



**Action incitative  
2017  
Créativité et innovation scientifiques (CRESI)  
Formulaire**

|                                      |   |  |
|--------------------------------------|---|--|
| CIRAD - DGD-RS                       | <b>AI – N°<br/>CRESI</b>                                  | Date de clôture de l'AI<br>Le 20/02/2017 |
| Département : ES<br>UR/UMR : ART-Dev | Nom du porteur de la<br>demande : Jean-Francois<br>Le Coq |  |

### **Intitulé du projet (Thème) :**

#### **BigdataPOL**

Mobiliser des approches de Big data pour l'analyse des processus et des effets des politiques publiques dans le milieu rural

### **Description du front scientifique exploré et justification de la pertinence et de l'originalité de la proposition (1 page maximum) :**

Le front scientifique exploré par ce projet est celui de l'usage des approches de Big data pour l'analyse des politiques publiques. Il s'agira d'explorer plus précisément deux dimensions de l'usage des approches et méthodes de Big data : 1) comment les mobiliser pour analyser les processus de politique publique et en particulier évaluer le rôle de l'information et des connaissances (notamment scientifiques) dans la prise de décision au niveau des politiques (*Big data pour le processus de politique – mise en agenda et décision*) ; 2) comment les mobiliser pour analyser les effets des politiques publiques sur les processus d'innovation, la situation socio-économique et environnementale des foyers ruraux et territoires (*Big data pour l'évaluation multi-critères et multi-échelles des politiques*).

L'originalité de cette proposition tient au fait que l'usage des méthodologies adaptées au « paradigme Big Data » est peu ou pas mobilisé pour l'analyse des politiques publiques à notre connaissance tant au niveau international qu'en interne au Cirad. Or, il serait très pertinent et urgent de les mobiliser car elles connaissent aujourd'hui un rapide développement dans plusieurs secteurs de recherche comme l'entrepreneuriat (Monino *et al.* 2016), dans le domaine biomédical et écologique (Kitchin, 2014; Jin *et al.*, 2015) et ouvrent des perspectives pour renouveler et compléter les approches et méthodes d'analyse traditionnelles des processus et des effets des politiques.

En effet, au niveau international les approches de Big data constituent un secteur de recherche en plein essor porté par les avancées technologiques (capacité des ordinateurs), la connectivité (Web, multiplication des bases de données et de l'*opendata*<sup>1</sup>, etc.) et les algorithmes s'appuyant sur ce nouveau paradigme. Ce concept, généralement caractérisé par les quatre V (Volume, Vitesse, Variété, Valeur), ouvre ainsi la voie à de nouvelles disciplines de recherche telle que la Science des Données qui mêle mathématiques, statistiques, informatique et visualisation, afin de répondre à des problématiques spécifiques en particulier en termes de traitement de données hétérogènes (*Variété*) et de qualités disparates (*Valeurs*).

Dans le champ de l'analyse des politiques publiques, bien que des avancées conceptuelles et méthodologiques permettent de mieux évaluer les politiques et d'appréhender les processus de

<sup>1</sup> Dans la plupart des pays émergents d'Amérique latine (Brésil, Chili, Mexique, Colombie), les états ont mis en place des portails de la transparence où il est possible d'avoir accès à l'ensemble des données sur les élections, les budgets des politiques publiques et les chiffres de leur mise en œuvre. Il est également devenu possible aux laboratoires de recherche de négocier des traitements spécifiques des recensements de population, des recensements agricoles ou enquêtes de budget des ménages auprès des instituts nationaux de statistiques.

politiques, des fronts de sciences subsistent que le Big data pourrait contribuer à résoudre. Concernant l'évaluation des effets des politiques, les approches quantitatives se sont raffinées et renforcées (notamment avec l'usage de variables randomisées, etc.) mais ces approches sont coûteuses et souvent difficiles à mettre en œuvre par manque d'intérêt et d'acceptation sociale (Baylis *et al.*, 2016). Le Big data pourrait dès lors permettre de mettre en place des protocoles moins coûteux, plus larges, voire plus robustes. Concernant la compréhension des processus politiques (mise en agenda, définition et choix des politiques), des cadres analytiques et méthodologiques ont été proposés afin d'appréhender de la complexité des processus de politique en intégrant notamment le rôle des variables cognitives (idées, discours,..) (Sabatier, 2007) en complément des variables de pouvoirs et d'intérêts (Surel, 1998) pour rendre compte de la complexité des processus de politique. Ces cadres d'analyse mobilisent ainsi une diversité de variables dont le rôle respectif demande encore à être testé de manière plus systématique et rigoureuse. L'usage d'approches de Big data semblent prometteuses pour dépasser ces limites en permettant de mieux capturer (et à moindre coût) certaines variables (idée, information, etc.) en mobilisant des sources jusque-là non mobilisées (base de données, open data, web) et de les mettre en perspective avec d'autres plus couramment mobilisées (base de données, enquêtes, entretiens).

