



PREMIER MINISTRE

Rapport de la mission Etalab sur les conditions d'ouverture du système Admission Post-Bac

Avril 2017

etalab gouv.fr

Synthèse

S'appuyant sur l'expérience acquise par une équipe de l'Institut national polytechnique de Toulouse dans la gestion commune de l'ensemble des concours d'entrée dans les écoles d'ingénieurs, le portail Admission Post-Bac (APB) est devenu depuis 2009 le point de passage obligé pour tous les bacheliers envisageant de poursuivre des études supérieures. Intégrant la plupart des formations supérieures reconnues par l'Etat, il permet de recueillir et de traiter les vœux de près de 850 000 candidats tout en dématérialisant leurs dossiers d'inscription.

Le système permet notamment aux futurs étudiants de recenser leurs vœux incluant:

- des filières sélectives (CPGE, DUT, STS), pour lesquelles l'établissement demandé dispose de la décision d'admission des demandeurs ;
- des filières non-sélectives ;
- certaines filières non-sélectives mais dans lesquelles les capacités d'accueil sont inférieures au nombre de demandes (PACES, STAPS...). Pour celles-ci, le système implémente des règles de priorisation, voire un tirage au sort entre les candidats.

Cette coexistence de modalités d'admission très différentes au sein d'un même système a engendré chez certains utilisateurs (lycéens, parents) des incompréhensions, qui ont parfois conduit à présenter le portail APB comme un système opaque déterminant l'affectation des futurs étudiants. Ces derniers témoignent par ailleurs de difficultés à renseigner de façon éclairée leurs vœux hiérarchisés, et adoptent face au portail des « stratégies » de choix plus ou moins informées.

L'association *Droits des Lycéens* s'est saisie de cette question en mars 2016, et a demandé au Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche (MENESR) des informations sur les critères employés et le fondement légal de leur usage. Suite à une première diffusion de documents d'explication par le Ministère, et à une décision de la Commission d'Accès aux Documents Administratifs (CADA) en juin 2016 favorable à l'association, le MENESR a communiqué en octobre le code source d'un des algorithmes concernés, dans des modalités jugées sous-optimales (code source non commenté, transmis sur support papier...).

Dans ce contexte, le Secrétaire d'Etat chargé de l'Enseignement supérieur et de la Recherche a mandaté la mission Etalab pour étudier les conditions d'ouverture

d'APB (code source et données), le cadre technique et juridique de cette ouverture, ainsi que l'organisation des échanges entre l'administration et les réutilisateurs.

Etalab a réalisé cette étude en lien avec les services du MENESR et le service APB de l'INP Toulouse qui assure le développement informatique du portail, avec l'Agence nationale de la sécurité des systèmes d'information (ANSSI), et avec l'appui d'autres départements de la DINSIC pour les questions techniques et juridiques.

▪ **A quels objectifs répond l'ouverture du système APB ?**

L'ouverture du code source d'APB est un prérequis pour améliorer la compréhension et l'appréhension du système par ses utilisateurs, et contribuer ainsi à lutter contre le sentiment de défiance qui s'est instauré.

Mais ce n'est qu'une première étape, qui devrait permettre de **mobiliser une communauté** (dont les prémices ont été démontrées lors de la première communication du code source), **dont l'intelligence collective permettra progressivement de proposer des améliorations du système.**

Cette communauté de développeurs, chercheurs, spécialistes des données jouera, en complément du nécessaire travail d'explication et de communication par le MENESR, un rôle de **médiateur**, explicitant pour le grand public les enjeux des algorithmes, des données et des paramètres utilisés.

A terme, de **nouveaux services** pourraient être développés pour les bacheliers ou les acteurs de l'éducation sur la base des données qui seront ouvertes (par exemple des personnalisations de conseils d'orientation en fonction du profil détaillé du candidat et des statistiques récentes).

L'ouverture du code source peut également être un **vecteur d'amélioration du système lui-même**, la communauté étant à même de pointer d'éventuels besoins d'améliorations ou corrections. Un objectif de long terme pourrait être d'ouvrir le code source à la contribution de développeurs externes à l'équipe projet actuelle, afin qu'ils puissent proposer directement des modifications dans le code.

Enfin, de la même façon que l'ouverture du code du calculateur des impôts et l'émergence d'OpenFisca permettent de simuler l'impact concret d'une évolution de la fiscalité, l'ouverture d'APB (code et données) permettrait aux spécialistes de l'éducation et de l'orientation, y compris au sein des ministères, d'étudier les avantages et inconvénients de changements envisagés dans les règles d'orientation, renforçant encore l'intelligence de la décision publique.

▪ **Quel est le cadre (juridique, sécurité) applicable à une telle ouverture ?**

Le cadre juridique de l'ouverture d'un système est globalement défini dans le **Code des relations entre le public et l'administration ("CRPA")**, et illustré par la doctrine de la Commission d'accès aux documents administratifs, à trois titres :

- **principe d'accès aux documents administratifs, application au code source d'un logiciel** : un code source produit par une administration dans le cadre de sa mission de service public est un document administratif au sens de l'article L300-2 du CRPA, et est donc communicable à toute personne qui en fait la demande (article L311-2). Depuis le 7 avril 2017 et la prise d'effet des nouvelles dispositions de l'article L312-1-1 de la loi pour une République numérique, les documents administratifs existant sous forme électronique et communiqués doivent être publiés en ligne, ainsi que leurs mises à jour. De plus, un code source ainsi publié devient librement réutilisable (article L321-1), sous réserve des droits de propriété intellectuelle éventuellement détenus par des tiers extérieurs à l'administration ;
- **principe d'accès aux documents administratifs, application aux données collectées et générées par le système cible** : Conformément à l'article L312-1-2 du CRPA, les documents comportant des données à caractère personnel ne peuvent être rendus publics que si des dispositions législatives ou réglementaires le prévoient, si les personnes concernées ont donné leur accord ou si les informations ont fait l'objet d'un traitement permettant de rendre impossible l'identification de ces personnes. A partir d'octobre 2018 au plus tard, les données non-réidentifiantes dont la publication présente un intérêt social devront être diffusées par défaut (article L312-1-1).
- **principe de transparence des algorithmes** : les administrations ont l'obligation d'avertir leurs usagers (article L311-3-1) lorsqu'une décision individuelle est prise sur le fondement d'un traitement algorithmique, et doivent publier les règles définissant les principaux traitements algorithmiques utilisés pour de telles décisions (article L312-1-3).

Concernant la sécurité, le CRPA prévoit que ne sont pas communicables les documents administratifs si cette communication porte atteinte, notamment (article L311-5), à la sûreté de l'Etat, à la sécurité publique, à la sécurité des personnes ou à la sécurité des systèmes d'information des administrations.

▪ **Comment ce cadre s'applique-t-il au cas particulier d'APB ?**

APB est un système complexe, composé de plusieurs modules et interfaces. Un travail a donc dû être effectué avec les services du ministère et les équipes de l'INP afin d'identifier un périmètre de code source cohérent qui puisse répondre au mieux aux objectifs de lisibilité (possibilité d'exécution complète de l'algorithme) et de limitation aux composants nécessaires ne présentant pas de risque de sécurité.

Ce périmètre comprend 35 fichiers répartis en deux blocs :

- algorithme de classement (14 fichiers PL/SQL de traitements de bases de données permettant de réaliser le classement des filières non sélectives, et quelques fonctions support) ;
- algorithme d'affectation (21 fichiers Java permettant de réaliser les traitements d'affectation des candidats).

