



Investissements d'avenir

Transition numérique de l'Etat et modernisation de l'action publique

Programme : Industrialisation de la mise à disposition de données ouvertes

Appel à projet : « Développement de l'open data au sein des territoires »

PIC'EAU

**Potentialiser l'Intelligence Collective
pour comprendre et valoriser les
données sur l'EAU**



DESCRIPTION DETAILLEE DU PROJET



1. LA SOLUTION PROPOSEE

• Périmètre, produits, services visés par le projet

La connaissance de l'eau et des milieux aquatiques en général est au centre de nombreux défis environnementaux et sociaux (pollutions, partage de la ressource, changement climatique,...). Organiser les données publiques et les diffuser pour répondre à ces défis est une mission majeure de l'Etat. Le projet PIC'EAU s'inscrit dans ce cadre, pour faciliter la diffusion et la valorisation des données sur l'eau.

Le projet vise ainsi à stimuler l'utilisation des données sur l'eau en offrant :

- des outils d'analyse, d'agrégation et de croisement de données brutes provenant aussi bien des bases de données nationales que de bases de données locales couplés à ...
- ... des mécanismes permettant aux utilisateurs de faire remonter auprès des utilisateurs et des producteurs de données des anomalies identifiées lors de leur exploitation pour améliorer la qualité des bases de données, ce qui en améliorera leur compréhension. Ces indications pourront s'appuyer aussi sur les dîres d'experts.

Ces composants, accessibles via des Interfaces de Programmation Applicative (API), seront exposés publiquement sur le web afin de faciliter leur intégration dans les outils existants utilisés par les différents acteurs de la gestion de l'eau, que ce soit des portails web (visualiseur du Système d'Information sur l'Eau : Cartograph¹, portails locaux – ex : SIGORE², gironde.fr, Gironde Scopie, portails thématiques – ex : SIGES³...) ou des outils métiers (client SIG OpenSource type QGIS, client ad hoc).

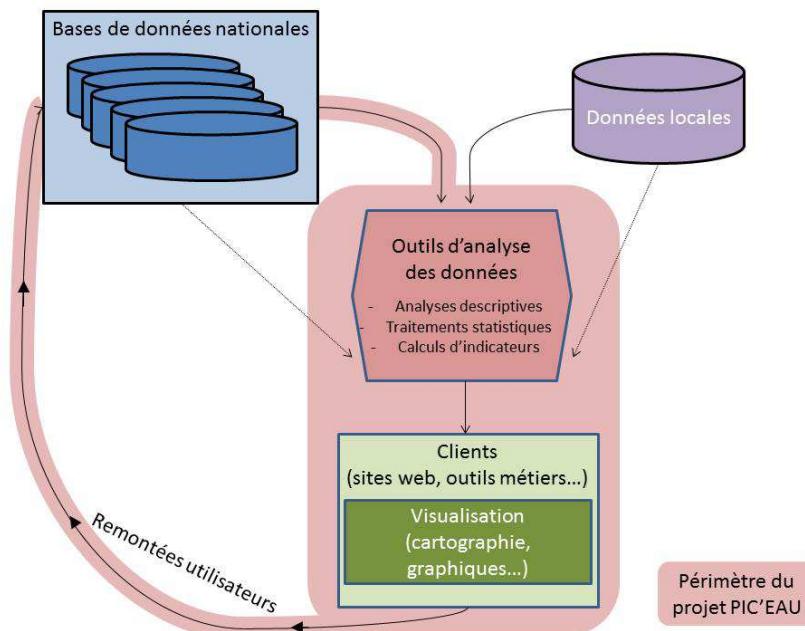


Figure 1 : Périmètre du projet PIC'EAU

¹ <http://www.cartograph.eaufrance.fr/>

² <http://cartographie.nature33.fr/>

³ <http://sigesaqi.brgm.fr/>

L'action opérationnelle est ancrée dans le territoire du Département de la Gironde (33) qui servira de terrain d'expérimentation pour l'amorçage des innovations. Les solutions nouvelles de diffusion d'outils d'analyse seront développées et testées à partir des données sur la quantité et la qualité des eaux souterraines. La plateforme de diffusion des données du Système d'Information de l'Eau (SIE), Hub'Eau, et notamment la base de données nationales sur les eaux souterraines, ADES⁴, constitueront le socle des développements réalisés durant le projet PIC'EAU. A la suite du projet, les solutions innovantes de diffusion d'outils d'analyse pourront être déployées sur l'ensemble des données du domaine de l'environnement en général, et sur l'ensemble du territoire national.

Les outils développés visent en premier lieu un public professionnel, mais les retombées du projet seront importantes pour les citoyens à plus d'un titre, notamment par la mise à disposition d'outils de calcul des données sur l'eau et par l'amélioration de la qualité des bases de données publiques, qui par leur meilleure compréhension en deviendront plus précises.

- Description détaillée de la solution adoptée, innovations et ruptures technologiques impliquées**

Le projet a pour objectif de changer le rapport des utilisateurs à la donnée sur l'eau en passant d'une consommation statique de la donnée, où l'utilisateur télécharge les données des bases nationales et applique dessus ses propres traitements, à une utilisation dynamique, dans laquelle il devient un acteur à part entière de la création et du partage de l'information sur l'eau.

Le projet vise à créer un « cercle vertueux » représenté en Figure 2 dans lequel la qualité des données est continuellement améliorée.

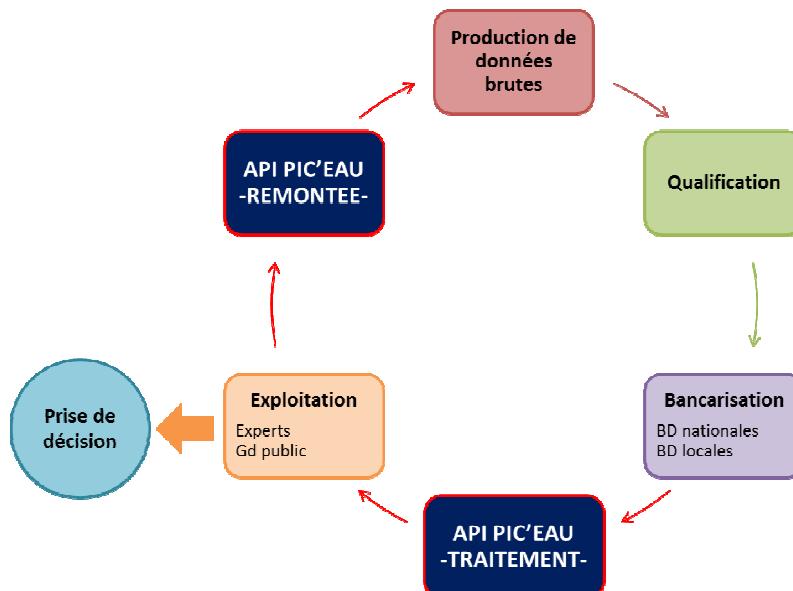


Figure 2 : Cercle vertueux de réutilisation des données sur l'eau selon le projet PIC'EAU

Le projet repose en effet sur la conviction que plus la diffusion et l'exploitation des données sont facilitées, notamment par la mise à disposition d'outils de visualisations et de traitements, plus les personnes les utilisent et plus ils sont susceptibles d'identifier et de faire remonter, via des outils

⁴ <http://www.ades.eaufrance.fr/>

spécifiques, des informations auprès des producteurs qui qualifient leurs données. La qualité des bases de données est ainsi améliorée stimulant d'autant leur réutilisation.

