



ALLISTENE

**l'alliance des sciences
et technologies du numérique**

Apport des sciences du numérique dans les sciences participatives

ALLISTENE - GT "Sciences Participatives"

19 Novembre 2015

Préambule

Sollicitée par François Houllier dans le cadre de la mission sur les Sciences Participatives¹ pour apprécier la place de ces sciences dans le champ de l'Alliance, ALLISTENE a mobilisé un groupe d'experts, animé par Jean-Gabriel Ganascia et Mokrane Bouzeghoub pour préciser les apports mutuels entre Sciences et Technologies du Numérique et Sciences Participatives. Le groupe d'experts composé de Sihem Amer-Yahia, Mokrane Bouzeghoub, Jean-Gabriel Ganascia et Jean-Marc Ogier a produit ce rapport.

¹ <http://www.sciences-participatives.com>

Groupe de travail sur l'apport des sciences du numérique dans les sciences participatives

Contributeurs: Sihem AMER-YAHIA, Mokrane BOUZEGHOUB, Jean-Gabriel GANASCIA, Jean-Marc OGIER

Recommandations d'actions

A l'issue des travaux engagés à l'initiative d'Allistène sur l'apport des sciences du numérique dans les sciences participatives, notre groupe de travail tient d'abord à insister sur le caractère mutuel de cet apport.

L'apport du numérique apparaît évident, tant les sciences participatives bénéficient de l'utilisation des réseaux informatiques (calcul et stockage) que ce soit pour collecter et agréger les contributions individuelles ou pour distribuer le travail et l'organiser sur un mode coopératif (allocations de tâches, synchronisation, validation, ...). Réciproquement, les sciences du numérique recourent aux techniques participatives pour l'entraînement de logiciels d'apprentissage machine, pour le recueil des retours d'usage dans la conception centrée sur l'utilisateur et pour la résolution de problèmes distribués (identification de compétences, décomposition de tâches, modèles de gratification, ...).

Afin d'améliorer le parti que les sciences du numérique peuvent tirer des techniques participatives et de leur permettre, en retour, de contribuer au perfectionnement des outils et des méthodes de recueil participatif, notre groupe fait quatre propositions :

1. Aider au développement de **plateformes pérennes de « crowdsourcing »** et de logiciels de traitement adaptés afin de permettre aux chercheurs d'avoir la maîtrise des différents composants, tout en évitant deux écueils, la centralisation excessive et la dispersion des efforts.
2. Mettre en place des « **maisons des sciences participatives** » bien identifiées sur le territoire national. Ces maisons abriteront des plateformes technologiques avec des **équipes supports d'ingénieurs** qui assureront la pérennité de plateformes de « crowdsourcing », leur maintenance, l'accès libre et la fourniture de ressources logicielles pour l'acquisition collaborative de données et leur traitement.
3. Financer des **axes de recherche prioritaires dans les sciences du numérique** qui porteront, entre autre, sur la modélisation des compétences des participants, la décomposition et l'allocation des tâches, la construction et la validation des connaissances, la conception d'interfaces, la mise au point de modèles économiques et le traitement de grandes masses de données issues d'une acquisition distribuée.
4. Encourager à la faveur de projets communs **l'émergence de compétences interdisciplinaires** entre les sciences du numérique et certains secteurs des sciences de l'Homme et de la société, en particulier la sociologie des usages, la psychologie sociale, la psychologie cognitive et l'économie.

Table des matières

Recommandations d'actions	1
Table des matières	2
1 Objectifs et plan	3
2 Impact sociétal et questions éthiques.....	4
2.1 Apports	4
2.2 Risques liés aux sciences participatives.....	5
3 Apport des sciences participatives aux sciences du numérique.....	6
3.1 La collecte de données.....	6
3.2 Résolution de problèmes.....	7
4 Problématiques scientifiques dans les sciences du numérique	9
4.1 Déploiement des tâches	9
4.2 Motivation des participants.....	10
4.3 Validation et détection d'anomalies	11
4.4 Modèles et algorithmes	11
4.5 Plateformes	12
5 Recommandation d'actions	13
5.1 Infrastructure d'accompagnement de la recherche.....	13
5.1.1 Plateformes technologiques.....	13
5.1.2 Centre de ressources et support en ingénierie	14
5.2 Axes de recherche prioritaires	14
5.3 Emergence de compétences interdisciplinaires.....	15
5.4 Résumé des recommandations	15
6 Références	16

Sciences Participatives

Groupe de travail sur l'apport des sciences du numérique dans les sciences participatives

Contributeurs: Sihem AMER-YAHIA, Mokrane BOUZEGHOUB, Jean-Gabriel GANASCIA, Jean-Marc OGIER