L'analyse juridique a été menée sur cette base, sur laquelle il apparaît que :

- Le code source d'APB est produit conjointement par une administration et un établissement public à caractère scientifique, culturel et professionnel (EPCSCP), dans le cadre d'une mission de service public. La publication de ce code source d'APB et de ses mises à jour est donc possible, puisqu'il s'agit clairement d'un document administratif et qu'aucun tiers extérieur à l'administration détenant des droits de propriété intellectuelle n'a été identifié;
- Sur ce périmètre, l'étude relative à la sécurité n'a pas mis en évidence d'atteinte parmi celles mentionnées au 2° du L311-5 ;
- L'administration a l'obligation de prévoir une mention explicite informant l'intéressé de l'existence d'algorithmes participant à la décision individuelle, et de publier les règles définissant ces algorithmes ;
- Les données traitées par APB sont pour certaines des données à caractère personnel (candidats, vœux) et pour d'autres ne présentent pas de risque de ré-identification (établissements, offres de formation). Dans le premier cas, un travail d'anonymisation est nécessaire avant publication en *open data*, dans l'autre de nombreux jeux de données ne présentant pas de risque de ré-identification peuvent d'ores et déjà être publiés.

▪ **Quel plan d'actions pour l'ouverture d'APB ?**

La mission formule, pour l'ouverture d'APB, les recommandations suivantes :

Recommandation n°1 Publier, sans attendre une nouvelle demande de communication, le code source de la dernière version d'APB et les données non-réidentifiantes associées (cf. <i>recommandation n°6</i>). Accompagner cette ouverture d'un événement avec la communauté (cf. <i>recommandation n°9</i>).
Recommandation n°2 Mettre à disposition le code source sur un système de gestion de version décentralisé (par exemple git) au moyen d'une plateforme ouverte au public. La documentation existante doit être partagée.
Recommandation n°3 Soumettre le code source à une licence de réutilisation. La mission Etalab invite le ministère à choisir une licence dite "permissive" permettant toute forme de réutilisation.
Recommandation n°4 Pour simplifier et garantir l'efficacité de la réutilisation, faire signer par chacun des agents susceptibles de détenir des droits de propriété intellectuelle un "Developer's Certificate of Origin".
Recommandation n°5 Au-delà de l'ouverture du code, prévoir à terme le passage d'Admission Post-Bac dans un modèle de développement ouvert aux contributions, et étudier les conditions de l'ouverture de la gouvernance du système à l'ensemble des parties prenantes.
Recommandation n°6 Ouvrir les données non-réidentifiantes concernant les candidats, les établissements, les vœux et les propositions, et en premier lieu celles qui permettent une meilleure information des candidats et des familles.
Recommandation n°7 Travailler sur l'anonymisation des données issues d'Admission Post-Bac, en s'appuyant sur l'expertise de l'Administrateur général des données et le retour d'expérience d'autres administrations (dont l'Assurance Maladie).
Recommandation n°8 Créer un ensemble de jeux de données fictives permettant de simuler et de "rejouer" le système à partir du code source qui sera publié.

Recommandation n°9

Organiser, avec l'appui d'Etalab et au plus tard à la rentrée 2017, un événement de type hackathon à l'occasion de la publication du code source, avec pour objectif de faciliter la réutilisation du code et des données.

Recommandation n°10

S'engager dans une démarche ouverte et collaborative autour de l'amélioration d'Admission Post-Bac, notamment avec les acteurs de la recherche (*open science*).

▪ **Quels sont les enseignements à tirer du cas emblématique APB ?**

Le système Admission Post-Bac constitue à plusieurs titres un cas d'école, dont l'Etat devrait tirer des enseignements utiles dans les années à venir.

L'ampleur du public concerné (chaque année plus que la totalité d'une classe d'âge, en comptant les réinscriptions) et l'importance de l'enjeu pour les intéressés (qui tendent toutefois à sous-estimer le rôle joué par les responsables de filières sélectives) imposent à l'Etat des exigences sur ce système mettant en jeu des algorithmes : il doit opérer sur des bases transparentes et compréhensibles par tous, et rendre des comptes à ceux qui sont concernés par ces traitements.

L'ouverture du système (code et données) est, plus qu'une obligation légale, une façon de répondre à ces exigences, mais elle n'est pas réalisable sans effort de la part de l'administration, notamment s'agissant d'un système hérité, qui n'a pas été initialement conçu pour être ouvert.

Des cas similaires ne manqueront pas de se présenter à l'avenir, et ces situations devraient, autant que possible, être anticipées. Il faudra notamment doter les administrations d'une culture de l'ouverture (*open data*, *open source*) et des compétences associées. Par ailleurs, les futurs systèmes d'information devront, dès leur conception, intégrer cette dimension d'ouverture et de redevabilité. Enfin, il conviendra de faire converger l'approche juridique, qui historiquement fait des textes et des réglementations la base de l'action publique, et le code informatique qui prend aujourd'hui une part de plus en plus importante dans la mise en oeuvre de l'action publique.

Synthèse	3
1. Rappel de la lettre de mission et du contexte d'APB	10
2. Les bénéfices attendus de l'ouverture d'APB	14
2.1 L'ouverture du système APB, prérequis pour sa meilleure compréhension par l'ensemble des acteurs du système	14
2.2 La création de nouveaux services pour accompagner les familles et les étudiants	15
2.3. L'amélioration à terme du système APB lui-même	16
3. Analyse technique et juridique	17
3.1 La définition du périmètre de l'ouverture du code	17
3.2 L'analyse technique du code source	20
3.3 L'analyse juridique	20
i. Modalités de conception et statut juridique d'Admission Post-Bac	21
ii. Communication et réutilisation du code source	21
iii. Communication des données	23
iv. Obligations relatives aux traitements algorithmiques lorsqu'ils fondent une décision individuelle	23
3.4 L'analyse des risques de sécurité	24
4. Recommandations et plan d'actions sur l'ouverture du système APB	25
4.1 Recommandations pour l'ouverture du code source	25
4.2 Recommandations pour l'ouverture des données d'APB	27
4.3 Recommandations pour l'animation et la mobilisation des écosystèmes	30
5 - Conclusion: APB, cas d'école	32
Glossaire	34
Annexe 1 - Lettre de mission	35
Annexe 2 - Liste des personnes et organismes auditionnés	37
Annexe 3 - Proposition de DCO developper certificate of origin	38
Annexe 4 - Cinématique de l'algorithme	39

1. Rappel de la lettre de mission et du contexte d'APB

Le Secrétaire d'Etat chargé de l'Enseignement supérieur et de la Recherche a saisi la mission Etalab, par courrier en date du 1er février 2017, d'une mission d'étude des conditions d'ouverture du système Admission Post-Bac¹.

Dans ce cadre, la mission a mené une série d'entretiens et d'analyses, notamment auprès des équipes techniques qui ont développé le système Admission Post-Bac. Ce document vise à présenter ces travaux, ainsi qu'à formuler un ensemble de recommandations sur l'ouverture du système APB.

Depuis 2009 et son déploiement au niveau national, le portail Admission Post-Bac (APB) est devenu le passage obligé pour près de 850 000 candidats souhaitant postuler dans la plupart des formations post-bac publiques ou privées reconnues par l'Etat.

Le logiciel APB est un système de pré-inscription, il permet de recueillir et de traiter les vœux des candidats et de dématérialiser les dossiers d'inscription. Par son ampleur, APB constitue aujourd'hui une source de données unique sur les vœux et les parcours d'orientation des étudiants.

Comparativement à la situation qui prévalait avant 2009, le système permet aux bacheliers d'accéder à une plus grande offre de formations post-bac et dispense le plus souvent de réaliser un dossier papier ou de réaliser une démarche de pré-inscription en se rendant physiquement dans les établissements d'enseignement supérieur. Du côté des filières sélectives (classes préparatoires, DUT, STS), APB est également un outil permettant d'aider à la décision, cette dernière relevant toutefois de la responsabilité des filières elles-mêmes.

La situation semble aujourd'hui paradoxale: d'un côté le système APB offre des gains réels pour l'ensemble des acteurs, et de l'autre il souffre d'un déficit de compréhension de la part des familles, voire parfois d'une défiance de la part de certains acteurs du système.

¹ La lettre de mission figure à l'annexe 1 de ce document

Une partie des difficultés provient de la genèse même d'Admission Post-Bac et des conditions de sa généralisation progressive.