La solution proposée repose sur le développement de 2 différents types d'API qui pourront s'intégrer dans les clients existants de tout type (applications Web, clients SIG, outils métiers) utilisés par les acteurs locaux :

- **Des API de « traitement »** dédiées au calcul d'indicateurs temporels et spatiaux, de mise en confrontation et de traitements statistiques sur les données des bases nationales ainsi que sur des données locales fournies par l'utilisateur. Les traitements proposés permettront aux utilisateurs de mieux s'approprier les données afin d'en tirer un maximum d'information mais aussi de les enrichir ;
- **Des API de « remontée »** d'expérience sur l'exploitation des données par l'utilisateur afin de faire partager l'expertise individuelle auprès des autres utilisateurs et de faire remonter les anomalies identifiées aux producteurs de la donnée. A noter que les anomalies susvisées ne sont pas des erreurs de cohérence de la donnée, mais des indications précises issues de l'expertise locale, permettant de fiabiliser la teneur de l'information. Ce dernier type d'API demande des innovations techniques importantes. Dans le cadre du projet PIC'EAU, il sera principalement question de tester la faisabilité d'un tel type d'API.

Le premier type d'API consiste à offrir des outils d'analyse des données publiques disponibles ainsi que des données locales en possession de l'utilisateur. Les traitements proposés dans le cadre de ce projet reposeront sur des méthodes connues, déjà mises en œuvre et ayant montré leur intérêt au travers de projets antérieurs. L'enjeu réside donc principalement dans la mise en place d'une plateforme d'exposition de traitements faisant ses preuves, pour la durée du projet, sur les traitements envisagés à court terme, mais devant offrir un niveau de généricité et de souplesse suffisant pour ouvrir ensuite à la plupart des traitements utiles à la communauté, en minimisant les risques d'incompatibilité technologiques ou les obligations de portage.

La seconde partie du projet consiste à concevoir et mettre en place un composant central capable « d'écouter » et par la suite bancariser les actions ou événements (et non directement les nouveaux états des données manipulées, conformément au pattern propre au CQRS⁵) survenus lors de toutes les utilisations de la donnée. Ce composant sera in fine le détenteur de toutes les suspicions d'anomalies, les propositions de corrections et d'enrichissement (comblement de lacunes dans les données) déclenchées par les utilisateurs finaux lors de leurs études locales, et sera accessible via une API d'écriture (remontée) et une API de lecture (consultation du journal d'événements, total ou par objet, par type, par utilisateur...). Ce composant pourra également gagner à se doter d'un mécanisme de publication / souscription (PubSub), permettant à tout système de s'abonner à certains types d'événements.

Au final, la solution détaillée proposée par le projet trouve une description synthétique dans les 2 schémas en Figure 3 et Figure 4, décrivant respectivement la situation actuelle sans la plateforme PIC'EAU, et la situation visée par le projet. On y voit une situation actuelle marquée par la non

⁵ *Command Query Responsibility Segregation*, cadre d'architecture qui repose sur un principe de séparation, au sein d'une application, des composants de traitement métier de l'information (“command” / écriture) et de restitution de l'information (“query” / lecture)

capitalisation des traitements, des corrections et des enrichissements réalisés lors des exploitations locales, ainsi qu'une perte évidente d'efficience dans le croisement des données nationales / locales et dans la mutualisation des traitements de référence (peu ou pas accessibles pour les utilisateurs locaux, et donc « forkés » ou réimplémentés). La Figure 4 omet volontairement la présentation de l'exploitation du journal d'actions issues des utilisations / corrections des données. Ce composant ouvre à de nombreuses utilisations toutes très différentes, dont la pertinence dépendra à coup sûr de chaque domaine et en particulier de son organisation entre partenaires et de sa gouvernance.

La solution proposée présente l'avantage de laisser de nombreuses possibilités et choix d'intégration. Dans un cas, un tel journal pourra servir à retourner vers les producteurs afin de les informer des utilisations et des anomalies détectées en toute fin de chaîne et les inviter à les prendre en compte en tout début de chaîne (cycle plus vertueux s'il est appliqué sur la base d'une référence, en l'occurrence celle des producteurs). Dans un autre cas, la simple exposition de ce journal sur un portail public (en complément de la fiche descriptive de chaque objet) revêtira déjà un grand intérêt (ou la simple analyse du nombre de remontées par objets...). Dans ce dernier cas, on pourra se poser la question de la pertinence de la projection (toujours selon les termes consacrés dans une approche CQRS) de ces enrichissements directement sur les données servies par HUB'EAU afin de créer un second niveau d'accès à ces données, sur une base « chaude des dernières remontées d'anomalies ».