Au sein du Cirad, l'élaboration de méthodes permettant de renforcer la connectivité entre les bases de connaissances que propose les Big data et leur mise en usage pour l'analyse ou pour l'orientation des politiques publiques permettrait de créer et tirer parties de synergies entre deux communautés scientifiques qui n'ont pas (ou très peu) d'interactions. Ceci permettrait de renforcer *in fine* la capacité d'analyse collective du Cirad face aux grands enjeux des sociétés du Sud, que représentent l'orientation des politiques publiques pour faire face aux défis du renouvellement des objectifs du développement durable : réduction de la pauvreté, de la dégradations des ressources naturelles et de la mise en place de système d'alimentation durable (axe 5, et en partie 3 et 6 de la stratégie actuelle du Cirad). En effet, d'un côté, les chercheurs de différents départements et unités mobilisent des approches de Big data dans des domaines très divers tels que la veille en épidémiologie animale, les analyse de parcours migratoires (M. Roche) ou la génomique (M. Ruiz). Ces travaux ont permis d'acquérir une expertise en termes de traitement de la variété (ou l'hétérogénéité) et de la valeur des données (cf. projet SONGES <http://textmining.biz/Projects/Songes>). De l'autre, la communauté des chercheurs travaillant sur les politiques publiques (principalement situé à ES, ART-Dev, E&F, GREEN, INNOVATION, etc...) mobilisent des approches principalement qualitatives (Science po – cas de ART-Dev, Impact pathway cas de INNOVATION), de la modélisation participative (Green) ou des méthodes quantitatives (E&F, Moisa). Par ailleurs, ces approches nécessitent des enquêtes de terrain importantes qui ne valorisent pas le potentiel d'information et de connaissances que propose le Big data (nouvelles sources de données, bases de données hétérogènes issues d'institutions, d'acteurs ou des ressources en open data, et qui sont en plein essor<sup>2</sup>). Le Cirad dans le cadre du chantier IMPACT développe également une approche méthodologique (ImpresS) pour évaluer la contribution de la recherche à l'innovation et aux mécanismes de développement. Cette démarche méthodologique aborde le référencement des relations entre les outputs et outcomes générés par la recherche et propose des outils pour établir ces causalités. Cependant l'analyse numérique des conditions de mobilisation par les politiques publiques des bases d'information et de connaissances scientifiques reste encore exploratoire d'un point de vue méthodologique. Cette initiative apparaît donc comme une contribution permettant de solidifier les capacités de l'institution à rendre compte de ces impact sur le développement au prisme des orientations données aux politiques publiques selon différentes échelle d'analyse possibles aux niveau régional ou local porté par les

---

<sup>2</sup> On peut mentionner à titre d'exemple l'initiative du CGIAR de mise en place d'une plateforme Big data afin de rendre accessible toutes les données produites les centres de recherche agronomiques.

collectivités territoriales. Enfin, d'un point de vue partenarial et de positionnement scientifique, l'élaboration d'approches Big data pour l'analyse des politiques renforcerait les collaborations possibles avec l'initiative de plateforme Big data du CGIAR et doterait d'outils d'analyse innovants auprès des partenaires du réseau des politiques publiques en Amérique Latine (DP PPAL).

### **Déroulement du projet (nature des actions et leur capacité à explorer le front scientifique ciblé, calendrier prévisionnel des activités programmées, résultats attendus) :**

Le projet se déroulera en 4 étapes : la réalisation d'un état de l'art, la tenue d'un atelier d'échange et de définition collective, la mise en place d'un test de méthodologie exploratoire sur des cas d'étude en Amérique Latine, et une phase réflexive et de valorisation. Les données identifiées et collectées sur les terrains respecteront un plan de Gestion de Données (PGD) associé au projet.