S'appuyant sur l'expérience acquise par l'INP de Toulouse pour la gestion commune de l'ensemble des concours d'entrée dans les écoles d'ingénieurs, le portail admission post-bac (APB), initialement conçu pour les classes préparatoires aux grandes écoles, a fait l'objet d'une généralisation à la quasi-totalité des formations de première année de l'enseignement supérieur français en 2009.

APB, qui avait été initialement conçu pour traiter des filières sélectives (où le niveau académique d'un candidat et ses motivations constituent les principaux critères de sélection) gère aujourd'hui, via le même outil, des filières non sélectives (qui ne peuvent donc pas, par définition, utiliser de critères de sélection). Cette coexistence, au sein d'un même outil, de deux univers de règles du jeu si différentes (voire antagonistes) crée une certaine confusion dans l'esprit des utilisateurs du système, et en premier lieu les lycéens et les familles.

Ainsi, à son corps défendant, APB est aujourd'hui perçu par une partie de ses utilisateurs comme un outil qui décide de leur affectation, passant sous silence par exemple le rôle des responsables de filière sélective. Ce n'est par exemple pas APB qui accorde ou refuse une place en classe préparatoire à un étudiant.

Le traitement des licences des filières en tension a créé une difficulté supplémentaire. Pour certaines d'entre elles, par exemple PACES ou STAPS, le nombre de demandes excède sensiblement les capacités d'accueil. Le Code de l'Éducation a bien introduit des critères pour prioriser les demandes, mais ils ne sont pas suffisants quand le déséquilibre entre l'offre et la demande est trop important. APB pratique alors un tirage aléatoire entre ces candidats. Bien qu'il ne concerne qu'une minorité de candidats, ce tirage au sort est souvent mal compris et est également parfois difficilement accepté.

Cette procédure a attiré l'attention de l'association *Droits des Lycéens*, qui a saisi en mai 2016 la Commission d'accès aux documents administratifs (CADA) pour obtenir communication de la partie du code source d'APB concernant le traitement des filières non-sélectives sous tension. La Commission d'accès aux documents administratifs a donné raison à l'association et le ministère a publié en juin les règles générales de fonctionnement du système².

Or, de doctrine constante de la CADA, et aux termes de la Loi pour une République numérique, la transparence sur les règles de fonctionnement d'un système ne saurait être considérée comme suffisante. Le code source lui-même représente un document administratif à part entière et doit être communiqué s'il n'est pas protégé par l'un des secrets prévus par la loi.

² <http://www.education.gouv.fr/cid102649/apb-cloture-de-la-phase-de-formulation-et-de-classement-des-voeux.html>
<http://www.youscribe.com/BookReader/Index/2734749?documentId=2913196>

Le Ministère de l'Enseignement supérieur et de la Recherche a bien transmis les éléments demandés, mais dans des conditions (fourniture de 20 pages de code non documenté sous format papier) qui ont donné l'impression qu'il répondait à contrecœur à la demande de transparence.

Parallèlement à la demande de l'association *Droits des Lycéens*, le système APB est devenu un objet de débat, relayé par les médias spécialisés et généralistes.

Les termes du débat sont maintenant connus: APB apparaît comme un système complexe et difficilement compris, ce qui laisse place à de nombreuses spéculations sur les "stratégies" à mettre en place. A ce constat s'ajoute une demande, de plus en plus forte, de transparence et de redevabilité quant aux fondements des décisions publiques dans la société, à laquelle les engagements de la France au sein du Partenariat pour un gouvernement ouvert, puis les dispositions de la loi pour une République numérique, semblent faire droit.

Il n'est guère étonnant qu'Admission Post-Bac soit l'un des systèmes sur lesquels cette pression s'avère la plus forte: l'outil est perçu, plutôt à tort, comme le lieu où se décide l'orientation d'un étudiant et, par extension son avenir personnel et professionnel.

L'ouverture du code source d'Admission Post-Bac, ainsi que des principales données collectées et produites par le système, s'inscrit dans ce contexte, dans lequel la mission Etalab apparaît comme un partenaire afin d'accompagner cette ouverture.

Service interministériel rattaché au Premier ministre, la **mission Etalab**³, au sein du SGMAP, est en charge de l'ouverture des données publiques ("open data") et de la promotion d'une action publique plus transparente et collaborative ("open government" ou "gouvernement ouvert").

Dans le cadre de cette politique visant à mettre à disposition un maximum de ressources produites par les administrations à des fins de transparence, de collaboration et d'innovation, et outre son expertise en matière de données, Etalab a acquis une expérience en matière d'ouverture de plusieurs codes sources de logiciels et d'accompagnement des administrations dans cette démarche.

La mission a notamment travaillé en 2016 avec la Direction générale des Finances Publiques sur l'ouverture du code source de la calculatrice des impôts (voir encadré), ou encore avec la Direction de l'information légale et administrative sur l'ouverture de plusieurs simulateurs⁴.

³ <https://www.etalab.gouv.fr/qui-sommes-nous>

⁴ <http://www.dila.premier-ministre.gouv.fr/actualites/actualites/hackathon-codegouv-ouvrir-les-codes-sources-des-simulateurs-de-l-etat>

L'ouverture du code source du calculateur des impôts et le hackathon #CodeImpot

Les 1er et 2 avril 2016, Etalab a organisé, avec la Direction générale des Finances Publiques (DGFIP) un hackathon baptisé #CodeImpot sur l'ouverture du code source du calculateur des impôts⁵.

Préalablement à cet événement, Etalab a d'abord collaboré avec les services de la DGFIP afin de préparer l'ouverture du code source de ce logiciel calculateur de l'impôt sur le revenu, qui a été rendu disponible en libre accès le 1er avril.

Grâce à cette ouverture, chacun peut ainsi désormais consulter les rouages du logiciel qui permet de fixer le montant de l'impôt des particuliers.

Cette mise à disposition du code source répond à une volonté de transparence, de pédagogie et de modernisation de l'administration fiscale. Elle s'inscrit également dans l'intention du législateur, l'article 2 de la loi pour une République numérique ayant introduit la mention des codes source des logiciels dans la liste des documents administratifs soumis au droit d'accès.

Lors du hackathon, développeurs, économistes, fiscalistes ou simples curieux étaient invités à se réunir pour créer des projets innovants basés sur le code source du calculateur de l'impôt sur le revenu.

Le code source a ainsi pu être traduit vers de nouveaux langages et donner naissance à de nouvelles interfaces destinées au grand public : applications, pages web, tableaux... Les travaux ont également fait émerger des pistes d'optimisations du temps de calcul du simulateur, sources d'amélioration pour l'administration elle-même. Ce code source ouvert peut également servir à des économistes pour tester leurs cas types, évaluer les politiques publiques ou encore réaliser des simulations de réformes fiscales.

Pour la présente étude, la mission Etalab a collaboré avec le pôle "Gouvernance et maîtrise des risques" de la DINSIC, qui comprend différentes compétences en matière d'audit de grands systèmes, de sécurité logicielle et d'audit, avec l'incubateur des services numériques (beta.gouv.fr), ainsi qu'avec l'Agence nationale de la sécurité des systèmes d'information (ANSSI).

Les chapitres suivants précisent les bénéfices attendus de cette démarche d'ouverture, détaillent l'analyse technique et juridique menée en lien avec les services du Ministère quant au périmètre du code source du logiciel à ouvrir, et formulent une liste d'actions à engager.

⁵ <http://www.economie.gouv.fr/hackathon-calculateur-impots>

2. Les bénéfices attendus de l'ouverture d'APB

2.1 L'ouverture du système APB, prérequis pour sa meilleure compréhension par l'ensemble des acteurs du système

L'opacité du système Admission Post-Bac génère un sentiment d'incompréhension, voire de défiance de la part de ses utilisateurs, et en premier lieu des étudiants et des familles.