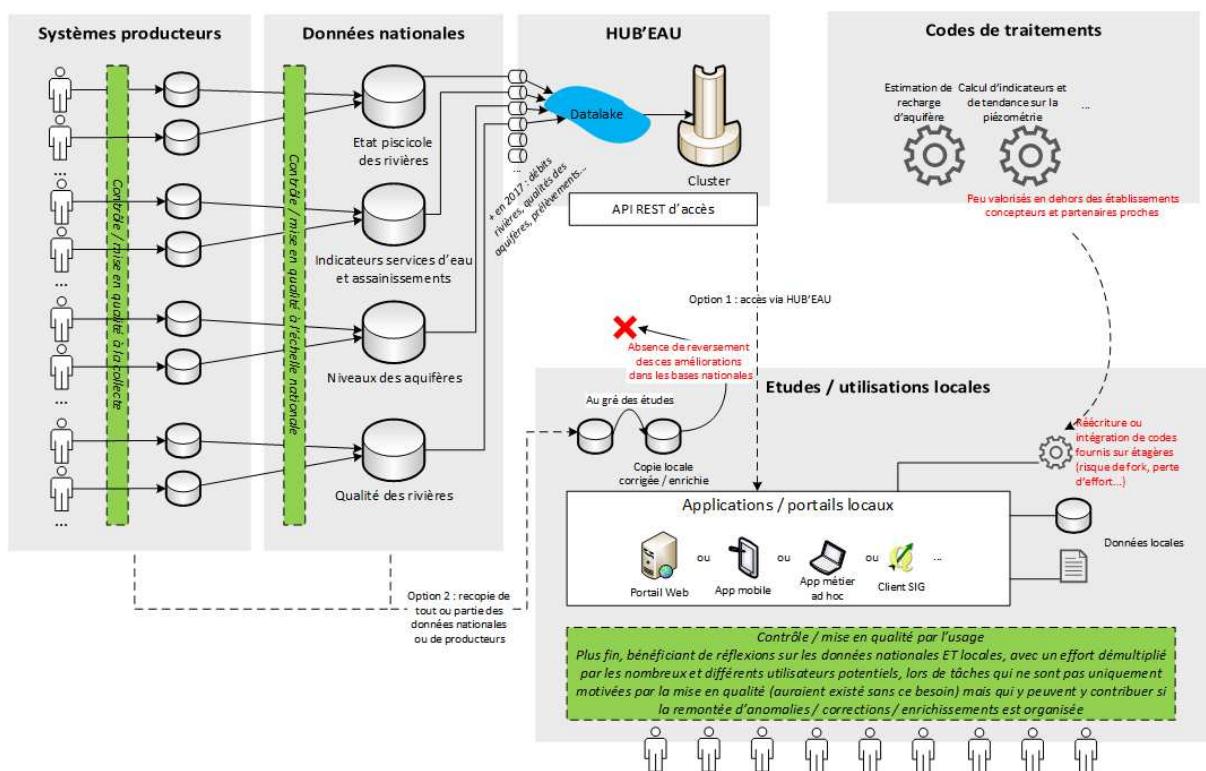


Figure 3 : Situation actuelle

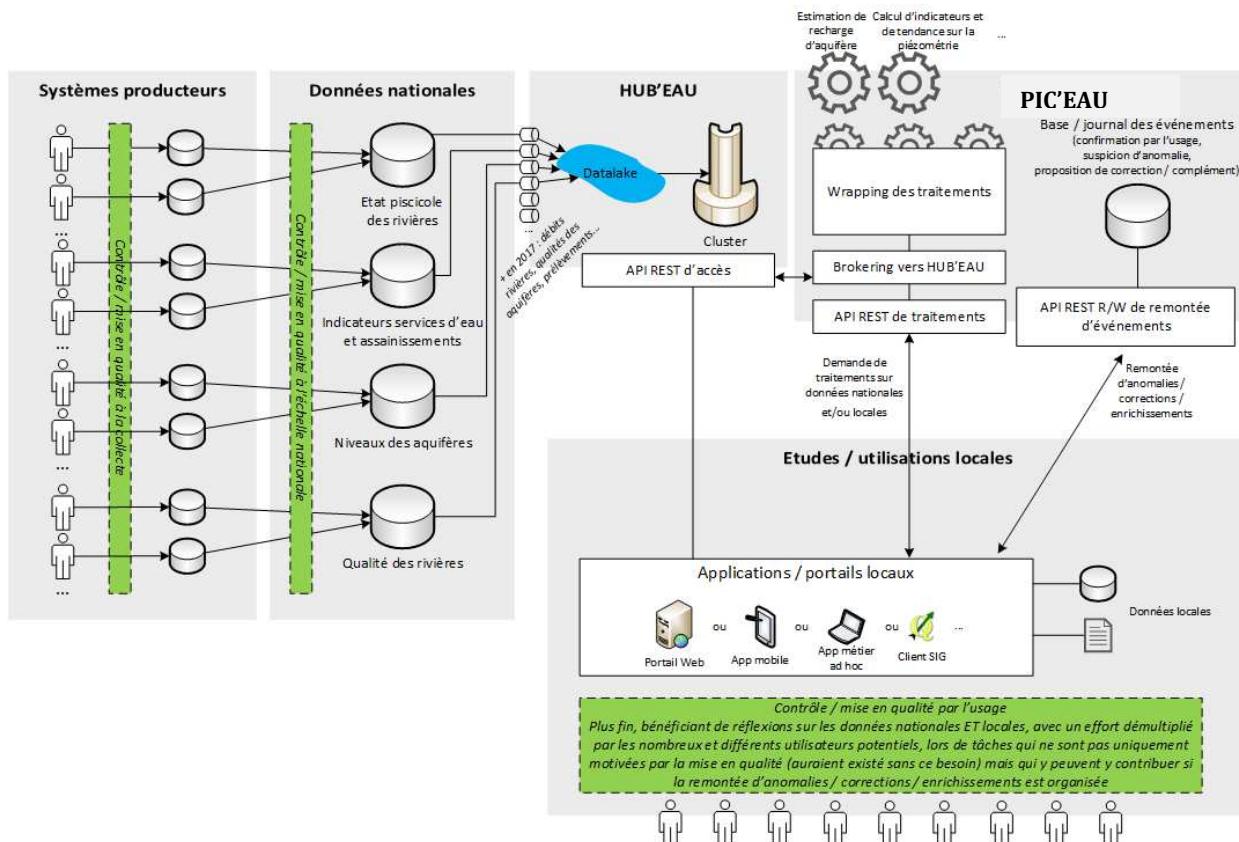


Figure 4 : Situation après PIC'EAU

• Etat des lieux, enjeux et difficultés présentes

En France, les données publiques sur l'eau sont organisées, partagées et mises à disposition en grande partie grâce au Système d'Information sur l'Eau (SIE). Depuis 2016, l'accès à ces données a été facilité grâce au développement de la plateforme Hub'Eau, financée pour partie dans le cadre des programmes d'investissement d'avenir. Cette plateforme contient aujourd'hui des données sur l'état piscicole des rivières, la qualité des rivières, le niveau de l'eau dans les aquifères et des indicateurs des services d'eau et d'assainissement ; elle est toujours en cours d'enrichissement, et de nombreuses données complémentaires devront prochainement être mises à disposition à travers cette infrastructure.

Les données brutes sur l'eau sont (ou seront) ainsi facilement accessibles. En revanche, à l'instar de beaucoup de données environnementales, ces données ne sont pas aisément analysables. Hormis le Système d'Evaluation de l'État des Eaux (SEEE) dont le périmètre se limite au calcul d'indicateurs réglementaires, peu d'outils d'analyses ou de calculs d'indicateurs sont diffusés librement auprès des utilisateurs et du grand public. Certains traitements sophistiqués sont développés pour des études particulières mais il n'y a pas de structure permettant de générer de la donnée agrégée, d'en simplifier leur diffusion et leur application à de nouveaux cas d'étude. Le projet PIC'EAU ambitionne de pallier à cette lacune en **facilitant l'analyse croisée de données locales et de données nationales**, partant du postulat que la grande quantité de données publiques disponibles pourrait être plus largement exploitée et valorisée si des outils de traitement, d'agrégation et de croisement de données provenant de différentes sources existaient. Il en résulterait une meilleure compréhension et permettrait aux territoires de mieux s'approprier les données sur l'eau.

D'autre part, les bases de données du SIE, peuvent contenir des anomalies, lacunes ou erreurs qui sont autant de freins à la réutilisation des données. Il convient de bien distinguer la cohérence de la donnée brute, qui est réalisée à plusieurs niveaux dans le schéma de production de ces données,

avec la compréhension de ces données. La qualification de données est un processus chronophage et fastidieux, surtout lorsqu'il est conçu comme un processus central réservé aux producteurs de donnée ; expliquant pourquoi il est souvent sous-dimensionné. Ce phénomène tend à s'accentuer avec l'augmentation de volumétrie des données collectées et manipulées.

Il est donc crucial à court terme de trouver des **leviers capables de démultiplier les efforts de qualification et de compréhension de l'information pour la fiabiliser**. Le projet PIC'EAU propose de **rechercher ces leviers auprès des utilisateurs locaux** (services déconcentrés de l'Etat, collectivités territoriales, acteurs privés locaux, associations...), afin de tirer profit de l'important panel d'utilisateurs qu'ils représentent, en les faisant intervenir dans le processus de qualification, sciemment ou non. Car ces utilisateurs locaux n'ont pas nécessairement plus de temps à consacrer que les producteurs à cette phase de qualification. L'objectif n'est donc pas de superposer de nouveaux outils de mise en qualité au-dessus de ceux existants, mais de centrer les API sur les besoins déjà exprimés de longue date par ces utilisateurs en développant (en central, afin d'éviter les re-développements à l'échelle locale) les outils de traitements ou de visualisation capables de les satisfaire. Ces outils intégreront les mécanismes capables de remonter et agréger les qualifications induits par ces travaux (idéalement sans les alourdir). Actuellement, les acteurs sont parfois souvent tentés d'avoir recours à des répliques des bases nationales, sur lesquels ils peuvent agir, par exemple pour corriger des valeurs anormales. Dans ce cas, aucune de ces corrections d'anomalie n'est remontée au niveau national, et ce travail n'est donc pas capitalisé. Le projet ambitionne que tous ces efforts soient valorisés.