1 Objectifs et plan

En premier lieu, rappelons que ce groupe de travail s'est constitué à la demande de la présidente d'Allistène, Brigitte PLATEAU, qui répondait à une sollicitation de François HOULLIER, président de l'INRA qui, lui-même, faisait suite à une requête conjointe de la Ministre de l'Education nationale, de l'Enseignement supérieur et de la Recherche et de la Secrétaire d'Etat chargée de l'Enseignement supérieur et de la Recherche. Les deux Ministres ont exprimé le souhait de recevoir un guide de bonnes pratiques sur les sciences participatives d'ici le 30 novembre 2015¹. Notons que leur requête porte sur **l'apport des citoyens aux progrès scientifiques par les données ou les connaissances qu'ils aident à collecter**, et non sur la diffusion de la culture scientifique dans la population, même si l'implication des citoyens dans le recueil de données ou de connaissances y contribue certainement. Quant aux questions liées à la participation des citoyens à la définition des orientations scientifiques ou aux bénéfices qu'ils tirent de la recherche scientifique, elles ne relèvent pas non plus de cette étude.

Emanant d'Allistène, ce groupe de travail doit porter naturellement sur **l'implication des sciences du numérique dans les sciences participatives**. A l'issue de nos travaux, il nous apparaît qu'en cette matière le rôle des sciences du numérique est double :

- ◆ D'un côté, les sciences du numérique bénéficient des sciences participatives lorsque celles-ci aident à construire des bases de données, des annotations ou des ensembles d'apprentissage et qu'elles permettent de résoudre collectivement des problèmes.
- ◆ D'un autre côté, le numérique donne un essor aux sciences participatives en facilitant la collecte et l'échange à travers les réseaux informatiques. Dès lors, les sciences du numérique se trouvent tout naturellement convoquées pour aider à systématiser et à améliorer ces pratiques communautaires.

¹ La lettre de mission envoyée le 19 février 2015 était signée de Madame Najat Vallaud-BELKACEM, Ministre de l'Education nationale, de l'enseignement supérieur et de la Recherche, et de Madame Geneviève FIORASO, Secrétaire d'Etat chargée de l'Enseignement supérieur et de la Recherche.

A cela, il convient d'ajouter un volet sur l'impact social des sciences participatives, que ce soit pour les contributeurs qui risqueraient de subir une forme d'exploitation, comme c'est parfois le cas avec certaines applications du « crowdsourcing » (ou de ce que certains appellent la *myriadisation*² en français), ou pour certains chercheurs qui y voient une science molle dénuée de fondements méthodologiques ou une science dont le contrôle échapperait à leurs prérogatives.

Ces remarques préliminaires nous conduisent à adopter un plan en quatre parties : (i) impact sociétal des sciences participatives et questions éthiques liées à la collecte des données, (ii) apport des sciences participatives aux sciences du numérique, (iii) problématiques scientifiques posées par les sciences participatives aux chercheurs dans les sciences du numérique, et enfin, (iv) recommandation d'actions.

2 Impact sociétal et questions éthiques

2.1 Apports

En préalable, soulignons que les sciences participatives contribuent à l'avancée des connaissances, du moins c'est l'une des motivations de leurs promoteurs, et qu'elles ont une vertu éducative. Elles forment ceux qui s'y adonnent à la méthode et à la rigueur scientifique. Ce faisant, elle aident à mieux faire comprendre les enjeux et les procédures de la science. On peut dès lors espérer qu'elles fassent naître ou confirment des vocations de scientifiques chez des jeunes filles et des jeunes gens tout en diminuant les méfiances à l'égard de la science qui se font trop souvent jour dans la population. Elles se présentent donc comme étant bénéfiques pour tous.

A cette motivation générale s'ajoute un intérêt pratique dans les sciences du numérique où beaucoup d'applications recourent, dès à présent, aux méthodes participatives pour l'acquisition de données. Mentionnons, à titre d'illustration, quatre applications déjà largement répandues :

1. L'annotation distribuée d'images ou de vidéos pour faciliter la mise en œuvre de techniques d'apprentissage supervisé à des fins de caractérisation de scènes.
2. L'annotation de gazouillis (*Tweets*), en vue, là encore, d'entraîner par apprentissage machine, des logiciels de catégorisation ou d'extraction de contenu.

² Le terme de « crowdsourcing », littéralement les ressources de la foule, a été forgé sur le modèle d'« outsourcing » (externalisation en français) pour désigner une forme de délégation du travail au grand nombre. Cela a conduit un certain nombre de linguistes du CNRS à proposer le terme de *myriadisation* en français. Quoique le mot nous apparaisse heureux, nous ne l'utiliserons pas ici pour ne pas dérouter le lecteur du rapport global dans lequel ce texte devrait être inséré ultérieurement.

3. Le recueil des retours d'utilisateurs à l'identification des CAPTCHA³ pour améliorer la reconnaissance automatique de caractères manuscrits, là encore par apprentissage machine.
4. L'obtention des retours d'usage pour des interfaces ou des logiciels dans le contexte d'une conception dite « centrée sur l'utilisateur ».