L'ouverture est ainsi un prérequis pour une meilleure compréhension et appréhension du système par ses utilisateurs, notamment sur la portée et le périmètre de l'outil. La coexistence de deux univers aux règles différentes (filières sélectives et non-sélectives) pourrait être rendue visible et vérifiable par la mise en ligne du code source. De même, cette ouverture permettrait d'identifier l'ensemble des nombreux intervenants qui interviennent autour d'APB et leurs responsabilités respectives.

Si cette ouverture est un prérequis, elle représente une condition *nécessaire* mais cependant *non suffisante* de la compréhension par tous du système APB.

Le code source n'est pas, par nature, immédiatement compréhensible par le grand public. Des tiers compétents (développeurs, chercheurs...) devront s'en emparer afin de le rendre plus intelligible et assurer ainsi un travail de médiation. La communication, en octobre 2016, d'une partie du code source par le Ministère, sous la forme de document imprimé, a d'ailleurs permis d'illustrer la capacité de mobilisation de ces communautés. En quelques heures des développeurs l'ont retranscrit en langage informatique et publié sur une plateforme de partage de code (*GitHub*)⁶. En quelques jours, ils ont pu réaliser une première analyse du code et interpeller le Ministère sur les traitements applicables aux candidats inscrits dans des lycées français à l'étranger.

Une ouverture plus large du code source et des données, réalisée dans des conditions qui correspondent à l'état de l'art (ce qui n'a pas été le cas de la première transmission du code) doit permettre de mobiliser ces communautés en partie déjà constituées. La médiation autour de ce code fera appel à des profils et

⁶ Voir notamment https://github.com/jeantil/admission_post_bac et <https://github.com/arnaudriegert/comprendre-apb>

des compétences multiples: informaticiens, mathématiciens, chercheurs spécialistes de l'éducation et de l'orientation, sociologues...

Il faut toutefois souligner que la publication du code source n'exonère pas le Ministère d'un effort de transparence, d'explication et de pédagogie non seulement quant à la procédure elle-même, mais aussi quant aux finalités poursuivies par Admission Post-bac et aux choix politiques qui sous-tendent le fonctionnement de la plateforme (par exemple le recours à un tirage aléatoire pour les licences non sélectives sous tension).

2.2 La création de nouveaux services pour accompagner les familles et les étudiants

Depuis sa généralisation en 2009, près de 5 millions d'étudiants ont eu recours à l'outil APB, qui est devenu de fait un observatoire de l'orientation tout à fait unique. Les données collectées et produites par Admission Post-Bac, ainsi que le code source pourront permettre de créer de nouveaux services pour les familles et les étudiants. Aujourd'hui de très nombreux acteurs fournissent des informations sur la procédure APB et plus généralement sur l'orientation : le Ministère, via les guides APB et le portail lui-même, les services académiques d'information et d'orientation (SAIO), l'ONISEP mais aussi des acteurs associatifs et privés, entreprises de médias notamment (Le Monde, le groupe L'Etudiant).

L'ouverture des données doit tout d'abord permettre de renforcer l'information des candidats et des familles. La plupart des informations aujourd'hui disponibles concernent soit la procédure elle-même (les dates, le calendrier), soit une information plus généraliste sur les métiers et les filières de formation pour y parvenir.

A l'aide des données, il sera possible de passer d'une information généraliste ("comment intégrer un IUT gestion des entreprises et des administrations") à une information factuelle et surtout directement utilisable par les candidats, par exemple: "En 2016, l'I.U.T de Roanne qui affichait une capacité d'accueil de 100 places dans sa filière gestion des entreprises et des administrations a reçu 501 demandes dont 45 exprimées en premier vœu. 73% des candidats ayant sélectionné cette filière en 1er vœu ont été admis. Parmi l'ensemble des admis, 35% provenaient de la filière ES et 19% de la filière S". Les candidats devraient pouvoir accéder à ces données en amont de la procédure.

Au-delà de cette dimension d'information, les données d'Admission Post-bac doivent permettre d'offrir de nouveaux services personnalisés. On peut notamment imaginer un service qui compare le profil d'un candidat à des profils similaires et lui indique ainsi des filières dans lesquelles il pourrait candidater à proximité de son

domicile⁷. Tout ce qui peut permettre d'éclairer les options et les choix des candidats représente une avancée.

2.3. L'amélioration à terme du système APB lui-même

L'ouverture du code source et des données d'Admission Post-Bac peut permettre d'améliorer le système lui-même.

Cette amélioration concerne tout d'abord le code du programme: les développeurs du service APB de l'INP Toulouse pourraient bénéficier des regards extérieurs sur leurs développements, et accepter à terme des améliorations proposées par des tiers. Il faut toutefois souligner que la taille réduite de l'équipe en charge du développement ne permet pas, dans un premier temps, d'envisager de basculer le système APB dans un modèle de développement open source qui puisse accepter aisément les contributions extérieures (« commit »).

Les discussions menées avec les équipes de développements de l'INP Toulouse ont permis de présenter certaines meilleures pratiques généralement de mise dans les communautés open-source. Des questions sur les licences, les outils de développement ou les processus de code review ont pu être abordées.

De manière plus générale, ouvrir le système APB doit permettre à des tiers, et notamment à des chercheurs, d'imaginer et de tester d'autres options pour le système Admission Post-Bac. Que se passerait-il par exemple si on changeait les règles concernant les licences non-sélectives sous tension ? Quel impact pourrait avoir la limitation du nombre de choix possibles ? Quelles logiques alternatives d'appariement (au cœur de la plateforme) peut-on imaginer et quelles en seraient les conséquences, par exemple en termes de mobilité étudiante ?

L'ouverture permettra alors non seulement d'objectiver le débat sur APB, mais aussi de discuter les choix qui ont précédé à la mise en œuvre de l'outil.

Le Ministère doit considérer cette discussion comme une opportunité pour améliorer le système en place, et non comme une remise en cause de son autorité. Comme le soulignait la lettre de mission qui nous a été adressée (reproduite en annexe 1), APB est un cas d'école qui illustre aussi *"la complexité à transiter de façon rapide d'un système hérité et historiquement fermé à un système ouvert, car ce sont aussi bien des questions juridiques, techniques que politiques qui se posent alors"*.

⁷ Cf. par exemple le site inspire-orientation.org proposé par Frateli Labs.

3. Analyse technique et juridique

APB est un système complexe comportant de nombreux modules et interfaces, son code source est mis à jour tous les ans afin de prendre en compte les modifications souhaitées par le ministère (prise en compte de la situation familiale pour le traitement des licences sous-tension en 2017 par exemple) ou l'évolution des filières.

L'étude des conditions d'ouverture du code source devait donc démarrer par une première analyse de périmètre, celle-ci a été effectuée sur la version 2016 du code, qui nous a été communiquée par les équipes de l'INP courant mars.

3.1 La définition du périmètre de l'ouverture du code

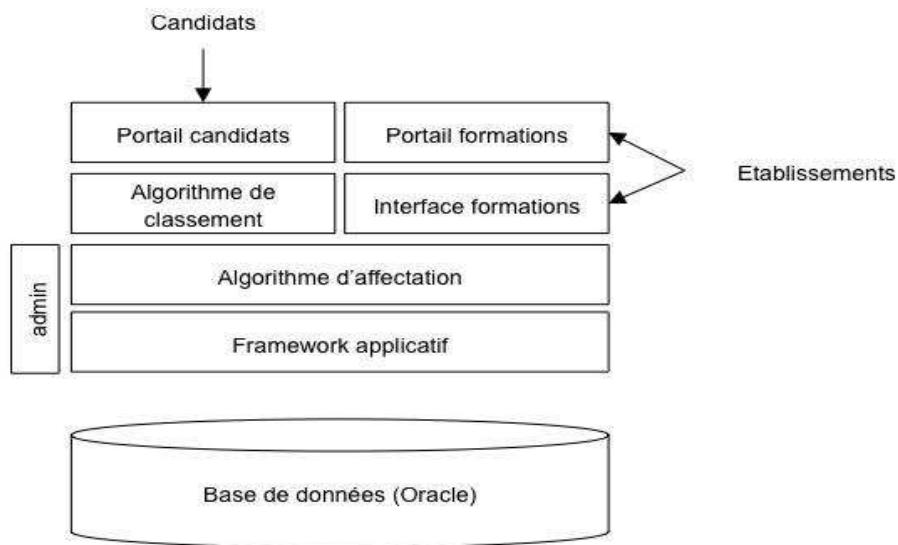
La première communication du code source d'APB, en réponse à la demande de l'association *Droits des Lycéens* concernait uniquement l'algorithme de classement des filières L1 non sélectives sous tension, ce qui ne représente qu'une partie de la plateforme.

