facile de copier des données, de les visualiser sans droit, de se les approprier mais aussi de les falsifier. La sécurité des données visuelles peut être abordée par des aspects insertion de données cachées (tatouage, stéganographie, ...) ou par des aspects cryptographie (chiffrement sélectif, signature perceptuelle, ...). Cela concerne principalement la confidentialité, l'authentification et l'intégrité. Tous ces aspects de sécurisation doivent bien sûr respecter les standards internationaux de compression tels que JPEG ou H264. Il faut aussi adapter les techniques aux nouveaux formats multimédia (HdR, multi-composantes, stéréo, ...). De nouvelles questions voient également le jour comme l'identification et l'authentification d'une image afin de pouvoir être utilisée comme preuve juridique recevable.

Un autre versant consiste à détecter si un message est caché dans une image (stéganalyse) ou si le chiffrement est suffisant (cryptanalyse). Il s'agit de la « sécurité à partir de données visuelles ». Dans ce domaine sont abordées également les problématiques de traçabilité de données visuelles et de détection de falsification tels que des copier-déplacer dans une image ou des copier-coller d'une image vers une autre. Dernièrement il est apparu des aspects phylogénie des images qui consistent à partir d'un ensemble d'images très similaires (récupérées sur des réseaux sociaux par exemple) de retrouver qui est le parent de qui afin de remonter à l'image source. Il existe aussi des méthodes d'identification d'appareil photo numérique à partir d'analyse de bruit et de distorsion contenus dans une image.

La vidéosurveillance et la biométrie correspondent également à des problématiques importantes. Il s'agit d'identifier des personnes pour des accès sécurisés ou analyser des comportements afin de détecter des comportements malveillants. Il s'agit ici de « sécurité des personnes », et non de sécurité des données. La vidéosurveillance amène aux problèmes de reconnaissance de personnes, reconnaissance de gestes, analyse de foules. La reconnaissance de visage basée vision est un sujet très étudié, considéré comme mature et pour lequel on annonce désormais un taux de reconnaissance supérieur à 99%<sup>5</sup>. En revanche, l'analyse de geste et de comportement reste un problème ouvert même si l'arrivée des caméras 3D a permis d'améliorer fortement les résultats. Cette problématique rencontre un grand succès aux niveaux national et international.

A l'échelle plus macroscopique, l'analyse de foule est également un sujet d'étude qui rassemble un nombre important d'acteurs. Dans le cas d'une foule peu dense, les méthodes de détection et de suivi de personnes ont atteint des performances très satisfaisantes même en présence de fortes occultations. Par contre, l'analyse de foule dense reste un problème difficile à résoudre et encore ouvert.

La sécurité de la couche physique est une autre des grandes thématiques actuelles en théorie de l'information. Le principe de la sécurité pour la couche physique est d'exploiter le bruit présent naturellement dans tous les canaux de transmission pour garantir la sécurité inconditionnelle des communications. L'idée repose sur le fait que les signaux interceptés par des écouteurs indésirables subissent généralement une dégradation différente de ceux reçus par les récepteurs légitimes (par exemple, une atténuation différente qui induit un rapport signal-à-bruit différent). Afin d'adresser cette problématique il est nécessaire d'étudier la capacité maximale d'un canal sous contrainte de sécurité. En effet, la théorie de l'information donne une limite fondamentale (théorique) de la capacité maximale d'un système de transmission. L'ajout de l'aspect sécurité (garantir que l'on ne peut rien obtenir comme information à partir d'une séquence d'observation sur un canal) rajoute des contraintes sur la transmission. Sur ces aspects sécurité de la couche physique, la France a des bonnes productions sur

---

<sup>5</sup> Spreuwers, L.J. (2015). "Breaking the 99% barrier: optimisation of 3D face recognition". IET Biometrics 4 (3): 169–177

les aspects théoriques, moins sur les aspects pratiques de mise en œuvre sur lesquels beaucoup d'actions doivent encore être menées.