La première étape consistera donc en **une revue de littérature** spécifique sur les thèmes d'intersection entre Big data et l'analyse des politiques publiques. Il s'agira de faire le point sur les avancées récentes sur cette interaction, en prenant soin d'élargir la recherche au-delà des politiques rurales (politiques d'innovation, politiques de santé,...). Ce travail sera réalisé par un binôme d'étudiants de « science sociale (science politique) » et « science des données » (informaticien/statistique), qui mèneront de manière parallèle et complémentaire leur stage pendant l'ensemble de la durée du projet sous la supervision d'un binôme de chercheur (J.F. Le Coq, M. Roche)<sup>3</sup>. De manière complémentaire, une première réflexion sera conduite pour proposer un protocole exploratoire d'analyse et identifier des études pilotes pour le tester. Pour ce faire, un outil d'échange sera mis en place pour inciter les chercheurs du projet à apporter de la littérature et partager leurs réflexions sur le protocole et le choix d'études pilotes, et servir ainsi de support à une animation virtuelle préparatoire à l'atelier. Cette première étape donnera lieu à un premier rapport faisant l'état de l'art des recherches et méthodes de Big data, des limites et difficultés rencontrées, de leurs potentiels pour l'analyse des politiques publiques (Livrable L1).

La deuxième tâche consistera en la réalisation d'un **atelier d'échange d'expérience et de construction collective d'un protocole exploratoire d'étude**. Réalisé sur 2 jours en juin 2017, il visera à présenter les résultats de l'état de l'art (L1) et faire une revue d'expériences en terme d'analyse des politiques publiques (questions de recherche, cadres d'analyses, leurs apports et limites). Cet atelier sera animé par des chercheurs en politique publique et en Big data. Ces derniers dresseront un panorama des méthodes automatiques de collecte de données, de fouille de données, fouille de textes, Web sémantique utiles pour le projet. Sur la base d'un dialogue entre ces communautés scientifiques, cet atelier permettra de définir et valider une méthode exploratoire à tester sur des cas d'étude pilotes. Il s'agira en particulier, à partir des variables considérées pertinentes, à prendre en compte pour l'analyse des processus et des effets des politiques publiques aux yeux des chercheurs en sciences sociales, d'identifier avec les chercheurs de Big data : les indicateurs, les sources d'informations, et méthodes de traitement à mobiliser dans un protocole pilote. Un rapport d'atelier (L2) sera réalisé. Il présentera les échanges et précisera le protocole pilote (modèle conceptuel, variables retenues, sources de données, méthode d'extraction et d'analyse) à tester dans les études de cas retenues.

La troisième tâche consistera en un **test de la méthode exploratoire** définie et validée pendant l'atelier. Trois études de cas pilotes potentielles ont été pré-identifiées en lien avec des dynamiques de travail existantes dans le cadre du DP PPAL et au CIRAD : la thématique

---

<sup>3</sup> Il s'agira de leur première phase de stage, qui se poursuivra avec la réalisation des cas d'études.

retenue est celle des politiques visant à la promotion de l'agroécologie et les trois terrains retenus sont :

- le Brésil, où une politique nationale d'agroécologie est ancienne et bien établie et où de multiples partenariats et des dispositifs existant permettent d'envisager l'accès à des informations de sources multiples
- la Colombie, où des dispositifs visent à promouvoir des pratiques agroécologiques en lien avec les enjeux climatiques sont mis en place et où un partenariat sera établi au CIAT avec des chercheurs des deux communautés scientifiques, Big Data et sciences politiques
- la Guadeloupe où le Cirad dispose de dispositifs et de travaux de recherche sur les thèmes de l'agroécologie et où un lien sera établi avec le projet RIVAGE traitant plus spécifiquement des problématique de pesticides<sup>4</sup>.