Enfin, personne n'est actuellement capable de mesurer statistiquement les niveaux et tendances d'utilisation des données du SIE et des bases de données locales : quelles sont les données les plus exploitées, géographiquement ? Quels sont les thématiques « à la mode » ? La remarque est la même pour les traitements sur ces données et, encore davantage, sur la qualité de ces données : quels sont les indicateurs en vogue ? Quelles sont les données les plus suspectes ? La mise en place de mécanismes permettant de bancariser les actions ou événements réalisés permettra d'apporter une réponse à ces questions.

- **Attentes des acteurs concernés (usagers, administrations)**

Ce projet de recherche et de développement trouve une application opérationnelle avec le cas particulier du Département de la Gironde, en raison de la grande densité d'informations sur les eaux souterraines et le nombre important d'acteurs largement impliqués dans ce domaine. Les cas pratiques appliqués au contexte de la Gironde permettront ainsi de tirer des enseignements sur l'opérationnalité attendue.

Les outils développés dans le projet PIC'EAU répondent au besoin des usagers des données sur l'eau de mise à disposition non seulement des données brutes mais également d'outils permettant de mieux les appréhender afin notamment de générer des données agrégées.

Les administrateurs sont aussi dans l'attente d'un mode d'utilisation de la donnée plus collaboratif, où l'utilisateur est pris en compte dans le cycle de vie de la donnée, et où ses remarques peuvent être facilement valorisées. Ce nouveau mode de consommation et de valorisation des données sur l'eau bénéficie non seulement à l'utilisateur qui s'approprie mieux les bases de données et peut ainsi en extraire beaucoup d'informations, mais aussi aux producteurs de données qui trouvent une aide à leur qualification. L'ensemble de la communauté aura aussi accès au retour d'expérience sur les utilisations faites des données permettant à terme d'adapter voire de rationaliser les suivis en fonction des attentes des utilisateurs et du grand public. La plus-value de l'expertise du territoire permettra d'augmenter la valeur ajoutée de la donnée bancarisée.

- **Conditions et facteurs clés de succès / risques principaux**

Le succès du projet PIC'EAU sera fortement et quasiment exclusivement lié à sa capacité à pleinement répondre aux besoins des acteurs territoriaux. Pour cela, un comité de suivi composé d'un panel représentatif de la diversité des acteurs locaux du monde de l'eau sera étroitement lié au projet. L'Agence de l'eau Adour-Garonne, l'Université de Bordeaux (via l'UMR EPOC⁶), le SMEGREG⁷ et l'AFB⁸ ont déjà exprimé leur intérêt à participer aux comités de suivi du projet (Cf. lettres de soutien jointes à la proposition). Les solutions proposées seront pensées pour pouvoir s'intégrer dans les outils existants et déjà utilisés par ces acteurs. Ces derniers veilleront, tout au long du déroulement du projet, à la mise en cohérence des développements techniques et thématiques avec leurs besoins opérationnels et les besoins du territoire.

D'autre part, une action importante d'animation est prévue dans le projet pour inciter les acteurs locaux à utiliser les outils développés. Ils constitueront ainsi une masse critique d'utilisateurs nécessaire à l'initiation de la mise en mouvement du « cercle vertueux » dans lequel plus nombreux sont les utilisateurs à mobiliser les données sur l'eau, plus leurs travaux et leurs expertises permettront de comprendre les données brutes par des données agrégées dans les bases de données nationales et locales, stimulant ainsi leur utilisation.

Le risque lié à la nouveauté des techniques employées, notamment pour la prise en compte des anomalies et des propositions de correction et d'enrichissement remontées par l'expertise des utilisateurs du territoire, n'est pas à négliger. Il est intégré dans l'organisation du projet en prévoyant de développer les outils de remontée des expertises locales selon plusieurs niveaux d'ambition et donc de risque : dans le cas le plus simple, les remontées sont simplement exposées tandis que dans le cas le plus ambitieux, mais aussi plus risqué, les remontées sont directement envoyées aux producteurs de données. Ce dernier schéma de remontée automatique permettrait que les amendements pour optimiser les données brutes et agrégées soient faits directement dans les bases de référence.

Le projet étant très novateur sur le plan des technologies employées mais aussi sur la place qu'il veut donner aux utilisateurs dans le cycle de gestion des données sur l'eau, il est prévu un développement en cycles courts (cf. § Principales étapes et méthodologie retenue et § Gouvernance du projet) qui permettra de réajuster à intervalles réguliers les orientations choisies en fonction des retours utilisateurs et des difficultés rencontrées.

- **Impact attendu en termes de modernisation de l'action publique**

Dans le domaine de l'environnement, la transparence de l'État est une attente forte des Français. Les données brutes sont aujourd'hui pour la plupart accessibles et c'est un premier pas, mais il faut aller au-delà pour permettre aux citoyens de s'approprier réellement les données. En ce qui concerne le domaine de l'eau, le développement de la plate-forme Hub'Eau a fait nettement progressé la mise à disposition des données en offrant des formats, structures et modalités d'accès aux données plus simples et plus faciles d'emploi. Le projet PIC'EAU permet de compléter cette démarche en offrant d'une part des outils d'aide à l'interprétation des données et d'autre part des outils de qualification des données qui permettent à l'utilisateur de mieux appréhender et donc de pouvoir s'approprier les données publiques. Le projet ambitionne de réaliser un travail de démocratisation de cette information. La transformation de ces données en interfaces graphiques, des infographies très visuelles et esthétiques, permettra un meilleur accès à leur appréhension par le citoyen in fine.

⁶ Unité Mixte de Recherche (UMR 5805) Environnements et Paléoenvironnements Océaniques et Continentaux.

⁷ Syndicat Mixte d'Etudes et de Gestion de la Ressource en Eau du département de la Gironde

⁸ Agence Française de la Biodiversité

Les outils de remontée de qualification des données par les utilisateurs permettront un changement important dans le positionnement des utilisateurs de la donnée : ils ne sont plus seulement « consommateurs » de données mais deviennent acteurs à part entière. Le territoire devient acteur des données produites, et in fine, participe à en être garant. Cette intégration importante des utilisateurs se traduit dans le projet PIC'EAU également par le rôle central du comité d'utilisateurs, ancré dans le contexte girondin, qui orientera les développements proposés.

A travers les cas d'étude prévus et présentés au paragraphe « Terrains d'expérimentation pressentis », le projet PIC'EAU permettra le développement d'outils relatifs à une meilleure connaissance des eaux souterraines en Gironde. Les cadres de développements des API de traitement comme des API de remontée de la qualification des données seront pensés pour pouvoir intégrer d'autres types de données et, si possible, d'autres systèmes d'informations environnementaux liés à l'application des politiques environnementales. Les développements réalisés dans ce projet pourront servir d'exemples et être réutilisés sur d'autres territoires et d'autres thématiques, dans un effort de mutualisation des outils dans les différentes administrations.