2.2 Risques liés aux sciences participatives

Cependant, indépendamment de leurs vertus putatives, les sciences participatives suscitent un certain nombre de craintes, plus ou moins justifiées. Tant la perception que la réalité des risques dépend, bien évidemment, des domaines d'application et de la nature active ou passive, rémunérée ou bénévole, du recueil d'information. Enumérons les, pour essayer d'en évaluer la dangerosité.

Il existe parfois une crainte de voir, avec les sciences participatives, des scientifiques du dimanche remplacer, à moindre coût, des chercheurs en poste que l'on pourrait alors débaucher par mesure d'économie. Rappelons, s'il en est besoin, que cela n'a aucun fondement. En effet, même s'ils ne se réduisent pas à être de simples « capteurs de données », les contributeurs bénévoles ne font pas le même travail que les scientifiques de métier.

En revanche, avec la collecte massive de données, deux questions d'ordre éthique demandent à être prises très au sérieux.

L'une porte sur la **protection de la vie privée**. Avec le recueil passif de données de santé ou d'usage au moyen d'objets connectés, des capteurs font irruption dans l'intimité individuelle et permettent désormais de suivre à la trace les individus. Si l'utilisation de ces données personnelles à des fins de recherche peut se justifier lorsque les personnes donnent leur accord, leur vente ou leur diffusion serait très problématique. Il convient donc de rappeler le caractère exceptionnel des données personnelles collationnées et traitées à des fins de recherche, tout en exigeant des chercheurs qu'ils prennent conscience de leurs obligations morales de discrétion et qu'ils s'y soumettent.

Ajoutons à cela que, l'utilisation des données personnelles à des fins scientifiques peut requérir une **anonymisation**, c'est-à-dire un effacement du lien entre données individuelles et personnes. Empêcher de rétablir ce lien et, le cas échéant, caractériser le degré d'anonymisation, c'est-à-dire la difficulté à reconstruire ce lien, posent des problèmes intéressants et délicats aux chercheurs. En effet, il ne suffit pas de supprimer les

³ Destinées à distinguer les hommes des robots virtuels (bots) sur des tâches simples de reconnaissance de caractères, les CAPTCHAS sont largement utilisés sur Internet pour prémunir les sites contre des attaques massives.

noms, car on peut bien souvent retrouver, par croisement de bases de données, des informations individuelles, sans compter que cela fait parfois perdre leur sens aux données. Les critères et les techniques d'anonymisation font donc l'objet de nombreuses recherches dans les sciences du numériques.

La seconde est relative au **recours à un travail distribué** et faiblement rémunéré, sans contrat ni assurances sociales, qui peut conduire à des formes d'exploitation difficilement admissibles dans nos pays^[1]. Malheureusement, la mondialisation des réseaux fait qu'il n'est pas aisé d'introduire des législations nationales dans ces activités^[2]. On franchit parfois le pas qui fait passer des « sciences participatives » bénévoles à un « travail distribué » réparti et rétribué sur mode tayloriste. On ne peut alors empêcher que les méthodes développées initialement pour accroître les connaissances scientifiques n'aient, comme finalité, qu'un accroissement des gains.

3 Apport des sciences participatives aux sciences du numérique

Les approches participatives se développent dans beaucoup de domaines scientifiques. Dans le contexte du numérique, elles contribuent essentiellement à deux problématiques majeures : la collecte de données et la résolution de problèmes.

3.1 La collecte de données

C'est l'une des contributions importante des sciences participatives, au point où on pourrait en réduire leur portée à ce seul point. Cette collecte peut prendre diverses formes : *relevés et transmission d'observations de terrain* (photos de plantes, audios de champs d'oiseaux, traces d'animaux, ...), *annotation interactive de corpus* (étiquetage ou tagging d'images, reconnaissances de caractères, identification de personnages dans une scène, ...), *retours explicites ou implicites d'usages* (évaluation qualitative de scènes ou d'objets, fréquence d'accès à une ressource, durée d'affichage d'une page, ...).

Le processus de collecte peut être actif ou passif et engendre de ce fait des types de manipulation et des traitements différents. Dans le premier cas, le participant est à l'initiative de l'action de collecte (prise de photo, mesure, étiquetage, appréciation). C'est le cas des amateurs en astronomie^[3], ou des clubs de randonnées intéressés par les espèces de champignons. On peut recourir aussi à une participation passive de contributeurs dès lors qu'ils sont porteurs d'équipements susceptibles d'acquérir automatiquement des données. C'est le cas de smartphones, des montres connectés, des tablettes et ordinateurs portables, voire même des implants de capteurs qui enregistrent et envoient des flux d'information. Il existe une autre forme d'acquisition passive qui est en relation avec

l'analyse du comportement (nombre de clics sur Internet, durée entre deux clics, durée d'affiche d'une page, vitesse de frappe sur un clavier, les sites fréquemment visités,...).