La composante matérielle de la sécurité en sciences de l'information est passée depuis quelques années au premier plan en raison du développement et de l'utilisation intensive des systèmes électroniques embarqués et aujourd'hui du développement de la cybersécurité, de l'internet des objets et des objets connectés. Outre les attaques par analyse de fuites apparues dans les années 90, de nouvelles menaces sur les systèmes électroniques sont apparues récemment du fait d'une filière de conception et de fabrication globalisées au niveau mondial. Les risques associés comprennent l'insertion de matériel malveillant (chevaux de Troie matériels, portes dérobées...), la contrefaçon de circuits intégrés et le vol d'IP (Intellectual Property). D'autre part, les attaques sur le matériel cryptographique évoluent aussi en efficacité, par exemple en couplant l'analyse sur canaux cachés à l'injection de fautes (notamment sur le canal électromagnétique), ou en visant plusieurs points simultanément lors des attaques en fautes. De nombreux verrous scientifiques et technologiques doivent encore être levés dans les prochaines années pour relever le défi de la sécurité matérielle de l'information.

Si les algorithmes standards de chiffrement à clé symétrique ou à clé publique sont toujours au cœur des efforts de recherche et de développement, de nouveaux schémas de chiffrement sont de plus en plus étudiés. C'est notamment le cas du chiffrement authentifié, du chiffrement post-quantique et du chiffrement homomorphe. Ce dernier type de chiffrement est particulièrement intéressant, cependant les réalisations actuellement disponibles ne sont pas utilisables en pratique. Des réalisations efficaces doivent être proposées rapidement par la communauté pour augmenter l'acceptabilité de ces nouveaux schémas. C'est pourquoi, les liens entre les algorithmes, les représentations des données, les architectures matérielles et logicielles et la résistance aux attaques physiques seront aux cœurs des travaux scientifiques en sécurité dans les années à venir.

La complexité et l'hétérogénéité des Systems-On-Chip (SoC) grandissants, des méthodes efficaces pour leur conception et pour l'évaluation de la sécurité des moyens de protection logiciels et matériels doivent être étudiées. Pour cela, il faut disposer d'outils de conception conjointe logicielle-matérielle et de synthèse de haut niveau prenant en compte la sécurité comme une contrainte de conception, ces outils n'existent pas encore. Durant les phases de conception, des outils de vérification holistique formelle de la sécurité portant à la fois sur le logiciel et le matériel sont aussi nécessaires. Le développement de la sécurité par conception reste un objectif fort dans les années à venir avec de nouveaux enjeux comme par exemple la gestion des droits de propriété des IP et des circuits, la surveillance comportementale des systèmes, la résilience des systèmes matériels sensibles, la garantie des fonctions de test, de diagnostic et de débogage sans préjudice sur la sécurité et la proposition de mécanismes sophistiqués d'authentification et d'identification. Le développement de chaînes de confiance allant du système au composant est un enjeu fort de la sécurité aujourd'hui. La vérification formelle (multi-niveaux et multi-technologies) de l'ensemble de la chaîne de confiance est à développer dans un futur proche (notion de « sécurité prouvée »).

Pour de nombreuses applications il existe un besoin fort en réalisations matérielles légères et même ultralégères d'algorithmes de chiffrement devant apporter des garanties élevées de sécurité vis-à-vis des attaques physiques. Il faut aussi pouvoir coupler des systèmes cryptographiques sûrs et légers avec des éléments de traçabilité et d'identification intrinsèque des composants matériels. Les fonctions physiques non clonables qui bénéficient aujourd'hui d'une attention très forte de la communauté, ne sont qu'une réponse partiellement satisfaisante. Il convient donc d'étudier précisément leur modélisation afin d'apporter les garanties de sécurité inhérentes au processus de certification de la sécurité.

Enfin, de façon connexe à la sécurité en sciences de l'information, il est également nécessaire d'élaborer des méthodes pour assurer un fonctionnement robuste et stable du système dynamique global lors de catastrophes naturelles, de pannes accidentelles (pannes d'équipement, erreur humaine, bug de logiciels), ou à la suite d'attaques malveillantes. Ceci se fait au travers de certificats (garanties a priori)

de bon fonctionnement, de mécanismes de détection de défauts et de diagnostic de pannes, de gestion des risques et de reconfiguration dynamique. Ces dimensions doivent également être appréhendées afin de garantir la disponibilité des systèmes quelles que soient les conditions de fonctionnement.