Le schéma ci-dessous donne une vision des différentes composantes du système APB et des algorithmes impliqués.

Après l'enregistrement des vœux, deux étapes doivent être distinguées:

- Le classement des candidats :
 - Pour les filières sélectives, cette étape est à la charge des établissements qui sont autonomes pour décider du classement des candidatures reçues via APB. Le système envoie pour chaque formation la liste des candidats et l'établissement doit retourner cette liste classée.
 - Pour les filières non-sélectives, c'est le système APB qui effectue ce classement uniquement pour les filières sous tension à l'aide d'un algorithme qui a fait l'objet d'une publication partielle.
- L'affectation des candidats : Une fois les classements terminés, les candidats sont affectés dans l'ensemble des formations sélectives et non-sélectives en fonction de données d'appel de chaque établissement et des vœux des candidats selon un algorithme dit des "mariages stables". La meilleure proposition est faite au candidat qui peut la confirmer définitivement, mettre une option, ou la refuser. Cette étape est jouée trois fois (les phases) pour stabiliser les résultats en fonction des réponses des candidats.

- Schéma du système APB -



Les développements initiaux d'APB, effectués pour les seuls besoins métier, n'avaient jamais pris en compte une possible ouverture du code source.

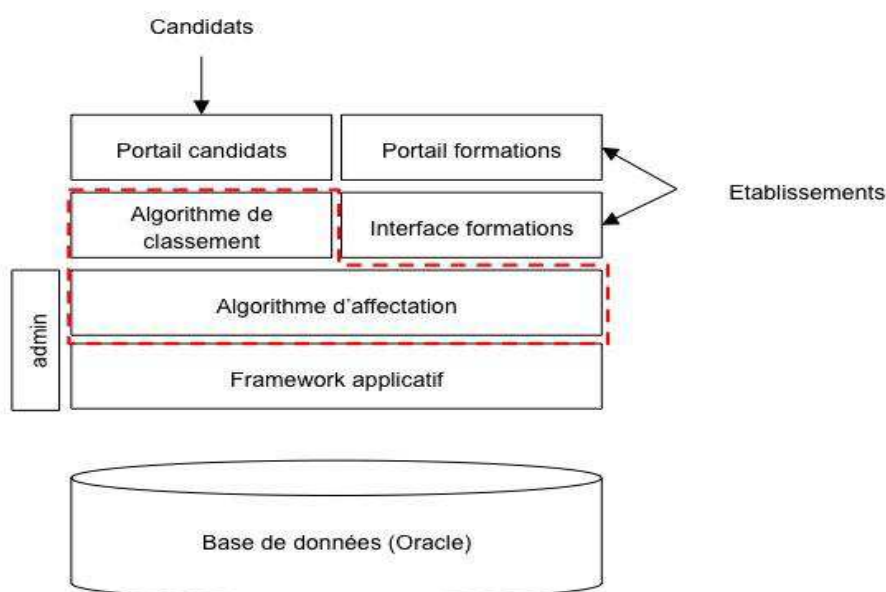
De plus, l'extension progressive du périmètre d'application d'APB, depuis la gestion de la filière sélective du concours commun polytechnique à la généralisation pour toutes les filières post-bac, a fortement complexifié le code d'origine.

Un travail d'identification du périmètre d'ouverture a donc été nécessaire pour déterminer un ensemble cohérent qui réponde aux attentes de transparence et permette une compréhension du système. Les objectifs en étaient les suivants:

- Couvrir le périmètre complet de classement et d'affectation de tous les étudiants, afin que le code ouvert soit compilable et permette d'exécuter l'algorithme,
- Limiter le périmètre aux composants nécessaires et éviter au maximum un effort de réécriture pour les équipes de développement du fait de l'historique,
- Exclure les modules présentant des risques de sécurité,
- Garantir une découpe de périmètre nette au niveau du code source, afin d'éviter toute opération manuelle pour les publications ultérieures, le code ouvert devant correspondre directement au code exécuté en production.

Deux journées de travail avec les équipes de développement (service APB à l'INP de Toulouse) ont permis d'identifier un périmètre de code source cohérent qui puisse répondre au mieux à ces différents objectifs.

Ce périmètre se décompose de la manière suivante (en rouge):



Une fois ce premier périmètre identifié, des itérations ont été nécessaires pour comprendre le code source associé, vérifier les dépendances et ajuster le périmètre du code à ouvrir afin de vérifier que les objectifs étaient bien respectés.

Au final, le périmètre de base du code source qui a fait l'objet de cette étude comporte **35 fichiers** et est décomposé en deux blocs :

- 14 fichiers PL/SQL: en charge du classement des filières non-sélectives et de fonctions support
- 21 fichiers Java: en charge de l'affectation des candidats.

A ceci s'ajoute également le schéma de la base de données sur laquelle s'appuie l'algorithme pour pouvoir fonctionner.

A noter que le périmètre et le nombre de fichiers considérés s'appuient sur la version 2016 d'APB, le code d'APB pour la version 2017 n'étant pas encore finalisé au moment de notre analyse.

Le fonctionnement de l'algorithme avec le nom des principaux scripts appelés est détaillé en annexe 4 (cinématique de l'algorithme). Un parallèle avec la procédure décrite sur le site APB⁸ est également donné pour faciliter la lecture avec les dates de l'année 2017. Toutes les saisies d'information (choix des candidatures et ordre

⁸ <https://www.admission-postbac.fr/index.php?desc=comment>

des vœux par les candidats, classements effectués par les filières sélectives, réponses des candidats) sont indiquées par des étapes de chargement de données.

L'ouverture de l'intégralité du code source d'APB n'est pas réalisable à court terme : Le périmètre identifié ci-dessus pour une première ouverture couvre le cœur du système, à savoir le classement et l'affectation des étudiants. Sans présager de failles de sécurité, le volume important de lignes de code que représentent les nombreuses interfaces à destination des candidats, des établissements et des gestionnaires de l'application nécessiterait un travail d'audit approfondi au préalable. L'ouverture n'ayant pas été pensée dès l'origine, des mécanismes de sécurité « en dur », et donc visibles dans le code source, peuvent avoir été appliqués par facilité. De plus, ces interfaces représentent une surface exposée aux attaques.

A terme, il conviendra d'évaluer le coût d'audit et de mise à niveau de l'intégralité du code source au regard de celui d'une refonte du système qui pourrait être déjà planifiée et qui intégrerait dès l'origine l'objectif d'ouverture du code source.

3.2 L'analyse technique du code source

L'analyse technique du code source n'a pas levé de point bloquant pour l'ouverture.

- Pour la partie PL/SQL le périmètre de chacun des fichiers source est clair et la sélection des fichiers et des dépendances n'a pas posé de difficulté particulière. Le code est documenté, bien qu'un dictionnaire des abréviations courantes des noms de variables puisse faciliter la lecture.
- Pour la partie Java, le périmètre des classes est également clair et facilement identifiable. Toutefois, un travail d'extraction des classes du *framework* utilisées est nécessaire pour assurer sa bonne compilation.

Le script de création de la structure de la base de données doit quant à lui faire l'objet d'un travail spécifique pour initialiser l'ensemble des tables nécessaires au bon fonctionnement de l'algorithme.