A titre d'illustration, nous citons deux exemples d'application aux sciences du numérique de cette forme de collecte des données

- *Annotation de corpus pour l'apprentissage et l'évaluation* : Aujourd'hui, on a de plus en plus souvent besoin de grands corpus annotés pour mettre en œuvre des techniques d'**apprentissage supervisé** dans tous les domaines de l'informatique, par exemple pour la reconnaissance de motifs, d'objets, ou de scènes sur les images ou pour l'indexation de textes avec la reconnaissance d'entités nommées et de relations, ou encore pour l'interprétation de textes courts comme les « gazouillis » (*Tweets*) ou de textes anciens. Identiquement, dans beaucoup de domaines où ne préexiste pas de mesure formelle, comme là où il est question de sémantique, l'évaluation de la **performance d'algorithmes** recourt aussi à des corpus annotés.
- *La conception d'interfaces centrée sur l'utilisateur* fait appel depuis longtemps à des retours d'usage. On conçoit dès lors l'apport des techniques participatives pour recueillir ces retours et les agréger en vue de valider par exemple une interface utilisateur. Plus généralement, les sciences participatives peuvent servir d'outil pour concevoir ou valider des modèles d'interaction, présentés par exemple comme des "jeux sérieux" (*serious games*), dans différents domaines où l'analyse des usages est un enjeu dans la création de la valeur ou l'acceptation d'un système ou d'un objet.

Dans tous les cas, la contribution essentielle des sciences participatives à ce processus de collecte d'informations réside dans leur capacité à mobiliser des foules autour d'un thème, à établir les règles de gouvernance des données collectées (y compris le droit de propriété), à définir un modèle économique respectueux des lois et de l'éthique dans le cas de la rétribution des participants. Il est à noter aussi que les sciences participatives ont constitué un levier à l'amélioration et la généralisation des plateformes d'acquisition.

3.2 Résolution de problèmes

L'usage des sciences participatives dans la résolution de problèmes peut paraître comme la partie "la plus noble" de ce nouveau courant, pour peu que la méthodologie utilisée soit bien formalisée et reconnue comme outil scientifique par les pairs. L'enjeu est donc double : (i) formuler un problème de telle sorte que sa résolution soit distribuée entre une myriade d'acteurs, (ii) inscrire cette résolution dans une méthodologie scientifiquement validée.

La question de la formulation du problème est essentielle. Il s'agit d'expliciter le but à atteindre, d'esquisser le processus avec ses tâches élémentaires, ses étapes et ses points de synchronisation, d'établir les profils des participants, de définir le modèle économique ou les modalités de mobilisation des participants, d'élaborer la procédure de recrutement de ces derniers et de lancer la campagne de recrutement. Ce travail constituera un cahier des charges préalable où l'élicitation de tâches dépend fortement du domaine d'application. Les sciences du numériques prendront ensuite le relai pour traduire ce cahier des charges en termes technologiques, c'est-à-dire en mise en oeuvre de plateformes et en une implémentation du procédé esquissé précédemment.

La définition et l'implémentation d'un tel procédé n'ont de valeur que si la communauté scientifique du domaine correspondant les reconnaît comme des pratiques scientifiques valables et conformes aux usages du domaine. Un regard particulier sera porté sur la crédibilité de l'approche, sa tolérance aux fautes (en particulier les comportements potentiellement malveillants des participants), les biais qu'elle peut induire dans les tâches élémentaires ou dans l'agrégation de ces tâches, la mesure de la qualité des résultats intermédiaires ou finaux, l'interprétabilité des résultats, la reproductibilité du processus, etc.

Bien que ces deux points soient connus dans les sciences du numérique, en particulier dans la conception des logiciels (analyse, décomposition, développement coopératif, intégration, test, validation, ...), ils n'ont jamais été à ce point « externalisés » vers des acteurs non spécialistes du domaine, et n'ont jamais connu une implication aussi nombreuse d'acteurs (milliers, centaines de milliers, voire millions de participants). D'où l'importance que prennent les sciences participatives en articulation avec les sciences du numérique et les ouvertures que cette articulation offre pour adresser des problèmes nouveaux ou jamais résolus par défaut de méthode ou en raison de leur complexité.

A titre d'illustration, citons un exemple de plateforme et un exemple de projet :

- Les plateformes *InnoCentive*^[5] et *IdeaConnection*^[6]. En l'occurrence, on mentionne sur le site de la plateforme *IdeaConnection* la "success story" d'une société Canadienne qui serait parvenue à découvrir une mine d'or^[7] en partageant ses données géologiques et en les faisant prospector de façon distribuée par des mathématiciens, des géologues, des militaires, des étudiants etc.
- Le projet *Graph Isomorphism Game*^[8] de la plateforme *crowdcrafting*^[9] propose de résoudre des instances du problème d'isomorphisme de graphes (GIP - Graph Isomorphism Problem) de façon distribuée. On peut aussi consulter sur le site "next stop design"^[10] la référence à une expérience de conception distribuée qui a porté sur les abris bus. Plus généralement, il existe des tâches d'optimisation dans des

domaines spécifiques comme la manipulation de gros graphes où le coup d'œil humain peut aider considérablement dans la recherche d'une solution à des problèmes de croisement, d'appariement ou d'identification de sous-graphes ayant certaines propriétés.