On comprend à travers ces enjeux la complexité et les interactions entre les problématiques liées à la sécurité. Il est donc essentiel de permettre aux différentes communautés scientifiques de se rencontrer pour développer des solutions allant de la théorie de l'information jusqu'à la mise en œuvre de solution logicielle et matérielle tout en intégrant la problématique des usages et de l'acceptabilité.

## Conseil scientifique de l'Institut des sciences d'information et de leurs interactions (INS2I)

### RECOMMANDATION

#### Egalité hommes-femmes à l'INS2I

L'INS2I compte 23% de femmes parmi ses chercheurs et 46% parmi ses personnels IT, la moyenne au CNRS étant, respectivement, de 33% et 51%<sup>1</sup>, et l'évolution n'est pas encourageante. C'est donc un institut peu féminisé, à la fois comparé à l'ensemble du CNRS et bien sûr, comparé à la population française. Pourtant, la société a largement évolué en matière de travail féminin, les femmes représentant aujourd'hui 48% de la population active et 55% de la fonction publique d'Etat<sup>2</sup>.

Le Conseil Scientifique de l'INS2I (CSI) recommande aux unités relevant des sciences de l'information et de leurs interactions d'évoluer pour présenter l'attractivité d'une communauté équilibrée.

Pour cela, le CSI invite les directeurs des laboratoires et l'ensemble des personnels relevant des sections 6 et 7 à s'engager dans une politique active en faveur de l'égalité. Le CSI recommande aux laboratoires de mettre en place des initiatives en ce sens. Afin d'alimenter la réflexion, le CSI propose une série d'initiatives qui peuvent être déclinées localement. Enfin, le CSI souhaite que les politiques d'égalité fassent systématiquement l'objet d'une discussion entre les unités et les tutelles lors des comités de suivi.

#### 1) Promouvoir les femmes dans les réseaux scientifiques

- favoriser la participation des doctorantes et jeunes chercheuses à la vie de leur communauté scientifique, et à des conférences comme ACM Women Encourage<sup>3</sup> ;
- promouvoir les femmes aux postes à responsabilités dans les laboratoires, les GdR et les GdT ;
- inviter les femmes aux postes de direction à participer aux réseaux professionnels pertinents pour leurs fonctions, par exemple l'association des femmes dirigeantes de l'enseignement supérieur, de la recherche et de l'innovation (Afdesri).

#### 2) Respecter la parité

- s'assurer que les projets présentés par les laboratoires ne sont pas déséquilibrés ;
- si c'était le cas, inviter les chercheuses à participer aux projets en cours de montage ou financés ;

<sup>1</sup> Source : bilan social du CNRS, 2014 [http://bilansocial.dsi.cnrs.fr/BSP-2014\\_web.pdf](http://bilansocial.dsi.cnrs.fr/BSP-2014_web.pdf)

<sup>2</sup> Source : INSEE 2013 [http://www.insee.fr/fr/themes/document.asp?ref\\_id=irsocmartra13#s2](http://www.insee.fr/fr/themes/document.asp?ref_id=irsocmartra13#s2)

<sup>3</sup> <http://womencourage.acm.org/>

## RECOMMANDATION

### Egalité hommes-femmes à l'INS2I

- lors d'une manifestation scientifique, s'assurer que les comités sont paritaires dans la mesure du possible en fonction du domaine scientifique ;
  - doter les unités de plus de 100 permanents d'un observatoire de la parité, en concertation avec leurs autres tutelles (le cas échéant, en s'appuyant sur les référents égalité qui peuvent être nommés dans les Universités ou organismes partenaires).
- 3) Promouvoir les figures féminines au-delà de l'INS2I**
- favoriser la participation des femmes à des actions visibles dans les médias, par exemple en s'inscrivant au guide "Les Expertes"<sup>4</sup> ;
  - inviter les membres des unités relevant des sciences de l'information et de leurs interactions à des actions en faveur de l'égalité en amont du monde scientifique, par exemple sur le modèle des bourses d'études "Informatique au féminin"<sup>5</sup> ou en participant à des associations comme "Femmes et Sciences"<sup>6</sup> ;
  - communiquer sur les disciplines de l'INS2I auprès des collégiennes et lycéennes, ou des étudiantes en premier cycle universitaire.

En conclusion, la recommandation du CSI est d'atteindre la parité, soit 40% au moins de chaque sexe, chez les nouveaux entrants pour les différentes catégories de personnels, à l'horizon de dix ans.