Le choix final des études pilotes sera réalisé/validé sur la base de la faisabilité (accès aux données, capacité à traiter sur un même cas l'analyse de processus et d'effets de politique) qui sera exploré pendant la phase préparatoire du projet (phase 1). Dans les cas d'étude pilote retenus, il s'agira d'élaborer un modèle explicatif spécifique à tester en mobilisant des cadres analyses et théoriques existants, d'identifier les sources d'information et mobilisables pour l'analyse, et enfin de tester le protocole d'analyse de ces données. Ce travail sera réalisé par le binôme d'étudiants de master ayant réalisé l'état de l'art et sous la supervision de chercheurs des deux disciplines. L'étudiant de Science Sociale se chargera plus particulièrement de l'identification des variables explicatives et des sources de données potentielles ainsi que de l'interprétation des analyses au regard de la science politique. Il s'agira pour lui *in fine* de discuter la pertinence d'un cadre d'analyse et d'un protocole mobilisant les Big data dans les sciences politiques. L'étudiant de Science des Données (informatique/statistique) se chargera du traitement et l'analyse des informations extraites, s'attachant à résoudre plus particulièrement les problématiques de Variété (ou l'hétérogénéité) et de Valeur (ou qualité) des données. Il s'agira pour lui 1) d'extraire automatiquement les indicateurs à partir des données en s'appuyant sur les outils existants (Stanford NLP, Polyglot, etc.) qui extraient les Entités Nommées standards (institutions, organisations, lieux, personnes, etc.). Par ailleurs, des méthodes originales adaptées aux indicateurs plus spécifiques et identifiées dans la tâche précédente seront proposées. L'ensemble de ces approches seront mise en œuvre et évaluées sur les données issues des cas d'étude retenus, 2) de mettre en relation les données hétérogènes et de qualités diverses (sur la base des indicateurs définis et extraits) pour tester le modèle conceptuel d'analyse. Pour cette seconde activité qui constitue un des verrou important du projet, il s'appuiera sur les connaissances acquises dans le cadre du projet SONGES et une thèse portée par TETIS (Cirad-Irstea) sur la mise en relation des données hétérogènes (thèse de Jacques Fize, TETIS). Par ailleurs, il s'appuiera sur des méthodes incrémentales pour intégrer des connaissances expertes (par exemple, par l'utilisation d'approche de type « Relevance Feedback » (Kelly & Teevan, 2003). *In fine*, ce stage permettra d'améliorer les méthodes de la littérature en terme d'usage et de mise en relation de données hétérogènes et d'évaluation de leur pertinence (Adderley *et al.* 2014).

La quatrième tâche consistera en une analyse réflexive et collective des résultats du test méthodologique. Il s'agira dans les terrains d'étude d'appréhender la réception et l'usage des résultats obtenus avec la méthode testée par les acteurs des politiques via des réunions de restitutions ou des entretiens ciblés. Puis, entre les chercheurs du projet, une visio-conférence

---

<sup>4</sup> Financé sur fonds européens FEDER et s'inscrivant dans l'axe environnement du DOCUP Guadeloupe, ce projet regroupe essentiellement des compétences en sciences biophysiques de PERSYST (et de l'INRA) et associe aussi TETIS sur une approche qui traite de la coordination d'acteurs au niveau d'un territoire, et de l'usage de l'information par les acteurs du territoire (J.P. Tonneau et J.P. Chery, TETIS). Ce projet offre l'opportunité de pouvoir financer des stagiaires en lien avec le test méthodologique qui sera développé par BigDataPOL.

sera réalisé afin de partager les résultats des études pilotes, d'analyser les points forts et les limites de la méthode testée et d'en proposer des améliorations. Cet échange donnera lieu à deux produits : un compte rendu de la réunion incluant un protocole consolidée, un synopsis d'article collectif de nature méthodologique (L4) sur l'usage de Big data dans l'analyse des politiques publique (L4), et une *concept note* (L5) pour un projet partenariale en vue d'une soumission à appel d'offre ou de négociation avec des bailleurs en 2018.