Les sciences participatives constituent donc pour le numérique une nouvelle approche d'analyse et de résolution de problèmes, originales dans leur capacité à formuler un problème en un ensemble de tâches distribuées, et à mobiliser les ressources humaines nécessaires pour réaliser ces tâches.

4 Problématiques scientifiques dans les sciences du numérique

La mise en œuvre de campagnes de collecte massive d'information auprès d'amateurs bénéficie grandement de l'apport des technologies du numérique. Cela tient à la fois au recueil en ligne qui tire parti des facilités offertes par le web et au traitement des masses de données (ce que l'on appelle aussi les *Big Data*) ainsi recueillies qui fait appel à des techniques d'apprentissage machine et de fouille de données.

Au delà de la mise en œuvre de techniques connues, cela ouvre sur des questions nouvelles auxquelles les sciences du numérique peuvent et doivent contribuer, soit seules, soit en collaboration avec différentes disciplines, en particulier avec les mathématiques, l'économie, la sociologie des usages et la psychologie sociale. Déjà, des ateliers de travail, des colloques ou des sessions spéciales dans de grandes conférences scientifiques abordent ces points. Cette section en énumère un certain nombre.

4.1 Déploiement des tâches

Une première question tient à la **décomposition des tâches** en sous-tâches qui doivent être acceptables par les participants. Quel degré de difficulté est admissible pour obtenir les meilleurs résultats ? Quelle doit être la nature et la durée de ces tâches ?

Une seconde question porte sur l'**assignation des tâches** aux utilisateurs : comment faut-il procéder ? Doit-on donner une seule tâche à chaque contributeur ou plusieurs ? Ainsi, si l'on annote des bandes dessinées, faut-il demander à la fois de reconnaître le contenu textuel des bulles et des phylactères, le nom de protagonistes et les lieux, ou spécialiser les contributeurs sur une seule de ces tâches ? De façon symétrique, doit-on assigner la même tâche à plusieurs pour confronter les résultats ? A cet égard, on peut aussi essayer de **faire coopérer** les gens. Cela pose des questions relatives à la taille des groupes et à la complémentarité des compétences, ce qui relève plus de la sociologie des organisations que des sciences du numérique. Indépendamment de ces questions d'ordre

sociologique, cela pose des questions intéressantes sur la **modélisation de processus coopératifs** avec des agents aux compétences variées et sur la **synchronisation** éventuelle des différentes tâches.

Enfin, comment repérer, parmi les participants, ceux qui sont les plus à même de réaliser une tâche ? Pour cela, il convient certainement de les **profiler**. Mais, comment construire ces profils et comment les **mettre en adéquation avec les tâches** et enfin comment aider les participants à choisir la tâche qui leur convient le mieux ?

4.2 Motivation des participants

Pour recruter des participants potentiels, il convient, d'abord, de les solliciter ce qui tient généralement à la plateforme et à son audience. On pourrait toutefois imaginer dans des opérations de science participative passer par des communautés savantes, des institutions éducatives ou des organismes de recherches pour attirer les participants potentiels.

Outre cette sollicitation, il faut les **motiver**, ce qui tient soit à l'intérêt intellectuel de la tâche, soit à son caractère ludique^[16], soit encore à la stimulation produite par l'envie de contribuer à l'avancée des connaissances ou à la soif d'apprendre. C'est le cas dans FoldIt^[17], une plateforme sur laquelle les participants sont invités à participer à un jeu de repliement de protéines. Cela peut aussi être un scénario pédagogique : des classes de collégiens contribuent, à la demande de leur professeur, à réaliser des tâches d'annotation ; le contexte scolaire induit alors une forme de motivation.

Cette motivation prend aussi parfois la forme d'une **rétribution financière**, ce qui pose tout naturellement la question du montant de cette rétribution. En effet, non seulement, cela a un certain coût mais, de plus, l'expérience montre qu'une indemnité importante attire une grande masse de personnes mues uniquement par l'appât du gain. Il en résulte un fort taux de réponses erronées, parce que hâtives. On assiste donc à des effets paradoxaux qui conduisent à obtenir de meilleures réponses avec une rémunération plus faible. A cela, il faut ajouter que des expériences ont prouvé que, plus que la rétribution totale, la comparaison des rétributions conduit à certaines frustrations^[18].

Notons que les recherches touchant à la motivation des participants débordent largement la compétence des seuls informaticiens : les *sciences de l'homme et de la société* sont invitées à examiner les dimensions psychologiques, sociologiques et économiques. Les "cobayes humains" ou les cohortes utilisés en médecine depuis des lustres constituent certainement une source d'information à exploiter et analyser pour en déterminer les motivations.