3.3 L'analyse juridique

Cette analyse rappelle le cadre juridique applicable au regard du statut d'APB (i), ainsi que les obligations en matière de publication et possibilités de réutilisation du code source et des données associées, et en précise les modalités pratiques. APB recouvre en effet à la fois les cas :

- (ii) de l'ouverture du code source d'un logiciel produit par l'administration, et de sa réutilisation,
- (iii) d'une transparence à apporter sur les algorithmes inclus dans ce code source et sur leur fonctionnement,
- (iv) d'un système qui produit lui-même des données.

Il convient de souligner, au-delà de l'analyse détaillée ci-après, que le cadre juridique délimitant APB est assez peu lisible. Hormis les dispositions prévues par le Code de l'Education et un arrêté du 8 avril 2011 relatif à la procédure de préinscription en première année d'une formation post-baccalauréat, sous la forme d'une déclaration de traitement de données à caractère personnel, il demeure difficile de tracer précisément le cadre complet de fonctionnement du système APB. Le Code de l'Education prévoyait à son article L612-3 que la « *réglementation établie par le ministre* » viendrait définir les modalités d'application de la loi pour les filières non sélectives sous tension. A notre connaissance, une telle réglementation n'a pas été pleinement formalisée. Il s'agira donc, au-delà de l'ouverture d'APB, de mieux faire converger les bases réglementaires et leur « traduction » dans le code source (à l'image de la prise en compte dans la version 2017 d'APB de la situation de famille du candidat).

i. Modalités de conception et statut juridique d'Admission Post-Bac

Le portail, et donc l'algorithme qui l'accompagne, est le fruit d'une collaboration entre l'Institut national polytechnique de Toulouse, établissement public d'enseignement supérieur à caractère scientifique culturel et professionnel, et le ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche formalisée par une convention annuelle. Ce projet s'inscrit très directement dans le cadre de leurs missions de service public respectives.

Il s'agit donc bien d'un *“document produit ou reçu, dans le cadre de leur mission de service public, par l'Etat, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission”* (Article L300-2 du code des relations entre le public et l'administration - CRPA).

ii. Communication et réutilisation du code source

Le code source est un document administratif communicable

Le statut juridique de l'algorithme d'Admission Post-Bac ne fait guère de doute aujourd'hui. En effet, suivant sa doctrine dorénavant bien établie, la Commission

d'accès aux documents administratifs (CADA) a rappelé dans son avis n°1200765 du 16 septembre 2016 *"que les fichiers informatiques constituant le code source ou algorithme sollicité produits [...] dans le cadre de leurs missions de service public respectives, revêtent le caractère de documents administratifs, au sens de l'article L300-2 du code des relations entre le public et l'administration. Ce code est, de ce fait, communicable à toute personne qui le demande, conformément à l'article L311-1 du même code."*

Cette doctrine a été ensuite, consacrée par le législateur à l'article 2 de la loi n°2016-1321 du 7 octobre 2016 pour une République numérique. Ainsi les codes-sources, et donc les algorithmes, sont des documents administratifs communicables. Les modalités d'exercice du droit d'accès aux documents administratifs sont définies au Titre Premier du Livre III du code des relations entre le public et l'administration.

Il convient de souligner par ailleurs que le nouveau cadre législatif prévoit à compter du 7 avril 2017, la publication en ligne des documents administratifs communiqués ainsi que leur version mise à jour, sous réserve que la communication du code ne porte atteinte à la sécurité du système d'information (sur ce point, voir la partie relative à l'analyse des risques de sécurité).

Le code source peut être librement réutilisé

La réutilisation des informations publiques, quant à elle, est régie par les dispositions du Titre II du Livre III du code des relations entre le public et l'administration.

En vertu de l'article L321-1, et à moins que des tiers à l'administration ne détiennent des droits de propriété intellectuelle sur ce code, il peut être utilisé par toute personne qui le souhaite à d'autres fins que celles de la mission de service public de l'administration.

Il ne ressort d'aucune pièce du dossier consulté par la mission Etalab l'existence de droits de propriété intellectuelle détenus par des tiers.

Licence de réutilisation appliquée aux documents diffusés

Les informations publiques peuvent être librement réutilisées. La réutilisation peut donner lieu à une licence (conformément à l'article L323-1 du CRPA), il ne s'agit pas là d'une obligation mais d'une bonne pratique.

L'article L323-2 du même code fixe les règles relatives aux licences. Il prévoit d'une part que la licence ne peut *"apporter de restrictions à la réutilisation que pour des motifs d'intérêt général et de façon proportionnée. Elles ne peuvent avoir pour objet ou pour effet de restreindre la concurrence"* et d'autre part que lorsque la réutilisation donne lieu à l'établissement d'une licence, l'administration choisit la licence appliquée parmi une liste fixée par décret.

Les obligations des réutilisateurs

Si la réutilisation d'information publique est libre, il convient à toutes fins utiles d'insister sur le fait que l'ensemble du cadre réglementaire doit être respecté par le réutilisateur, notamment celui relatif à la protection des données à caractère personnel. Ainsi, bien que le code-source de l'algorithme en tant que tel ne contienne aucune donnée à caractère personnel, si le traitement réalisé par le réutilisateur parvient à identifier directement ou indirectement des personnes, le réutilisateur est soumis au cadre juridique de la protection des données à caractère personnel. C'est donc sur ce dernier que pèseront les obligations de déclarations.

iii. Communication des données

Le système Admission Post-Bac contient également des données à caractère personnel, notamment concernant les candidats et les vœux qu'ils formulent. Conformément à l'article L312-1-2 du Code des relations entre le public et les administrations, les documents et données comportant des données à caractère personnel ne peuvent être rendus publics que si des dispositions législatives ou réglementaires le prévoient, ou que les personnes concernées aient donné leur accord (consentement) ou que les informations aient fait l'objet d'un traitement permettant de rendre impossible l'identification de ces personnes. Dans le cas d'Admission Post-Bac, aucune disposition législative ou réglementaire ne permet à ce jour la diffusion de ces données à caractère personnel, et le consentement à la diffusion n'a pas été recueilli auprès des personnes concernées. Dès lors, il convient d'anonymiser les données à caractère personnel avant de les rendre publiques.

Les autres données ne présentant pas de risque de ré-identification ont vocation à être diffusées en open data conformément à la législation en vigueur. A partir d'octobre 2018, les données dont la publication présente un intérêt social devront être diffusées par défaut (article L312-1-1 du CRPA).

iv. Obligations relatives aux traitements algorithmiques lorsqu'ils fondent une décision individuelle

Par ailleurs, le dispositif Admission Post-Bac est également concerné par les nouvelles dispositions relatives aux traitements algorithmiques introduites par la loi n°2016-1321 du 7 octobre 2016 pour une République numérique :

- d'abord, celles prévues à l'article L311-3-1 et suivants du CRPA : le ministère devra donc prévoir une mention explicite informant l'utilisateur de l'existence d'un algorithme participant à la décision individuelle à compter du premier septembre 2017. Cette mention devra expliquer les finalités de l'algorithme,

rappeler le droit de l'utilisateur d'obtenir communication et lui expliquer la procédure à suivre pour obtenir ces informations ;

- ensuite, celles prévues à l'article L312-1-3, qui vise à la publication "*en ligne des règles définissant les principaux traitements algorithmiques utilisés dans l'accomplissement de leurs missions lorsqu'ils fondent des décisions individuelles*".

3.4 L'analyse des risques de sécurité

L'évaluation de sécurité n'a porté que sur la version 2016 du code source. Les entrées-sorties associées aux algorithmes qui vont être publiés sont très majoritairement des entiers et présentent ainsi une surface d'attaque réduite. Le code associé présente quelques mauvaises pratiques, toutefois le contexte d'exécution de l'algorithme ne met pas en évidence de scénarios d'attaque majeurs.

-> A l'issue des analyses menées, nous ne voyons pas d'obstacle juridique ou technique à l'ouverture prochaine du système Admission Post-Bac dans le périmètre considéré par notre étude.