### Calendrier des activités

| Activité   | Mar. | Avr. | Mai | Jun. | Jul. | Aoû. | Sep. | Oct. | Nov.  | Déc.   |
|--|------|------|-----|------|------|------|------|------|-------|--------|
| <b>I. Revue bibliographique et Etat de l'art</b>   |      |      |     |      |      |      |      |      |       |        |
| I-1. Lancement : mise en place plateforme d'échange et réunions skype des partenaires  |      |      |     |      |      |      |      |      |       |        |
| I-2. Revue bibliographique (St.1 et St2)   |      |      |     |      |      |      |      |      |       |        |
| I-3. Rédaction d'un état de l'art (St.1 et St2)  |      |      |     | L1   |      |      |      |      |       |        |
| <b>II. Atelier d'échange et de construction collective d'une méthodologie exploratoire à tester</b>  |      |      |     |      |      |      |      |      |       |        |
| II-1. Préparation atelier (proposition méthodologique et étude de cas pilotes)   |      |      |     |      |      |      |      |      |       |        |
| II-2. Réalisation de l'atelier   |      |      |     |      |      |      |      |      |       |        |
| II-3. Rédaction compte rendu de l'atelier  |      |      |     |      | L2   |      |      |      |       |        |
| <b>III. Test de la méthodologie sur le cas d'études majeur identifié</b>   |      |      |     |      |      |      |      |      |       |        |
| III-1. Définition d'un Modèle conceptuel pour l'analyse de politique et identification des sources de données (St.1)<br>Evaluation d'outils d'extraction d'entités nommées et implantation / évaluation des règles d'extraction pour de nouveaux indicateurs (St2) |      |      |     |      |      |      |      |      |       |        |
| III-2. Analyse des données (mise en correspondance / relations de données) – (St2). et d'interprétation des données (St1, St2)   |      |      |     |      |      |      |      |      |       |        |
| III-3. Mission d'appui   |      |      |     |      |      |      |      |      |       |        |
| III-4 Rédaction rapports d'étude de cas  |      |      |     |      |      |      |      |      | L3a,b |        |
| <b>V Retour sur l'expérience, valorisation académique et préparation de projet</b>   |      |      |     |      |      |      |      |      |       |        |
| IV-1. Réunion de présentation de résultats des tests pilote  |      |      |     |      |      |      |      |      |       |        |
| IV-2. Rédaction d'un <i>outline</i> détaillé d'article méthodologique et d'une <i>concept note</i> de projet   |      |      |     |      |      |      |      |      |       | L4, L5 |

NB : St1 : stagiaire en science sociale, St2 stagiaire en science des données

### Financements demandés et temps chercheur consacré à l'action :

Les financements sont demandés pour la réalisation d'un atelier, de 2 missions d'appui méthodologique et de deux études pilotes (bourse de master et travail de terrain - test de méthodologie/analyse). Il s'agira en particulier de financer :

- Deux bourses de master pour avoir un binôme profils « science sociale » / « science des données - Big data » pour les recherches bibliographiques (état de l'art) et le test

de la méthodologie dans les cas d'étude (développement informatique/analyse de politique)

- Les frais de l'atelier d'échange et d'élaboration collective de méthodologie (billet avion et séjour de 4 participants « extérieurs » : collègues expatriés, partenaires et invité extérieurs) à Montpellier pendant 2 jours.
- Les frais des 3 études pilotes : frais de terrain et 2 missions d'appui de chercheurs Cirad

Par ailleurs, nous envisageons de mobiliser des ressources issues de l'Institut de Convergence "Agriculture numérique" (#DigitAg), en s'adossant aux axes 4 (SI et gestion de données) et 5 (big data et extraction de connaissances). Cet institut encourage les travaux interdisciplinaires qui sont caractéristiques de notre projet. Des financements de stages permettraient d'approfondir le traitement des données du projet ou développer la méthodologie sur des terrains secondaires afin de valider la généralité des approches.

| <i>Catégorie</i>  | <i>Montant (€)</i> | <i>Commentaires</i>   |
|---|--------------------|---|
| Deux bourses de master (un binôme étudiant « Science social » et « science de l'information - Big data ») | 11.000             | Ces étudiants en binômes seront chargés de la revue bibliographique et de la conduite des études de cas |
| Atelier de 2 jours à Montpellier sur les approches d'analyse de politiques et de Big data                 | 10.000             | Entre 20 et 30 participants dont 3 invités extérieurs   |
| Etude de cas pilote / test méthodologique   | 3.000              | Déplacement, entretiens (3x1000)  |
| - Frais de terrain locaux   | 6.000              | Mission de chercheur sur chacun des sites (3 x2000)   |
| - Missions d'appui méthodologique   |                    |   |
| <b>Total</b>  | <b>30 000</b>      |   |

Le projet fera appel donc à des collègues des deux communautés de recherche (« science social / politique publique » / « science des données - Big data ») de diverses disciplines et venant des différents départements et unité de recherche du Cirad. En outre, seront mobilisés des partenaires à travers le dispositif de PP-AL (politiques publiques et développement territorial en Amérique Latine) et en particulier du CIAT qui a, au sein de DAPA, un groupe de travail Big data. Des chercheurs hors de ces réseaux mais ayant une expérience en matière d'usage de Big data pour l'analyse des politiques publique seront identifiés et invités à participer à l'atelier. Au total, une vingtaine de chercheurs contribueront au projet.