4.3 Validation et détection d'anomalies

Une fois les tâches exécutées, il convient de **valider les réponses**, d'évaluer leur qualité et de **détecter les anomalies systématiques** ainsi que les **biais**. Pour déterminer une réponse acceptable, on peut soit recourir à un vote majoritaire, soit prendre en considération le profil de compétence des utilisateurs. Ce profil lui-même peut être appris automatiquement en fonction de la qualité des réponses apportées antérieurement. Au surplus, il convient d'éliminer le bruit inhérent à une collecte distribuée. La détermination des modèles de validation des réponses, le « nettoyage » des données acquises ainsi que la construction des profils de compétences apparaissent comme autant de sujets d'investigation intéressants pour des équipes de recherche dans les sciences du numérique.

A cela, ajoutons que les sites de sciences participatives ne sont pas à l'abri d'actions partissanes menées par des groupes de militants voulant, pour des raisons diverses, infléchir les résultats d'une étude. Là encore, la détection de ces actions individuelles ou collectives devrait faire l'objet de recherche.

4.4 Modèles et algorithmes

Certaines questions liées à la modélisation des participants et des tâches ainsi qu'aux algorithmes d'assignation de tâches aux participants et aux algorithmes d'apprentissage de profils des participants, ont déjà été abordées par des équipes de recherche. Un grand nombre de ces questions reste ouvert^[11]. La modélisation des participants nécessite de prendre en compte **des facteurs humains** tels que l'expertise des participants, leur disponibilité, la nature de leur motivation ou leur ratio de participation à des tâches antérieures. Ces facteurs doivent intervenir dans la conception d'algorithmes d'allocation de tâches aux participants et de validation des réponses des participants aux tâches.

De plus, comme une grande majorité des plateformes de crowdsourcing se basent sur le volontariat, et sur la capacité des participants à s'auto-assigner des tâches, elles bénéficieraient de travaux antérieurs dans le domaine de la **recommandation en ligne**. Cela permettrait d'**orienter les participants** et de les **aider à trouver les tâches** qui les intéressent le plus, tout en gérant au mieux le flux des tâches à traiter.

L'**apprentissage des compétences des participants** soulève de nouveaux problèmes d'ordre algorithmique. Aujourd'hui, pour établir les compétences de participants potentiels, la majorité des plateformes recourent à des tests de pré-qualification, sous forme de questionnaires remplis par les utilisateurs eux-mêmes préalablement à l'exécution des tâches. Ces questionnaires sont coûteux, fastidieux et doivent être adaptés aux besoins de chaque tâche. Une approche basée sur l'apprentissage

automatique des qualifications des participants à partir de tâches exécutées auparavant, serait plus robuste.

Alors que certains modèles et algorithmes pour le crowdsourcing ont été testés sur des simulateurs, il convient aussi, pour certaines évaluations, de les **déployer avec des participants humains**. La disponibilité de tels environnements est essentielle pour (1) le passage à l'échelle des algorithmes, et (2) leur confrontation aux facteurs humains ^[12]. Cela permet de valider les hypothèses scientifiques sur les motivations et les pratiques collectives. Par exemple, l'hypothèse selon laquelle les tâches collaboratives (ex., rédaction d'un texte à plusieurs) sont mieux effectuées par des groupes de moins de trois membres (masse critique) peut être ainsi validée. Également, le déploiement de tâches avec des participants humains permet d'aborder des questions de motivation ou les questions de remplacement incrémental de participants au cas où certains seraient indisponibles au moment du déploiement ce qui n'est évidemment pas possible dans un environnement simulé.

4.5 Plateformes

Dans chaque opération de science participative en ligne, il faut construire des sites pour mettre en œuvre le recueil informatique de données et de connaissances. Ce travail de programmation assez lourd et coûteux est grandement facilité par l'**utilisation de plateformes génériques de « crowdsourcing »**. Cependant, beaucoup reste à faire pour alléger le fardeau de ceux qui souhaitent se lancer dans cette entreprise, en particulier pour les aider à décomposer les tâches, à valider les réponses, à motiver les participants, etc. Aujourd'hui, la plateforme la plus utilisée semble être *Amazon Mechanical Turk*^[13] (AMT) du fait du nombre de participants potentiels. Cependant, celle-ci semble plus dédiée à des opérations commerciales qu'aux sciences participatives, car elle repose d'abord sur une rétribution financière des exécutants. Il en existe d'autres comme *Crowd4U*^[14] et, en logiciel libre, *PyBossa*^[15] mais le recrutement y est moins aisé puisqu'il n'est pas rémunéré et qu'il repose sur des milieux plus restreints que la clientèle Amazon tel que le milieu académique pour *Crowd4U* et les cercles d'amis pour *PyBossa*. À cela, il faut ajouter que les plateformes commerciales ne donnent pas la maîtrise de tous les composants. Les chercheurs éprouvent donc des difficultés à modifier un module spécifique ou à intégrer une fonctionnalité qu'ils désirent tester. Et, quand bien même ils seraient parvenus à ajouter un module original, ils n'arrivent pas à le mettre à disposition de la communauté d'utilisateurs de la plateforme et à faire connaître leur travail.