**M. Christian BARILLOT**  
Président du CS INS2I

**Recommandation adoptée le 29 février 2016**  
**Voté à l'unanimité /19 votants**

#### Destinataires :

- M. Alain FUCHS, Président du CNRS
- M. Michel BIDOIT, Directeur de l'INS2I
- M. Philippe BUTTGEN, Président de la CPCN.
- M. Bruno CHAUDRET, Président du Conseil scientifique
- Mme Anne PEYROCHE, Directrice générale déléguée à la science
- M. Christophe COUDROY, Directeur général délégué aux ressources
- Les unités des sciences de l'information

---

<sup>4</sup> <http://expertes.eu/>

<sup>5</sup> <http://femmes.fil.univ-lille1.fr/>

<sup>6</sup> <http://www.femmesetsciences.fr>

# Plateformes d'envergure en Sciences de l'Information

Conseil Scientifique d'Institut de l'INS2I

Fiche de synthèse – 24 novembre 2015, v1

Rédigée par le groupe de travail Plateformes :

*Michel Beaudouin-Lafon (animateur), François Chaumette, Julien Gossa, Inbar Fijalkow, Anne-Catherine Letournel, Serge Torres, Hélène Touzet*

## 1. Introduction

Les recherches en sciences de l'information nécessitent de plus en plus l'utilisation de plateformes matérielles et logicielles spécialisées. Cela est dû à plusieurs facteurs, parmi lesquels :

- Une approche expérimentale de plus en plus souvent nécessaire pour valider des résultats de recherche (analyse de performance par exemple), mais aussi pour conduire l'ensemble du processus de recherche (interaction humain-machine par exemple) ;
- Le passage à l'échelle imposé par l'évolution des enjeux scientifiques dans de nombreux domaines, qui nécessite des infrastructures dédiées (HPC, big data, etc.) ;
- L'importance croissante des recherches qui combinent le monde physique et le monde numérique (robotique, internet des objets, etc.).

En conséquence, un nombre conséquent de *plateformes d'envergure*, c'est-à-dire qui dépassent le cadre d'une équipe ou d'un laboratoire, ont vu le jour depuis quelques années dans les unités de l'Institut, dont certaines dans le cadre des appels EQUIPEX du PIA.

Le CSI de l'INS2I a conduit le 1<sup>er</sup> juin 2015 une journée d'auditions et de réflexion sur le thème des plateformes. Afin de focaliser la réflexion, cette journée n'a pas porté sur les plateformes purement logicielles, et n'a que peu abordé les plateformes de calcul ou de stockage, dont l'importance pour la recherche en informatique est désormais bien établie. Les présentations et les réflexions ont donc porté surtout sur les plateformes ayant une forte composante matérielle et d'envergure régionale ou nationale, voire internationale. Ont été auditionnés :

- Jean-François Gibrat, directeur de l'IFB, pour l'infrastructure nationale de bioinformatique
- Eric Fleury, professeur à l'ENS Lyon, pour la plateforme FIT (Internet des objets)
- Michel de Mathelin, directeur de iCube, pour l'Equipex ROBOTEX
- Jean-Marc Vézien, ingénieur au LIMSI, pour l'Equipex DIGISCOPE
- Vivien Quema, professeur au LIG, pour la plateforme d'usages DOMUS

## 2. Bilan des réflexions

Les auditions ont montré la diversité des plateformes, qu'il s'agisse des types d'équipements déployés, de leur portée géographique, des publics visés, ou des moyens, notamment humains, de support. Elles ont montré également un certain nombre de points communs.

### 2.1 Finalités des plateformes

D'une part ces plateformes sont très souvent utilisées à la fois comme *outils* de recherche et comme *objets* de recherche. En tant qu'outil, une plateforme fournit une infrastructure permettant aux chercheurs de réaliser leurs travaux, au même titre qu'un

microscope pour un biologiste. En tant qu'objet de recherche, une plateforme est conçue et développée dans le cadre d'un projet de recherche pour permettre certains types de travaux et apporter des réponses à des questions de recherche, mais aussi pour poser de nouvelles questions, selon une démarche exploratoire. Ces deux finalités ne sont pas exclusives, et peuvent évoluer au cours de la vie d'une plateforme : un outil peut devenir un objet de recherche en l'adaptant ou en le modifiant, et un objet de recherche peut se stabiliser et, même s'il devient obsolète pour cette finalité, devenir un outil utile pour d'autres chercheurs. Il est donc important de considérer une plateforme non pas seulement comme une ressource matérielle avec des objectifs et des utilisateurs bien identifiés, mais comme une construction socio-technique de ses différents acteurs amenée à évoluer dans le temps (certains diraient un « écosystème »).