#### **Chercheurs principaux (Cirad)**

| <i>Chercheur</i>                               | <i>Temps (mois)</i> | <i>Domaine de recherche</i>   |
|--|---------------------|---|
| Jean-François Le Coq<br>(ES – <b>ART-Dev</b> ) | 3                   | <b>Coordination</b> , Agroéconomiste, Analyse politique publique agricoles et environnementales |
| Mathieu Roche<br>(ES – <b>TETIS</b> )          | 3                   | <b>Co-coordination</b> , Sciences des données, Spécialiste Big data (analyse textuelle)         |
| Eric Sabourin<br>(ES - <b>ART-Dev</b> )        | 1                   | Sociologie, Analyse politique publique rurales (agro-écologie, agriculture familiale,...)       |
| Gilles Massardier<br>(ES – <b>ART-Dev</b> )    | 1                   | Science politique ; Analyse des processus de politique  |
| Muriel Bonin<br>(ES – <b>TETIS</b> )           | 1                   | Géographe, analyse des politiques agro-environnementale   |
| Driss Ezzine De Blas                           | 1                   | Ecologie économique, Analyse impact politique publique  |

|   |   |   |
|---|---|---|
| (ES – <b>F&amp;S</b> )                          |   | environnementales   |
| Bruno Locatelli<br>(ES – <b>F&amp;S</b> )       | 1 | Analyse des politiques de changement climatique   |
| Marie Gabrielle Piketty<br>(ES – <b>GREEN</b> ) | 1 | Economiste, Analyse des instruments de politiques publiques environnementales                       |
| Ludovic Temple<br>(ES - <b>INNOVATION</b> )     | 1 | Economiste, analyse des processus d'innovation et des impacts                                       |
| Patricio Mendez del Villar (ES - <b>TETIS</b> ) | 1 | Economiste, systèmes d'information, veilles thématiques sur le fonctionnement des marchés agricoles |
| Manuel Ruiz<br>(BIOS – <b>AGAP</b> )            | 1 | Bioinformatique, génomique, Spécialiste Big data (Web Sémantique)                                   |
| Sandrine Auzoux<br>(PERSYST <b>AIDA</b> )       | 1 | Science des données, spécialiste Big data (Système d'Information), Agro écologie                    |
| Philippe Cattan<br>(PERSYST – <b>GECO</b> )     | 1 | Agronomie, impacts environnementaux   |

### Autres chercheurs impliqués

| <i>Chercheur</i>                                 | <i>Domaine de recherche</i>   |
|--|---|
| Chris Béné<br>(CIAT DAPA)                        | Analyse des Politiques publiques  |
| Daniel Jimenez<br>(CIAT DAPA)                    | Sciences des données, spécialiste Big Data (fouille de données)                               |
| Jean-Philippe Tonneau<br>(ES – <b>TETIS</b> )    | Géographe, analyse des politiques publiques de développement territorial                      |
| Maguelonne Teisseire<br>( <b>TETIS</b> - Irstea) | Sciences des données, spécialiste Big data (Fouille de données)                               |
| Dino Ienco<br><b>TETIS</b> – Irstea)             | Sciences des données, spécialiste Big data (Fouille de données et apprentissage)              |
| Sandra Luque<br>( <b>TETIS</b> – Irstea)         | Géographe, Écologie du Paysage  |
| Didier Babin<br>( <b>TETIS</b> /UQAM – Cirad),   | Comptabilité Ecosystémique du Capital Naturel   |
| Nicolas Merveille<br>(UQAM)                      | Anthropologue, Co-titulaire de la Chaire Internationale sur le Cycle de Vie - CIRAIG-ESG-UQAM |