2. METHODE RETENUE

• Présentation des entités porteuses de projets et de la pertinence du partenariat

Le projet sera mené par un consortium composé du **Département de la Gironde**, en qualité d'entité porteuse, et du **BRGM** (directions Eau, Environnement et Ecotechnologies d'une part, et, Infrastructures et Services Numériques d'autre part), au titre de porteur opérationnel. Le projet sera par ailleurs, de par son évidente filiation avec le projet Hub'Eau et l'intérêt qu'il peut présenter, en cas de réussite, pour le SIE, suivi et rattaché sur le plan technique et scientifique au pôle de recherche et d'innovation en systèmes d'information appliquées au domaine de l'eau : **INSIDE** (<http://www.pole-inside.fr/>).

Le **Département de la Gironde**, collectivité locale de 6 500 agents, intervient sur un territoire de près de 1 505 517 habitants. Ses missions sont vastes ; il s'occupe notamment de l'action sociale, des collèges, des routes, de la culture, du développement du territoire pour répondre aux préoccupations des Girondins. Il s'appuie sur un budget d'un 1,64 milliard d'euros et est administré par 66 conseillers départementaux, répartis sur 33 cantons. Depuis plus de 60 ans, le Département de la Gironde s'investit sur la question des eaux souterraines, au travers d'un partenariat historique avec le BRGM. Ce réseau a pour vocation de surveiller la ressource en eau souterraine tant en quantité qu'en qualité. Des milliers de données sont ainsi produites et bancarisées tant dans le SIE, que dans les bases de données locales. Elles offrent de plus une aide à la décision stratégique localement et cruciale pour l'organisation de la gouvernance de l'eau souterraine, utilisée à des fins d'eau potable dans le Département. La question de la donnée, sa qualité, son utilisation et sa bonne compréhension est ainsi un enjeu de première importance pour le territoire de la Gironde. En complément, le Département de la Gironde s'est investi depuis plus de 10 ans, en tant que chef de file de projets européens et innovants sur la question de l'eau (MAC'EAU, WAT,)

Le **BRGM**, établissement public de référence dans le domaine des sciences de la terre, compte parmi ses principales missions la recherche scientifique et l'appui aux politiques publiques en particulier sur les problématiques liées aux eaux souterraines. Il opère le réseau piézométrique national ainsi que plusieurs banques nationales de données (ADES, BNPE...) fondamentales pour le SIE, et développe par ailleurs des démarches et outils innovants (librairies de traitements statistiques HYPE et ESTHER, outil QUALINET, indicateurs type IPS, ou encore les SIGES comme vecteur de diffusion de ces connaissances, voir par exemple Croiset et Lopez, 2013 et Croiset et al., 2015), afin de toujours mieux comprendre, anticiper et partager les comportements et risques liés aux eaux souterraines. Le BRGM s'est par ailleurs distingué depuis 10 ans dans le domaine des sciences de l'information et leurs

applications dans les sciences de l'environnement. Très actif depuis le début dans la définition des standards d'interopérabilité techniques et sémantiques (OGC par exemple), ainsi que dans les instances porteuses de la directive INSPIRE, il a progressivement structuré une compétence forte à tous les niveaux du traitement de l'information scientifique : conception de grands systèmes d'information, calcul scientifique et visualisation, indexation de données et de métadonnées, représentation de connaissances et même hébergement de type Cloud (en particulier, gestion de la plupart des données sur l'eau en France ou des projets comme georisques.gouv.fr pour la DGPR). Cette expertise s'est de longue date combinée avec succès aux applications dans le domaine de l'Eau, comme le souligne encore récemment le projet Hub'Eau, projet d'amorçage mené dans le cadre d'un PIA sur 2015 – 2016 et qui, depuis 2017, connaît une industrialisation soutenue par l'AFB et le BRGM ainsi qu'une intégration opérationnelle au sein du SIE.

Autre synergie gagnante entre Eau et Science de l'Information, le pôle **INSIDE**, auquel ce projet sera rattaché, ce dernier bénéficiant de ses équipes et de son expérience dans le domaine. Créé en 2013 par l'ONEMA et le BRGM, ce pôle de recherche et d'innovation dans les Systèmes d'Information Distribués pour le domaine de l'Eau (INSIDE) s'est, dès le début focalisé sur les problématiques d'interopérabilité et de systèmes répartis. Il a pour objectif principal d'appuyer le Système d'Information sur l'Eau (SIE), dans toutes ses composantes, afin de préparer et d'accompagner son évolution méthodologique et technique. Le pôle regroupe des équipes du BRGM et de l'AFB, engagées sur trois domaines d'activités principaux : l'organisation et l'architecture de systèmes d'information répartis, l'évaluation de nouvelles solutions d'acquisition et de diffusion de données, ainsi que sur la mise en place et l'orchestration de services et de traitements interopérables. Dans le cadre de ce projet, le pôle INSIDE permettra de mettre en place les services de traitements de données dans le respect des règles et standards d'interopérabilité les plus appropriés, ainsi que d'expérimenter la manière et les technologies les plus efficaces afin de structurer et organiser les remontées d'anomalies, corrections et autres enrichissements de données par les utilisateurs finaux, dans un contexte de forte répartition des systèmes. L'étude et l'évaluation d'approches type Event Sourcing et du pattern CQRS⁹ sont tout particulièrement envisagées.

Au-delà du consortium présenté, le projet recourra très vraisemblablement à des partenaires technologiques –startups ou SSII – en capacité de fournir une expertise spécifique sur les verrous technologiques du projet (Event Sourcing / CQRS, exposition des codes de traitements sous forme d'API, intégration dans les clients sélectionnés...), et ainsi accroître l'efficacité du projet. Un comité utilisateurs rassemblera également les acteurs intéressés par les retombées potentielles du projet (Agences de l'Eau, établissements publics territoriaux de bassin, universités locales, Agence Régionale de Santé, bureaux d'étude, syndicat d'eau, éditeurs de solutions informatique pour l'eau...), dont certains implantés en Gironde ont d'ores et déjà été contactés et ont montré un intérêt marqué (Agence de l'Eau Adour-Garonne, SMEGREG...).

- **Principales étapes et méthodologie retenue pour mener le projet**

PIC'EAU est un projet d'amorçage qui doit rapidement démontrer des résultats probants pour aller vers un outil opérationnel. Les partenaires privilégient donc un calendrier sur 18 mois où les développements s'opèrent de façon itérative en s'appuyant sur une méthode de type AGILE.

Outre la phase préalable de lancement du projet, comprenant la définition précise des résultats attendus par phase, la définition des tâches détaillées au sein du partenariat et la mise au point d'indicateurs de réussite, les grandes phases de réalisation du projet sont les suivantes :

1. Définition des terrains d'expérimentation afin de schématiser les fonctions et API de PIC'EAU et définir le « produit ». T0

⁹ *Command Query Responsibility Segregation*

2. Analyse des architectures techniques et mise en œuvre des environnements techniques, en s'appuyant autant que possible sur des services disponibles. T+1M
3. Après une analyse du découpage en 3 cycles de développement, réalisation d'un premier cycle de développement et mise en expérimentation. T+5M
4. Premiers tests et retours par le réseau d'utilisateurs au regard des cas d'utilisation envisagés au lancement du projet. T+6M
5. Second cycle de développement et mise en expérimentation publique T+8M
6. Troisième cycle de développement. T+11M
7. Actions d'animation et valorisation dans des projets SIE et des outils locaux (SIGORE, GIRONDE SCOPIE et gironde.fr) dans une démarche de pérennisation T + 15 M

Le principe de ce séquencement du projet est double :

- a. garantir la mise en œuvre rapide d'une plateforme démontrant le concept (5 mois après le démarrage) puis consolider progressivement la plateforme sur les retours des utilisateurs.
- b. mesurer la montée en compétences progressive des équipes au regard des technologies employées de manière à garantir une pérennisation à terme du projet d'un an et demi.