Dans certains domaines, les plateformes peuvent ainsi contribuer significativement à l'animation et la structuration des communautés en favorisant les rencontres et les échanges autour d'outils partagés. Ce facteur va de pair avec un changement dans les rapports qu'entretiennent les plateformes et leurs usagers. On passe d'une relation client-fournisseur à une co-construction entre les gestionnaires et les utilisateurs, d'une définition statique à une vision dynamique de l'organisation et des prestations offertes.

Par ailleurs, un certain nombre de plateformes se trouvent à la limite interdisciplinaire entre les sciences de l'information et d'autres disciplines. La question se pose alors de l'intérêt d'un tel équipement pour la *recherche* en sciences de l'information. L'intérêt peut venir de la possibilité d'observer l'utilisation réelle d'une technologie et des données ainsi recueillies, qui peuvent alimenter des recherches futures. Il peut aussi venir d'un travail de co-conception de la plateforme, bénéfique aux différents partenaires. Il semble que nous soyons ici dans les questionnements traditionnels de l'interdisciplinarité, qui ne peuvent être tranchés que par une analyse fine de chaque situation.

D'un autre côté, les plateformes en sciences de l'information peuvent jouer un rôle de ressource et d'expertise au service de la construction des outils numériques dont se dotent les différents secteurs de la communauté scientifique—voir par exemple le rôle de Grid 5000 dans le développement de l'infrastructure pour la bioinformatique. Il s'agit alors d'une forme de valorisation.

Enfin, les plateformes sont également des vecteurs de diffusion de la connaissance et des supports à la formation. En informatique, où l'invisibilité du logiciel et des algorithmes qui le sous-tendent rend souvent difficile la communication sur nos pratiques et résultats de recherche, une plateforme matérielle fournit un point d'ancrage commode pour illustrer une problématique ou un résultat. Mais cela conduit aussi parfois les plateformes à devenir des « usines à démos », avec un coût (en temps) conséquent pour les chercheurs. Pour la formation, les plateformes peuvent être un élément d'attractivité pour les étudiants et, là aussi, un élément tangible des savoirs et des connaissances transmises. Mais le partage de ces ressources entre recherche et enseignement peut aussi être source de difficultés.

## *2.2 Organisation et gestion des plateformes*

Les plateformes considérées sont très souvent organisées en réseau et sont souvent ouvertes à des utilisateurs au-delà des partenaires de la plateforme. L'organisation en réseau, à une échelle géographique plus ou moins grande, apparaît comme une nécessité afin de justifier les investissements par un plus grand nombre d'utilisateurs, de mutualiser les moyens, et d'être compétitif au niveau international. L'organisation en réseau rend aussi une plateforme plus robuste. L'ouverture répond à la fois à un souci de visibilité et de partage, et parfois à une demande des bailleurs de fonds. Elle est aussi

un moyen de développer les transferts technologiques, en donnant accès à des démonstrateurs technologiques à des acteurs industriels. Cependant l'ouverture et la mise en réseau sont sources de contraintes et de problèmes : disponibilité des équipements pour les différents utilisateurs, robustesse de l'équipement pour des utilisateurs extérieurs, mais aussi les questions de propriété intellectuelle, de confidentialité et de facturation des prestations, notamment dans le cadre d'utilisateurs privés.

Un aspect critique de l'exploitation des plateformes, soulevé par tous les intervenants, est la quotité et la pérennité des moyens humains affectés à l'équipement. Des équipements pointus, mis à la disposition d'utilisateurs variés, nécessitent absolument des personnels qualifiés, dédiés à l'exploitation de la plateforme, sa maintenance et son évolution. De plus, à côté de la gestion technique, les tâches de relation avec la communauté (formation, communication, accompagnement de projets, intégration en continu des nouveaux besoins) prennent une place croissante, décisive pour la pertinence d'une plateforme dans le temps.