En outre, on pourrait essayer de rendre les attributions de tâches et les rémunérations transparentes sur les plateformes, de sorte que les citoyens contribuant à une activité scientifique aient une idée du but ultime de la recherche à laquelle ils contribuent, qu'ils sachent quelle est leur place dans l'ensemble du dispositif et qu'ils

connaissent leurs partenaires et éventuellement échangent avec eux. Cependant, cette vision ouverte peut être discutable si elle “rompt la neutralité” du processus scientifique et conduit à des “alliances d’activistes” souhaitant orienter la méthode ou les résultats scientifiques dans une direction qui leur est favorable.

5 Recommandation d’actions

Il existe, dans les pays développés, des initiatives et des financements importants pour des projets de recherche ayant trait aux dimensions numériques des sciences participatives. C’est en particulier le cas aux Etats-Unis avec le lancement d’appels d’offre de la NSF depuis plusieurs années, par exemple en 2014, sur les *Social-Computational systems*^[19]. De même, il y eu quelques initiatives européennes sur ces problématiques ; notons, par exemple, le projet Social-IST^[20] sur l’intelligence sociale collective (*Social Collective Intelligence*)^[21].

A cela, on peut ajouter, aux Etats-Unis, des laboratoires, comme le AMP Lab^[22] de l’université de Berkeley, dont l’activité contribue en grande partie au développement de techniques d’interprétation des données recueillies par les sciences participatives^[23]. Si l’on souhaite donner, dans notre pays, un essor aux sciences participatives, il apparaît indispensable de lancer et financer des **programmes de recherche** sur leur versant numérique, tant au niveau théorique qu’à celui, plus appliqué, des technologies mises en œuvre.

Compte tenu de l’état d’avancement des recherches et du domaine sur lequel ils portent, ces programmes doivent se répartir sur trois volets complémentaires : infrastructure d’accompagnement de la recherche, axes de recherche spécifiques et émergence de compétences interdisciplinaires.

5.1 Infrastructure d’accompagnement de la recherche

5.1.1 Plateformes technologiques

Nous avons vu que la recherche en ce domaine passait par un déploiement en ligne avec des participants humains. A cette fin, il faut recourir à des plateformes technologiques pour ne pas recommencer, à chaque fois, le travail de programmation. Or, outre qu’elles ne donnent pas pleinement satisfaction, les plateformes disponibles en libre accès aujourd’hui ne procurent pas aux utilisateurs la maîtrise sur tous leurs composants, ce qui constitue un verrou pour des chercheurs qui souhaiteraient procéder à des expérimentations spécifiques. Il faudrait donc **promouvoir des plateformes technologique de crowdsourcing** ouvertes à toute la communauté scientifique. Sur ce point, il apparaît important d’**éviter la fragmentation** des efforts et la dispersion des ressources en

mobilisant les forces sur **quelques sites bien identifiés** et en poursuivant les financements dans le temps, pour **assurer leur pérennité**. A cet égard, il ne semble pas non plus souhaitable d'imposer une centralisation excessive sur un seul équipement national, car tant l'émulation que la proximité thématique et géographique apparaissent bénéfiques.

5.1.2 Centre de ressources et support en ingénierie

Au delà des plateformes technologiques de crowdsourcing, les chercheurs recourent tous à des **équipements logiciels** qu'il serait bon de partager dans des centres de ressources mutualisées. Cela permettrait à la fois de valoriser des réalisations existantes, pour les offrir à tous, de les perfectionner, en tenant compte des contributions et des remarques des utilisateurs, et de faciliter la tâche des chercheurs en leur procurant des outils qu'ils n'auront pas eux-mêmes à reprogrammer.

De même, il faudrait partager les données recueillies, pour tester et comparer différentes techniques de traitement.

En complément à ces équipements logiciels, il faudrait mettre en place des **supports en ingénierie** qui s'occuperaient de la maintenance des équipements logiciels et des plateformes de crowdsourcing, tout en développant des outils et en les mettant à la disposition de toute la communauté. Ces supports sont essentiels pour éviter de voir le fruit d'efforts financiers, parfois considérable, se perdre, faute de suivi.

Ces équipes d'ingénieurs et ces centres de ressources pourraient prendre place dans des « **maisons des sciences participatives** » qui auraient pour vocation d'accompagner les chercheurs d'horizons différents et de contribuer à la pérennité des équipements. Soulignons, ici, le caractère incrémental de la recherche qui s'appuie sur les résultats et les apports antérieurs pour progresser.