- **Composition de l'équipe (acteurs et compétences mobilisées)**

L'équipe sera organisée autour de deux niveaux d'interventions :

Une **équipe stratégique** en charge de l'orientation générale de la plateforme et de sa communication auprès des utilisateurs ainsi que des partenaires institutionnels potentiels : le Département de la Gironde et le BRGM y seront représentés respectivement par Alain Renard/Stéphanie Larbodie et Nathalie Dorfliger/Pierre Lagarde ;

Une **équipe opérationnelle** en charge d'affiner les cas d'utilisation, de développer la plateforme et d'intégrer les services mis en place dans les outils clients sélectionnés. Cette équipe sera coordonnée par le BRGM (Nolwenn Croiset/Anthony Mauclerc) et le Département de la Gironde (Stéphanie Larbodie/Pascal Romain), et composée par du personnel des deux établissements mis à disposition pour la durée du projet, ainsi que renforts externes (sous-traitance) sur des besoins spécifiques d'expertise technique ou de réalisation cadrée. Una attention particulière sera accordée à faire travailler de façon très collaborative les équipes fonctionnelles (experts métier Gironde et BRGM), de communication (Gironde principalement), de développement (BRGM avec un suivi de la Gironde) et d'exploitation (principalement BRGM).

Personne	Organisme	Compétences clés	Rôle dans le projet
Alain Renard <i>Vice-Président</i>	Département de la Gironde	Gestion des eaux souterraines	Direction du projet Valorisation du projet auprès des acteurs locaux
Stéphanie Larbodie <i>Technicienne ressource en eau</i>	Département de la Gironde	Gestion des eaux souterraines	Chef de projet Coordination des moyens humains, techniques et scientifiques du projet
Nathalie Dörfliger <i>Directrice de la Direction Eau, Environnement et Ecotechnologies (D3E)</i>	BRGM	Stratégie scientifique Ressources humaines	Coordination de la stratégie du BRGM en termes de valorisation des données sur l'eau Valorisation du projet auprès des acteurs métiers nationaux Coordination des moyens humains, techniques et scientifiques du projet
Pierre Lagarde, <i>Directeur adjoint de la Direction Infrastructures et Services Numériques (DISN)</i>	BRGM	Système d'information, en particulier dans le domaine environnemental Politique de SI	Valorisation du projet auprès des acteurs SI nationaux, à commencer par le SIE Coordination vis à vis de la stratégie de l'établissement en terme de SI
Céline Debrieu-Levrat <i>Ingénieur ressource en eau</i>	Département de la Gironde	Traitement des données sur l'eau pour l'aide à la compréhension des hydrosystèmes	Aide au développement des algorithmes de traitement de données. Suivi et expertise r
Christian Gaffory <i>Chargé de mission observatoire eau</i>	Département de la Gironde	Système d'information, en particulier dans le domaine environnemental	Aide au développement des algorithmes de traitement de données. Suivi et expertise r
Frank Uteau <i>Ingénieur environnement</i>	Département de la Gironde	Traitement des données sur l'eau pour l'aide à la compréhension sur l'environnement	Aide au développement des algorithmes de traitement de données et garant de la compatibilité technique avec les lignes directrices du SIGORE
Catherine Clément <i>Chargeée de communication</i>	Département de la Gironde	Traitement de l'information pour la communication des données	Traitement de l'information pour la communication des données lors de la phase de démocratisation de l'information dans gironde.fr
Pascal Romain <i>Chargé de mission transition numérique</i>	Département de la Gironde	Système d'information Conception de SI	Aide à la conception technique de la plateforme et compatibilité avec la politique de mise à disposition de données ouvertes du département et de sa plate-forme de mise à disposition de données en open data
Nolwenn Croiset <i>Hydrogéologue</i>	BRGM	Traitement des données sur l'eau pour l'aide à la compréhension des hydrosystèmes	Développement des algorithmes de traitement de données. Suivi et expertise
Abel Henriot <i>Hydrogéologue</i>	BRGM	Conception d'outil statistiques et de traitements SIG	Développement des algorithmes de traitement de données.
Anthony Mauclerc, <i>Responsable de l'Unité Information et Services pour l'Eau (au sein de la DISN) et Responsable technique du pole INSIDE (AFB / BRGM)</i>	BRGM	Système d'information Conception de SI	Conception technique de la plateforme et garant de la compatibilité technique avec les lignes directrices du SIE
Vincent Labbe <i>Data Scientist</i>	BRGM	Conception de SI Science de la donnée	Conception technique du SI, suivi de la réalisation et expertise sur des problématiques de science de la donnée.

Au-delà de cette équipe, le projet se dotera d'un comité d'utilisateurs, avec un premier cercle d'ores et déjà identifiable composé des différents soutiens déclarés, à étendre en début de projet (les établissements publics territoriaux de bassin (EPTB) des nappes profondes de Gironde, de la Dordogne, de l'Estuaire de la Gironde et de la Garonne, l'Agence de l'Eau Adour-Garonne, l'ARS de la Nouvelle Aquitaine, le laboratoire EPOC de l'université de Bordeaux 1, AFB, ...).

• **Terrains d'expérimentation pressentis**

Dans l'optique d'un premier développement méthodologique et technique, les terrains d'expérimentations sont limités à un nombre réduit de thématiques, se rapportant à quelques cas d'application bien identifiés. Les expérimentations prévues dans le projet s'intéressent à des problématiques qui concernent la gestion de la ressource en eau souterraine et s'appuient essentiellement sur les données disponibles (ou très prochainement disponibles) en local et au sein de la plateforme Hub'Eau.

Trois cas d'usage ont été identifiés et sélectionnés pour leur capacité à répondre à des attentes fortes du territoire, tant pour les utilisateurs que pour les producteurs des données. Ces cas d'usages permettent le développement de traitements à forte valeur ajoutée d'un point de vue thématique et génèrent un gain substantiel sur l'usage des données et leur qualification.

Il est prévu que les API développées dans ces cas d'usage puissent s'intégrer dans les outils déjà existants et utilisés par les acteurs locaux. Par ailleurs, un démonstrateur sous forme d'un plug-in QGIS sera également développé au cours du projet.

Etude de cas n°1 : Etude des variabilités spatiales et temporelles des pollutions en nitrates

Des concentrations élevées en nitrates sont la cause principale de dégradation de l'état des eaux de surface comme souterraines dans le bassin Adour-Garonne. En effet, une pression importante s'exerce dans les secteurs des grandes cultures céréalières et d'élevages intensifs. Les pollutions en nitrates se caractérisent souvent par une variabilité importante, tant temporelle au niveau d'un point de suivi, que spatiale. Les concentrations en nitrates dans les eaux sont en effet contraintes par de nombreux facteurs eux-mêmes très variables comme la pression (surplus azotés) et les conditions hydroclimatiques et hydrogéochimiques du milieu.