Sans support ingénieur pérenne, l'obsolescence d'une plateforme est *très* rapide, la rendant caduque. Le recours à des personnels temporaires de même que la contribution de doctorants ou post-doctorants à des tâches qui ne relèvent pas de la recherche conduisent à des difficultés d'exploitation : sous-utilisation des équipements, manque de fiabilité, perte de savoir-faire, etc. En particulier, la capitalisation des connaissances est vitale pour qu'une plateforme existe dans la durée, or cette capitalisation ne peut pas se faire si elle repose seulement sur des personnels temporaires. Enfin, les plateformes en réseau posent des problèmes supplémentaires de dispersion géographique des compétences, qu'il n'est pas raisonnable de résoudre par la seule mobilité des personnels.

### **3. Propositions d'actions de l'Institut**

Il est important que l'Institut reconnaisse que les sciences de l'information ont une composante expérimentale de plus en plus grande, que l'approche expérimentale est critique dans un certain nombre de domaines si l'on veut rester à la pointe de la recherche internationale, et que des moyens substantiels doivent donc être affectés aux moyens expérimentaux. Il est également important de reconnaître que les contraintes et objectifs qui président à la construction des plateformes conduisent à des configurations extrêmement diverses, ce qui rend difficile la définition d'un modèle unique pour celles-ci. Des actions générales peuvent cependant être entreprises par l'Institut.

#### **3.1 Cartographie**

Une première étape devrait être la réalisation d'une cartographie des plateformes existantes, de leurs moyens et de leurs usages. Pour cela, une définition du périmètre considéré est nécessaire, notamment sur la taille des plateformes, ainsi qu'une description claire des buts visés par cette cartographie. L'un de ces buts devrait être de diffuser cette cartographie pour encourager des actions de mutualisation : mise en place de réseaux, partage de compétences, etc. Le but principal est l'aide au pilotage des moyens affectés aux plateformes (voir ci-dessous pour les moyens humains) et du cycle de vie des plateformes (création, utilisation selon diverses finalités, fin d'exploitation).

Une proposition de fiche est jointe à ce document pour établir cette cartographie. Elle contient d'une part des éléments factuels de description de chaque plateforme et des moyens qui lui sont affectés, d'autre part un résumé de ses objectifs scientifiques et techniques, enfin un ensemble d'indicateurs quantitatifs et qualitatifs.

### 3.2 Bonnes pratiques

Une autre action devrait viser d'une part l'aide à la valorisation des plateformes et de leurs résultats, et d'autre part l'aide à l'ouverture vers des utilisateurs extérieurs, notamment du monde industriel. Cette aide pourrait porter sur des questions de gouvernance, de management et d'évolution technique, des questions juridiques (PI, confidentialité), et des questions financières (location des équipements, accords de partenariat). Sur ce dernier point, l'évaluation, validée par l'administration, du coût complet d'une plateforme permettrait par exemple sa facturation, et la détermination d'un tarif de location.

Par ailleurs, les plateformes semblent un bon point d'entrée pour développer et mettre en œuvre les approches d'Open Science en traitant suffisamment tôt les aspects de collecte, de partage et de diffusion des données, en prenant en compte les aspects juridiques.

En pratique, il s'agirait de constituer un *guide des bonnes pratiques* couvrant notamment les aspects suivants :

- Gouvernance ;
- Moyens humains, mise en réseau des compétences ;
- Mode de fonctionnement, gestion des utilisateurs ;
- Veille technologique, évolution ;
- Evaluation du coût complet, facturation, valorisation ;
- Contribution à l'Open Science.

### 3.3 Moyens humains

L'action la plus attendue est certainement en direction des moyens humains. Il est indispensable que chaque plateforme d'envergure dispose de moyens pérennes sous forme d'au moins un ingénieur permanent affecté par l'un des organismes partenaire.

Compte tenu de l'extrême tension sur ces postes, une politique de recrutement ou d'affectation d'emplois existants devrait être affichée. La cartographie évoquée ci-dessus devrait permettre d'identifier les besoins les plus importants. L'intégration de ces personnels dans les laboratoires doit également être étudiée avec soin car les situations sont très variables. Idéalement, les personnels devraient être affectés à la plateforme et aux équipes qui les animent, pour une période donnée, et ils devraient également être rattachés à un pool technique au niveau du laboratoire ou du site afin d'assurer le bon déroulement de leur carrière.