## Suites envisagées à l'issu du projet financé par l'action incitative :

### PROJETS

La méthode d'analyse de politique élaborée et testée dans le projet BigDataPOL sera proposée pour intégrer des propositions de projets de recherche et de développement. Plusieurs pistes concrètes seront explorées : l'intégration d'une composante d'analyse de politique dans des projets existants, tels que la composante « agriculture résiliente et sécurité alimentaire » du programme européen « Euroclima + » (14 Millions \$US) ou le programme CRP CCAFS qui vise à identifier et développer la mise en place de pratiques agricoles pour l'adaptation au changement climatique. En outre, une proposition de projet sera soumise en réponse à un appel d'offre de fondations privées (Tinker, Futur Earth, Ford,...) ou publique de recherche (ANR, H2020) portant sur les politiques alimentaires, de gestion de ressources naturelles, ou de développement agricole. Enfin, l'approche sera présentée à des partenaires techniques et

financiers comme la BID, le FIDA ou le GEF, l'IICA, la CEPAL en vue d'obtention d'expertise concernant l'analyse des effets des politiques publiques.

### **PUBLICATION**

Un article collectif méthodologique sur l'usage des big data dans l'analyse des politiques publiques sera élaborée et soumis à une revue en Science de Données tel que « International Journal of Data Science and Analytic » et/ou de science politiques tel que « Policy Study journal ». Le cas échéant, les résultats spécifiques de l'application de la méthode aux études de cas pourront faire l'objet d'article dans des revues thématiques telle que « Global Environmental Change ». La participation à une conférence thématique ou disciplinaire sera également envisagée pour faire connaître la démarche et les résultats. Enfin, de manière complémentaire, et si cela apparaît nécessaire pour obtenir des contrats d'expertises, on pourra élaborer et publier un manuel méthodologique pour l'analyse des politiques afin de servir de base à des formations spécifiques.

### **POURSUITE DE LA REFLEXION**

Conscient que les champs ouverts par le projet BigDataPOL est large et prometteur, il s'agira également de trouver les moyens de poursuivre une animation scientifique méthodologique. Cela pourra prendre la forme d'école-chercheur en lien avec le DP PPAL ou Govin, l'élaboration d'un projet de thèse co-suivi par les chercheurs impliqués, ou de sessions spécifiques dans le cadre de congrès internationaux tel que l'Association Française des Science Politique, Association Latino Américaine de Science politique, le Réseau de Recherche sur l'Innovation. Ce projet est un levier scientifique qui fera émerger de nouvelles questions de recherche à l'interface entre les sciences sociales et le Big data. Ceci permettra de proposer des sujets de thèse dans le cadre l'Institut de Convergence #DigitAg en 2018.

### **Argumentaire justifiant de l'impossibilité de mobiliser des financements externes :**

Alors que la demande des tutelles de la recherche, des bailleurs de projets et des acteurs politiques est de plus en plus forte sur le fait de mieux caractériser et mesurer les effets des interventions (de recherche et des projets) sur les politiques et *in fine* sur la résolution des problèmes de la société, la réflexion pour l'amélioration de méthodes est difficilement finançable. Etant donné que l'application des approches d'analyse de politique dans un contexte Big data n'existe pas (ou pour le moins pour nos domaines de recherche – agriculture dans les pays du Sud), il faut d'abord démontrer que ce croisement est prometteur et peut apporter des solutions originales et efficaces. C'est l'objet de ce projet d'explorer et tester cette démarche pluridisciplinaire, afin d'avoir, vis-à-vis des bailleurs de coopération et de développement, des résultats de pilotes concrets. Les résultats scientifiques valorisés par une ou des publications scientifiques permettront d'établir les conditions d'adoption de cette innovation pour les bailleurs de recherche (type H2020, ANR).

---

Adderley R., Seidler P., Badii A., Tiemann M., Neri F., Raffaelli M.. (2014). Semantic Mining and Analysis of Heterogeneous Data for Novel Intelligence Insights. Proceedings of The Fourth International Conference on Advances in Information Mining and Management, p.36-40

Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., Lapeyre, R., Persson, U. M., Pfaff, A. and Wunder, S. (2016), Mainstreaming Impact Evaluation in Nature Conservation. Conservation Letters, 9:58–64.