Ce cas d'étude vise à fournir des outils permettant de mieux comprendre les pollutions azotées dans les eaux souterraines notamment grâce à l'analyse croisée des concentrations et des principaux facteurs explicatifs. Ce cas d'étude fait suite à de nombreux travaux du BRGM sur cette thématique (voir par exemple Lopez *et al.*, 2015).

Les traitements suivants seront proposés :

- Sélection de l'ensemble des données de concentration en nitrates relatives à la zone d'étude et visualisation des évolutions spatiales et temporelles ;
- Agrégation spatiale et temporelle des concentrations par la mise en œuvre de méthodes statistiques appropriées (krigeage et tendance Kendall régional par exemple) ;
- Analyse de la variabilité spatiale des concentrations et classification des points d'eau grâce à des tests de comparaison d'échantillon ;
- Etude de l'autocorrélation des chroniques par croisement avec d'autres banques mono-thématisques, pour comprendre la variabilité temporelle et des éventuelles cyclicités ;

- Analyse croisée des concentrations en nitrates et des chroniques météorologiques et/ou piézométriques ;
- Analyse croisée des concentrations en nitrates et de l'occupation du sol (Orientation technico-économiques des exploitants (OTEX), recensement agricole (RGA)...).

Dans ce cas d'étude, il est attendu que les résultats produits par les calculs et les confrontations de différentes données amènent l'utilisateur à avoir un regard critique et d'expert sur les données. Plusieurs types d'API sont ainsi envisagés pour faire remonter les anomalies détectées :

- Remontée de valeurs anormales dans les différentes bases de données utilisées ;
- Remontée d'anomalies et d'expertises du territoire, en ce qui concerne l'appartenance d'un point d'eau à une masse d'eau ou à une entité hydrogéologique BDLISA.

Etude de cas n°2 : Analyse des données sur les niveaux d'eau dans les nappes

L'analyse des chroniques piézométriques (évolution des charges hydrauliques dans les aquifères) est essentielle pour de nombreuses applications : clé de la compréhension du fonctionnement hydrodynamique des aquifères, indicateur de la pression quantitative s'exerçant sur les eaux souterraines ou encore marqueur de l'influence du changement climatique sur la recharge des aquifères...

Ce cas d'étude vise plus particulièrement à aider les professionnels du monde de l'eau à caractériser le fonctionnement hydrodynamique d'un aquifère d'intérêt. Pour répondre à ce besoin, les API nécessaires aux traitements suivants seront développées :

- a) Sélection de l'ensemble des données de piézométrie relatives à la zone d'étude et présentation de façon innovante pour l'utilisateur, en lui proposant un ensemble d'informations complémentaires calculées systématiquement, classées ou hiérarchisées selon des critères hydrogéologiques (référentiel des entités hydrogéologiques BDLISA notamment) :
 - durée des chroniques ;
 - nombre de données manquantes ;
 - distribution statistique des données ;
 - valeurs extrêmes (outliers) ;
 - tendances détectées dans les chroniques ;
 - ruptures et dates de ruptures associées ;
 - étude de l'autocorrélation dans les séries pour qualifier l'inertie et la(les) cyclicité(s).
- b) Mise en relation des données : corrélation croisée, graphiques de mise en confrontation et de comparaison des données, carte représentant les différents critères, éventuellement avec interpolation (sous réserve que les conditions d'application soient réunies notamment d'un nombre de points suffisant), au sein d'un objet hydrogéologique (BDLISA) par exemple, ou en croisant différents objets hydrogéologiques.

Lors de la réalisation de ces traitements, notamment grâce au croisement de différentes données, les utilisateurs pourront être amenés à identifier des anomalies dans les bases de données nationales. Des API de remontée d'anomalies seront ainsi développées pour capitaliser le travail effectué par les utilisateurs :

- Remontée de valeurs anormales dans les différentes bases de données utilisées ;
- Remontée d'anomalies en ce qui concerne la localisation d'un point d'eau, son rattachement à une masse d'eau ou à une entité hydrogéologique BDLISA.

Etude de cas n°3 : Estimation de la recharge des aquifères et confrontation aux données de prélèvements

La quantification de la recharge des aquifères est indispensable à une bonne gestion de la ressource en eau souterraine. Or, si les données de pluviométrie sont en général accessibles (bien que difficilement), le ratio infiltration/ruissellement à l'échelle d'un territoire, nécessaire au calcul de la recharge, est souvent pas ou mal connu et l'estimation de la recharge devient alors souvent un point critique. Des méthodes basées sur l'étude des chroniques piézométriques et les hydrogrammes (chroniques de débits de rivières) sont une alternative très intéressante pour le calcul de la recharge.

Afin d'établir un bilan hydrique, la recharge doit être comparée aux prélèvements dans le même aquifère. Une banque nationale des prélèvements en eau (BNPE) a été créée en 2015. Cette base de données comporte aujourd'hui d'assez nombreuses lacunes, notamment dans le rattachement de l'ouvrage (souvent un compteur, servant de base au calcul de la redevance) à une ressource. Compléter et améliorer la qualité de cette base de données serait une avancée très importante.

Ce cas d'usage se propose donc de mettre à disposition des gestionnaires de l'eau des outils d'estimation de la recharge des aquifères à partir des chroniques piézométriques et des hydrogrammes (chroniques de débit de rivière) et de confronter les résultats aux données de prélèvement grâce à la mise à disposition des traitements suivants :

- Identification des périodes de recharge, identification des périodes de tarissement ;
- Estimation de la recharge par l'intermédiaire de l'étude des fluctuations des niveaux de nappe (« Water Table Fluctuation », de l'étude des chroniques de débit de cours d'eau (calcul du « Base Flow Index » et des coefficients de récession) ;
- Présentation à l'utilisateur des données de prélèvement connues pour cette zone géographique (BNPE et/ou données locales)

Les API de remontée d'anomalies prévues dans ce cas d'étude sont les suivantes :

- Remontée de valeurs anormales dans les différentes bases de données locales utilisées ;
- Remontée d'une suggestion par l'utilisateur d'un lien entre un volume prélevé et une entité hydrogéologique (BDLISA), ou un lien vers un ouvrage souterrain (identifié par un code BSS), etc... Cette solution permettrait d'aider au rapprochement des bases de données ADES et BNPE et ainsi d'initier une montée en qualité d'une table permettant de faire le lien entre les référentiels utilisés pour estimer les volumes prélevés (basés sur la notion de redevance gérée par les agences de l'eau) et les référentiels hydrogéologiques (BSSeau, Ades, BDLISA, etc.).

L'ensemble de ces études de cas doit permettre de corrélérer les données brutes produites avec l'expertise du territoire et d'en consolider aussi leurs appropriations. Au-delà de leur fiabilisation, l'objectif est que ces données puissent être comprises par les acteurs du territoire qui les produisent afin de les utiliser de façon plus conséquente pour appuyer les prises de décision locales. Cet objectif

sera affiché pour chaque étude de cas, les API développées devant à chaque fois apporter des éléments innovants dans la manière d'administrer et de gérer les ressources en eaux du territoire.