Au-delà de cette organisation locale, l'Institut devrait encourager la constitution de réseaux de compétences permettant de mettre en relation des personnels sur l'ensemble du territoire et de mutualiser ainsi leurs connaissances et leurs expériences. Ces réseaux pourraient organiser des journées de formation et d'échanges, des visites croisées, maintenir une base documentaire et de veille technologique, etc.

Par ailleurs des actions plus ciblées, telles que des financements de CDDs, sont également bienvenues, en évitant le saupoudrage. Mais elles ne peuvent constituer la seule réponse aux enjeux des plateformes pour la recherche en sciences de l'information.

### 3.4 Chargé de mission

L'ensemble de ces actions pourrait être coordonné par un chargé de mission de l'Institut.

# Proposition de questionnaire Plateforme

## Identification

Nom (acronyme)

Titre (une dizaine de mots)

Site web

“Mission statement” (une phrase)

Partenaires (institutions et laboratoires)

Personne contact (nom et adresse mail)

Localisation(s) des équipements (si multi-site, indiquer s’il y a un site central ou coordinateur)

Etat d’avancement (en construction, opérationnel en interne, opérationnel avec utilisateurs extérieurs, ...), date de mise en exploitation

Ouverture à des extérieurs académiques et/ou industriels

## Objectifs

(texte d’une dizaine de lignes décrivant les objectifs scientifiques de la plateforme et les utilisateurs/usages visés)

## Moyens

Description succincte des moyens matériels et logiciels disponibles

Description des moyens humains affectés à la plateforme

Mode de gouvernance

Existence d’un label ou d’une démarche qualité

Evaluation du financement initial pour l’investissement matériel

Evaluation du financement de l’exploitation et de la maintenance annuelles

Description de l’éventuelle offre de service rémunérée

## Indicateurs

Evaluation du taux d'utilisation du ou des équipements (séparer utilisateurs faisant partie des partenaires et utilisateurs extérieurs). Selon la plateforme ce taux peut être un nombre d'utilisateurs, de connexions, de jobs, etc.

Nombre et origine des utilisateurs (un utilisateur = une équipe, une entreprise, etc.)

Nombre et origine des projets (locaux, régionaux, nationaux, internationaux)

Nombre de publications liées à l'utilisation de la plateforme et/ou publications marquantes

Participation à des actions de formation

Résultats marquants / Défis scientifiques atteints ou avancées remarquables

Exemples de projets conduits sur/avec la plateforme (objectif, partenaires, financements, résultats ...)

Autres indicateurs jugés pertinents



# Prospective report of the scientific council of CNRS-INS2I

## English Summary

The production of a scientific prospective report corresponds to one of the missions of the Institutes Scientific Councils (CSI) of CNRS. Such a report is both for internal use by the CNRS (the head of the institution, institutes, national review committees) and for the scientific community at large. In this context the Scientific Council of the INS2I Institute of CNRS has decided to direct this prospective report around the positioning of information sciences and their interactions in the global scientific landscape. In this way, we wish to highlight the importance and role of our discipline in a period when our field is undergoing deep technological, societal and scientific evolutions.

To bring this reflection to life, we have chosen to rely, among other things, on hearings of leading personalities from our domains. These hearings, conducted in small group sessions, gave rise to extremely rich discussions that fed the Board's thinking in preparing this report. They were conducted on the basis of a common framework structured around five themes:

1. How to define our discipline? What are the major achievements?
2. What are the possible evolutions in the next 10 to 20 years?
3. What are the biggest challenges and risks?
4. How will our discipline impact on other sciences? How to manage thematic mobility in research at the interfaces with other disciplines?
5. How will our discipline impact society at large, particularly in terms of education and popular instruction?

Finally, the forward-thinking carried out in this work is also informed by the work of the Board over the past four years, in particular with regard to the organization of thematic seminars. These seminars allowed us to reflect more specifically on about ten transversal or disciplinary themes. Many of the issues discussed during these seminars were also discussed through the preparatory stages for the preparation of this report, in particular around the place of women in our disciplines or even around ethical issues. The summaries of these seminars are given in the appendix to this document and the working documents are available for free access on the Council's website (<https://csins2i.irisa.fr/>).