4. Recommandations et plan d'actions sur l'ouverture du système APB

Les recommandations que nous formulons concernent les modalités d'ouverture du code source et des données d'Admission Post-Bac, ainsi que l'animation et la mobilisation des écosystèmes autour de cette démarche d'ouverture.

En effet, la manière d'ouvrir le code et les données influe aussi sur la dynamique de réutilisation. Les actions listées ci-dessous se fondent sur un dialogue réalisé avec les différentes parties prenantes, sur l'expérience acquise par Etalab et la DINSIC, et sur des bonnes pratiques en matière d'open source.

Recommandation n°1

Publier, sans attendre une nouvelle demande de communication, le code source de la dernière version d'APB et les données non-réidentifiantes associées (cf. *recommandation n°6*). Accompagner cette ouverture d'un événement avec la communauté (cf. *recommandation n°9*).

4.1 Recommandations pour l'ouverture du code source

L'emplacement de la publication du code source d'APB est un choix important qui doit respecter les critères suivants:

- Être un système de gestion de version décentralisé (eg. git, bitbucket, mercurial...)
- Permettre une consultation publique sans inscription du code source

Si une plateforme ministérielle peut être utilisée, le code source d'APB a vocation à recevoir des suggestions de mise à jour ou correction de personnes extérieures au ministère qui n'y sont généralement pas inscrites. Une plateforme publique comme github.com, gitlab.com ou framagit.org permet de toucher une plus large audience.

Recommandation n°2

Mettre à disposition le code source sur un système de gestion de version décentralisé (par exemple git) au moyen d'une plateforme ouverte au public. La documentation existante doit être partagée.

Concernant la licence à utiliser, le code source d'APB n'impose pas la nécessité d'introduire une notion de partage à l'identique qui servirait l'intérêt général. Les fonctionnalités apportées par l'algorithme de mariages stables, bien que contextualisées au besoin d'Admission Post-Bac, peuvent servir à d'autres applications qu'elles soient ouvertes ou propriétaires.

Recommandation n°3

Soumettre le code source à une licence de réutilisation. La mission Etalab invite le ministère à choisir une licence dite "permissive" permettant toute forme de réutilisation. La famille de licence permissive préconisée pour l'ouverture du code APB inclut la licence CeCILL-B⁹ ou la licence Apache 2.0¹⁰ qui permet de viser une diffusion à l'international.

Ne s'agissant pas d'une ouverture de code source dès le démarrage du projet, une traçabilité est nécessaire afin d'en garantir l'origine et l'authenticité. Dans le cadre d'APB, seuls deux développeurs sont les auteurs de l'algorithme d'affectation et peuvent attester de l'originalité de leur travail.

Recommandation n°4

Pour simplifier et garantir l'efficacité de la réutilisation, faire signer par chacun des agents susceptibles de détenir des droits de propriété intellectuelle un "Developer's Certificate of Origin" (cf. annexe 3)

Enfin l'ouverture ponctuelle du code source d'APB pour les versions 2016 et 2017 ne constitue qu'une première étape de l'ouverture. Un scénario reposant sur un phasage progressif, pour permettre plus de transparence jusqu'à une ouverture de la gouvernance (c'est à dire du processus de prise de décisions sur APB), pourrait reposer sur les **étapes suivantes**:

1. Etape 1 : Ouvrir dans un premier temps le code source en l'état, pour permettre de l'étudier,
2. Etape 2 : Permettre ensuite d'exécuter facilement le code source avec des technologies libres et faciliter des simulations,

⁹ Disponible à l'URL : http://www.cecill.info/licences/Licence_CeCILL-B_V1-fr.html

¹⁰ Disponible à l'URL: <https://www.apache.org/licenses/LICENSE-2.0>

3. Etape 3: Permettre les contributions extérieures et se donner les moyens de les accepter,
4. Etape 4 : Coder en mode ouvert pour les versions à venir,
5. Etape 5: Proposer une gouvernance ouverte avec l'ensemble des parties prenantes.

Les étapes 1 et 2 pourront être combinées à l'occasion d'un événement associant les services du Ministère, les développeurs de l'application ainsi que les communautés externes (cf. recommandation n°9).

Recommandation n°5

Au-delà de l'ouverture du code, prévoir à terme le passage d'Admission Post-Bac dans un modèle de développement ouvert aux contributions et étudier les conditions de l'ouverture de la gouvernance du système à l'ensemble des parties prenantes.

4.2 Recommandations pour l'ouverture des données d'APB

Depuis 2016, le Ministère a accès à l'intégralité des bases de données d'Admission Post-Bac géré par le service APB de l'INP Toulouse. Le projet APB Stat porté par le Département des études statistiques de l'enseignement supérieur (SIES / MENESR) vise à valoriser ces données pour répondre aux besoins des chercheurs et des services du Ministère.

Un important travail a été réalisé dans le cadre d'APB Stat pour documenter l'ensemble des variables du système APB, et préparer quatre bases de données distinctes, respectivement sur les candidats, les vœux, les propositions et les établissements d'accueil et de formation.

Le Ministère a mis en place un comité scientifique qui s'est réuni deux fois depuis le lancement du projet. A terme, les données disponibles dans APB Stat ont vocation à être accessibles aux chercheurs via le réseau Quêtelet.

Dans le cadre de notre mission, nous avons identifié trois types de données différentes:

- les données individuelles sur les candidats et les parcours: elles présentent un risque de ré-identification et n'ont pas vocation à être diffusées en mode *open data* sans anonymisation préalable,
- les données qui ne présentent pas de risque de ré-identification et qui ont vocation à être proposées en open data conformément à la législation en vigueur,
- enfin, un ensemble de jeux de données permettant de rejouer le système APB et de s'assurer ainsi de la bonne exécution du code.

Conformément à notre lettre de mission sur les conditions de l'ouverture des données d'APB, notre analyse s'est concentrée sur les deux derniers types de données, l'accès aux données individuelles à des fins de recherche étant par ailleurs déjà bien identifié par APB Stat et les services statistiques du Ministère.

De nombreuses données, qui ne présentent pas de risque de ré-identification, pourraient être publiées en open data sur la plateforme dédiée du Ministère et sur la plateforme ouverte des données publiques data.gouv.fr. Les services académiques d'information et d'orientation (SAIO) ont aujourd'hui eux aussi accès aux bases de données d'Admission Post-Bac. En pratique, tous les SAIO n'utilisent pas cette faculté d'accès de la même manière et ne partagent pas nécessairement une méthodologie commune. Ainsi, pour certaines académies, on ne retrouve qu'un bilan statistique général, alors d'autres services d'information et d'orientation, on retrouve en ligne des données très détaillées¹¹.

Ces données sont très utiles pour les candidats et les familles, elles permettent notamment d'éclairer des choix. Le format de leur diffusion actuelle (document PDF) ne permet pas cependant de les réutiliser et n'est pas conforme aux récentes dispositions de la loi pour une République numérique. En effet, toute mise à disposition effectuée sous forme électronique en application du Code des relations entre le public et les administrations se fait dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé.

Recommandation n°6

Ouvrir les données non-réidentifiantes concernant les candidats, les établissements, les vœux et les propositions, et en premier lieu celles qui permettent une meilleure information des candidats et des familles.

¹¹ L'académie de Lyon a diffusé en décembre 2016 un document intitulé "bilan de l'affectation post-bac". Ce document contient des données non-réidentifiantes mais très précieuses pour l'information des candidats. Il précise, pour chaque filière sélective de l'académie la capacité d'accueil, le nombre de candidats, filles et garçons ayant demandé cette filière en 1er vœu et tous vœux confondus, le nombre d'admis sur vœu 1 et le nombre d'admis quel que soit le rang de vœu.