5.2 Axes de recherche prioritaires

Indépendamment de ces infrastructures technologiques et de l'ingénierie, un certain nombre de problèmes ne pourront être résolus que si l'on promeut des recherches dans les sciences du numérique. Parmi ces recherches, certaines apparaissent aujourd'hui comme prioritaires. C'est en particulier le cas pour :

- ◆ La décomposition de problèmes en tâches et l'allocation des tâches aux utilisateurs,
- ◆ La construction dynamique des profils de compétences à partir des performances obtenues et la détection de « vandales », ce qui sert à la fois pour l'allocation des tâches, pour l'évaluation de la qualité des réponses et pour la validation des connaissances,

- ◆ La construction et la validation des connaissances acquises par croisement des réponses obtenues, avec l'utilisation de modèles de compétence des utilisateurs
- ◆ La mise au point de modèles économiques tenant compte des différentes formes de rétribution,
- ◆ Le traitement des grandes masses données provenant, en particulier, des objets connectés et des traces d'usage.

5.3 Emergence de compétences interdisciplinaires

Les sciences participatives ne se contentent pas de recourir à des infrastructures logicielles et à du traitement d'information. Si les sciences du numérique et les mathématiques y ont une grande part, elles ne suffisent pas pour comprendre les blocages ou les réussites de telle ou telle opération. En effet, les contributeurs y jouent un rôle central. En conséquence, pour concevoir les interfaces d'acquisition de données, pour motiver les participants et pour évaluer leurs compétences on doit aussi faire appel aux sciences de l'Homme et de la société, en particulier à la psychologie cognitive, à la psychologie sociale, à la sociologie des usages, à l'économie, etc. Cet apport des sciences de l'Homme et de la société ne peut se faire en parallèle, de façon disjointe. Il faut des équipes qui travaillent de concert, sur les projets communs, afin que la réflexion des sciences de l'Homme et de la société soit prise en compte dès la conception et qu'elle intègre les contraintes techniques et algorithmiques. Des liens entre disciplines doivent donc s'établir sur des projets conjoints. Les « maisons des sciences participatives » pourraient jouer là un rôle très positif en favorisant les échanges et en stimulant les initiatives.

5.4 Résumé des recommandations

En conclusion, les recommandations sont des quatre ordres :

1. Aider au développement de **plateformes pérennes de « crowdsourcing »** et de logiciels de traitement adaptés afin de permettre aux chercheurs d'avoir la maîtrise des différents composants, tout en évitant deux écueils, la centralisation excessive et la dispersion des efforts.
2. Mettre en place des « **maisons des sciences participatives** » bien identifiées sur le territoire national. Ces maisons abriteront des plateformes technologiques avec des **équipes supports d'ingénieurs** qui assureront la pérennité de plateformes de « crowdsourcing », leur maintenance, l'accès libre et la fourniture de ressources logicielles pour l'acquisition collaborative de données et leur traitement.
3. Financer des **axes de recherche prioritaires dans les sciences du numérique** qui porteront, entre autre, sur la modélisation des compétences des participants, la décomposition et l'allocation des tâches, la construction et la validation des

connaissances, la conception d'interfaces, la mise au point de modèles économiques et le traitement de grandes masses de données issues d'une acquisition distribuée.

4. Encourager à la faveur de projets communs l'**émergence de compétences interdisciplinaires** entre les sciences du numérique et certains secteurs des sciences de l'Homme et de la société, en particulier la sociologie des usages, la psychologie sociale, la psychologie cognitive et l'économie.

6 Références

- [1] <https://hal.archives-ouvertes.fr/hal-01067110/document>
- [2] https://en.wikipedia.org/wiki/Amazon_Mechanical_Turk#Labor_relations
- [3] <http://www.galaxyzoo.org/>
- [4] <http://slide-apps.liglab.fr/~crowdhealth/>
- [5] <https://www.innocentive.com/>
- [6] <http://www.ideaconnection.com/>
- [7] <http://www.ideaconnection.com/open-innovation-success/Open-Innovation-Goldcorp-Challenge-00031.html>
- [8] <http://crowdcrafting.org/project/graphisomorphism/>
- [9] <http://crowdcrafting.org/>
- [10] <http://nextstopdesign.com/>
- [11] <http://www.www2015.it/documents/proceedings/companion/p1531.pdf>
- [12] <http://wp.sigmod.org/?author=22>
- [13] <https://www.mturk.com/mturk/welcome>
- [14] <http://crowd4u.org/en>
- [15] <http://www.pybossa.com>
- [16] Les plateformes de « CrowdSourcing » prennent parfois la forme de jeux.
- [17] <http://fold.it/portal/>
- [18] <http://arxiv.org/abs/1008.1276>
- [19] http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503406
- [20] <http://social-ist.eu/>
- [21] <http://www.springer.com/us/book/9783319086804>
- [22] AMP Lab – Algorithmic Machine and People Laboratory
<https://amplab.cs.berkeley.edu/#>
- [23] Le site de l'AMP Lab décrit de façon très explicite ses objectifs, en faisant référence au traitement de masses de données issues du crowdsourcing : *“Working at the intersection of three massive trends: powerful machine learning, cloud computing, and crowdsourcing, the AMPLab is creating a new Big Data analytics platform that combines Algorithms, Machines and People to make sense at scale.”*