3. ACTIONS PREVUES POUR PERENNISER LA SOLUTION EN CAS DE SUCCES

- **Gouvernance du projet**

Le mode de gouvernance du projet PIC'EAU est construit selon trois niveaux :

- La gouvernance « opérationnelle » du projet sera portée selon une méthode d'inspiration AGILE, favorisant les développements en itérations courtes ("sprints" de 4 à 8 semaines) et incrémentales, de manière à éviter tout effet tunnel. La priorité est de garder une bonne faculté d'adaptation aux différentes opportunités et difficultés qui ne vont pas manquer de survenir pendant la vie du projet.
 - La gouvernance « utilisateur » sera portée par un réseau collectif d'utilisateurs et des acteurs du territoire. Ce réseau sera sollicité par l'équipe opérationnelle à chaque fin de cycle de développement, de manière à conserver une proximité forte avec la vision et les besoins du terrain. La confrontation régulière à l'outil en développement permettra d'ajuster le tir au fur et à mesure.
 - La gouvernance « stratégique » sera portée par l'équipe dirigeante, au travers de comités de pilotage réguliers, garantissant le contrôle des risques et les grandes orientations.
- **Acteurs institutionnels mobilisés pour pérenniser le projet en cas d'expérimentation réussie**

Le projet est porté par le Département de la Gironde, en tant qu'acteur territorial, et par le BRGM comme opérateur principal. Cette expérimentation constitue un projet d'innovation du SIE pour répondre à un double enjeu : l'amélioration de la réutilisation et de la qualité de l'information sur l'eau en Gironde et en France. Cette double amélioration concerne autant les données publiques nationales que les données locales : la plateforme proposée facilite leur exploitation croisée et ouvre la voie à la capitalisation des efforts de corrections menés localement sur les données nationales.

Dès que la plateforme aura démontré son efficacité technique dans les terrains d'expérimentation développés, il est fortement envisagé d'en faire un composant structurant du Système d'Information sur l'Eau, ce qui, en complément de son intégration au sein des systèmes locaux (SIGORE, SIGES, outils de menée d'étude propriétaires...), assure rapidement un niveau de pérennisation important. Des API de remontée de ces données, intégrées dans les outils du quotidien de ces utilisateurs ou dans de simples interfaces web, seraient aussi utiles et mettraient en valeur l'expertise du territoire

Au-delà, le concept visé par le projet pourrait être valorisé comme l'une des réponses génériques aux besoins de nombreux systèmes d'information distribués, comme celui du SIE mais aussi le Système d'Information Nature et Paysage (SINP) ou encore le futur Système d'Information sur les Milieux Marins (SIMM). De par ses relations avec ces deux domaines (biodiversité et milieu marin), le pôle INSIDE pourrait servir de catalyseur à cette mise en commun des visions et outils développés au travers du projet PIC'EAU.

Enfin, du point de vue technologique, la plateforme s'intègre dans les orientations du BRGM en termes de système d'information et de son infrastructure technique.

• Documentation prévue du projet en cours de réalisation

Dans le cadre du projet, la documentation sera produite par les équipes autour de trois objectifs :

- La traçabilité des spécifications et développements menés dans le cadre du projet
- La documentation pour la configuration de la plateforme
- La documentation des différentes API exposées
- Les tutoriaux et aides pour les utilisateurs de la plateforme PIC'EAU

Outre une documentation technique *ad hoc* à ce type de projet, des outils de promotion seront déployés afin de faire connaître et stimuler l'utilisation des outils développés. Des plaquettes seront éditées au lancement du projet afin d'expliquer les objectifs et le contenu du projet PIC'EAU : elles seront distribuées dans le cadre de la journée annuelle des eaux souterraines en Gironde. Le projet sera valorisé lors dans les journées d'information sur les souterraines au niveau local et au niveau national (GT Quantité et Qualité eau souterraine) afin de promouvoir le projet et fédérer les acteurs autours des solutions innovantes de diffusion des traitements des données sur l'eau développées dans PIC'EAU.

Chaque jalon majeur du projet sera validé par le comité de suivi sur la base de notes décrivant les développements réalisés pour l'atteinte des objectifs de chaque tâche. Un site internet dédié permettra à chacun de suivre les avancées du projet en cours de réalisation.

Un séminaire de restitution final à destination des acteurs locaux et nationaux sera organisé en fin de projet. Ce temps pourrait s'inscrire lors de la tenue annuelle de la journée girondine sur les eaux souterraines. Ce séminaire, au-delà de la présentation des API innovantes PIC'EAU, permettra d'engager la réflexion sur l'implémentation de ces modules dans les outils de gestion de la donnée en local et notamment tournés vers le citoyen girondin.

• RESULTATS ATTENDUS

Pour la collectivité, il est attendu une opérationnalité finale de la plateforme, qui permette de faire profiter aux citoyens de l'expertise du territoire et sa mise à disposition sur un spectre de données ciblées in fine vers le citoyen, du fait de leur fiabilisation confortée. Le projet PIC'EAU, en offrant des outils d'aide à la compréhension des données publiques sur l'eau, accompagnera le mouvement **d'ouverture et de partage des données publiques**. En effet, dans le domaine de l'eau comme dans celui de l'environnement en général, il ne suffit pas de diffuser des données brutes, trop complexes pour être interprétées par le citoyen, des outils d'aide à l'interprétation des données sont nécessaires.

D'autre part, il est attendu que les composants de remontées d'anomalies concourent à la montée en qualité des bases de données nationales et locales. Cette mise en qualité, réalisée de manière indirecte par les utilisateurs au travers de leurs missions quotidiennes (missions par ailleurs facilitées par l'utilisation des traitements mis à disposition), peut éviter, au moins en partie, de très coûteuses opérations de nettoyage de bases de données. A titre d'exemple, le rapprochement des données Prélèvement (cas d'utilisation n°3 prévu au présent projet) issues d'une part des Agences de l'Eau, et d'autre part des DDT, avait été envisagé dans le cadre du projet BNPE en 2012 ; le travail à mener initialement, envisagé en central, débouchait sur le « dédoublonnement » de plusieurs centaines de

milliers d'ouvrages (entre 300 000 et 400 000 environ à l'échelle nationale), et posait le problème de l'implication massive (possibilités de rapprochement automatique très limitées) d'agents par ailleurs surqualifiés ; sur la base d'une centaine d'ouvrages traités par jour, l'effort initial avait été dimensionné à hauteur de 5 millions d'euros, ce qui avait mis un point d'arrêt à cette ambition nationale. En parallèle, tout spécifiquement en Gironde, ces mêmes travaux ont été menés localement et a permis d'en améliorer la qualité. Le succès de la plateforme PIC'EAU, appliqué au cas girondin constituerait un formidable levier pour pérenniser ce travail, qui ne comporte pas à ce jour de routine. Enfin, la plateforme de traitements en ligne a pour objectif de fournir un point d'accès et de promotion des traitements issus de nombreuses actions de R&D, afin de proposer aux utilisateurs des traitements de référence clairement identifiables et compréhensibles, jusque dans leurs détails d'implémentation. En complément de cette mise à disposition en ligne, la plupart des traitements ainsi diffusés a vocation à être identifiés par des éléments descriptifs type fiches de métadonnées (dans les catalogues locaux et nationaux prévus à cet effet) et gagneraient in fine à ouvrir leur code (sur les nombreuses plateformes de partage type GitHub) afin de finir le cercle vertueux visé par le projet en améliorant également par l'usage la qualité du code de traitement lui-même.