Le système Admission Post-Bac contient également des données à caractère personnel, notamment concernant les candidats et les vœux qu'ils formulent. Conformément à l'article L. 312-1-2 du Code des relations entre le public et les administrations, les documents et données comportant des données à caractère personnel ne peuvent être rendus publics que si des dispositions législatives ou réglementaires le prévoient, ou que les personnes concernées aient donné leur accord (consentement) ou que les informations aient fait l'objet d'un traitement permettant de rendre impossible l'identification de ces personnes. Dans le cas d'Admission Post-Bac, aucune disposition législative ou réglementaire ne permet à ce jour la diffusion de ces données à caractère personnel. Dès lors, il convient d'anonymiser les données à caractère personnel avant de les rendre publiques.

Plusieurs administrations diffusent aujourd'hui des données qui ont fait l'objet d'un traitement d'anonymisation. Par exemple, dans le cadre de l'ouverture des données de santé, l'Assurance Maladie (CNAMTS) et la mission Etalab ont préparé la mise à disposition de données détaillées sur les remboursements de l'assurance maladie tous régimes confondus. Afin d'empêcher toute ré-identification d'un patient ou d'un professionnel de santé, le fichier DAMIR a fait l'objet d'agrégation en fonction du niveau de détail des données et de la granularité géographique¹².

Recommandation n°7

Travailler sur l'anonymisation des données issues d'Admission Post-Bac, en s'appuyant sur l'expertise de l'Administrateur général des données et le retour d'expérience d'autres administrations (dont l'Assurance Maladie).

Afin de permettre de rejouer le système APB et d'exploiter pleinement le code source qui sera mis à disposition, il faudra disposer d'un jeu de données qui simule les vœux des candidats et les propositions des établissements. La construction d'un tel jeu de données pourrait être l'un des sujets proposés lors du premier événement de type hackathon (cf. infra l'animation et la mobilisation des écosystèmes).

Recommandation n°8

Créer un ensemble de jeux de données fictives permettant de simuler et de "rejouer" le système à partir du code source qui sera publié.

¹² Deux versions du fichier sont disponibles: l'une contient des données avec un très grand nombre de variables et agrégées au niveau de grandes régions, l'autre est moins détaillé mais est fourni pour chaque département.

4.3 Recommandations pour l'animation et la mobilisation des écosystèmes

La mise à disposition en ligne du code source et des données d'APB n'est qu'une première étape. Afin d'exploiter pleinement le potentiel de réutilisation de ces éléments, nous recommandons de mettre en place des actions d'animation et de mobilisation des écosystèmes.

La première étape serait d'organiser au plus tôt un premier événement de type hackathon dédié à l'accompagnement de l'ouverture. Ce hackathon, organisé avec le soutien d'Etalab marquerait la **mise en ligne du code source et d'un premier ensemble de données**. Plusieurs sujets pourraient être traités: la documentation du code et des données, la création d'un environnement pour permettre l'exécution du code indépendamment de l'environnement actuel propriétaire (base Oracle), la préparation d'un jeu de données fictives à partir des spécifications du système APB.

Ce premier événement doit associer les services du ministère, les équipes de développement d'Admission Post-Bac ainsi que les communautés qui se sont déjà engagées autour de la première transmission du code source. Il conviendra d'être clair sur les règles d'engagement entre les participants, le Ministère et l'équipe de développement d'APB. Les moyens restreints de cette dernière équipe ne permettent pas d'envisager de traiter immédiatement l'ensemble des propositions d'amélioration du code qui auront pu être formulées.

Recommandation n°9

Organiser, avec l'appui d'Etalab et au plus tard à la rentrée 2017, un événement de type hackathon à l'occasion de la publication du code source, avec pour objectif de faciliter la réutilisation du code et des données.

Au-delà de la mise en ligne du code source et des données, une approche ouverte et collaborative (*open science*) autour d'APB pourrait être lancée avant la session 2018, en lien notamment avec les acteurs de la recherche. Plusieurs thèmes pourraient ainsi être abordés, citons par exemple:

- l'amélioration d'Admission Post-Bac du point de vue de l'algorithmique et de l'informatique,
- l'explicitation des finalités du système et de son fonctionnement,
- la pédagogie d'APB au-delà de l'approche aujourd'hui centrée sur la procédure APB et son calendrier,
- l'analyse des conditions de l'acceptation sociale d'un système technique de cette ampleur,
- le lien entre du code informatique et le corpus juridique existant.

Recommandation n°10

S'engager dans une démarche ouverte et collaborative autour de l'amélioration d'Admission Post-bac, notamment avec les acteurs de la recherche (*open science*).

5 - Conclusion: APB, cas d'école

Le système Admission Post-Bac constitue à plusieurs titres un cas d'école, dont l'Etat devrait tirer des enseignements utiles dans les années à venir.

Les exigences que la société formule à l'égard d'APB sont à la mesure de l'ampleur du système (qui concerne plus de 850 000 candidats) et de l'importance qu'il a pris pour les intéressés. La première exigence est que le système opère sur des bases transparentes et compréhensibles par tous. L'ouverture d'APB est un moyen de répondre à cette demande, et n'est pas une fin en soi. La seconde exigence est que le système APB puisse être redevable, c'est à dire qu'il puisse rendre des comptes à ceux qui sont le plus directement concernés, les candidats¹³.

Aussi efficace et puissantes soient-elles, l'ouverture et la transparence ne vont pas résoudre, à elles seules, l'ensemble des défis auxquels est confronté APB et plus généralement l'ensemble des systèmes dans lesquels des algorithmes interviennent dans la conduite de l'action publique.

5 axes de travail pour rendre APB plus intelligible

Dans le cadre du programme NosSystèmes, dont le SGMAP est partenaire, la Fondation Internet Nouvelle Génération a formulé un ensemble de recommandations pour rendre le système APB plus intelligible et concevoir les conditions d'un meilleur dialogue entre usagers et concepteurs de ces dispositifs. Ces recommandations sont formulées selon cinq axes de travail, qui recoupent en partie les conclusions de notre mission.

Axe	Recommandations
Responsabilité	Construire une Base de connaissance APB, avec forums où des experts répondent, offrent une assistance, et documentent l'ensemble avec des partenariats avec des médias externes, sur le modèle de service-public.fr. Créer une équipe dédiée à la qualité du dialogue sur la plateforme APB.
Explicabilité	Donner la "Big Picture", la vision et les motivations politiques. Ne pas esquiver les questions qui fâchent, comme celle de la sélection... Tisser les liens manquants entre le code d'APB et celui de l'Education nationale
Jouabilité	Proposer un outil de simulation pour aider à comprendre, tester des hypothèses, saisir les différences entre académies, permettant d'offrir une symétrie plus grande de l'information, des statistiques pour permettre aux élèves de mieux comprendre et d'anticiper.

¹³ C'est dans cette optique que la loi pour une République numérique a introduit une disposition qui permet d'informer un individu sur les traitements algorithmiques dont il fait l'objet.

Symétrie	Éclairer les critères, augmenter le niveau d'information : aires de recrutement des universités, tensions, lieux et résultats des élèves sélectionnés les années précédentes.
Open Science	Ouvrir des données, inviter des chercheurs, lancer un "HackYourAPB"... Un système parallèle sur une académie pour tester comment on pourrait l'améliorer.

APB est aussi un cas d'école d'un système hérité et dont le développement n'a pas été initialement conçu pour être ouvert et partagé. Le passage à un mode de développement ouvert demande donc un changement important pour l'ensemble des acteurs concernés, et en premier les équipes de développement d'APB. Ce changement sera nécessairement progressif, l'étape de publication du code source doit être pensée **comme un point de départ d'une démarche, et non un point d'arrivée.**

Enfin, Admission Post-Bac est une illustration du **lien étroit entre le code informatique et le droit**. "**Le code fait loi**" ("code is law") expliquait Lawrence Lessig il y a 17 ans déjà. Le droit dans lequel évolue APB est déterminé par le Code de l'Éducation. Mais face aux difficultés d'accueil des étudiants en licences non-sélectives sous tension et en l'absence de réglementation sur ce point¹⁴, le code informatique d'APB met aujourd'hui en œuvre des dispositifs qui n'ont pas été prévus par le législateur, à l'instar du dispositif de tirage au sort pour les filières non-sélectives sous tension.

Cela pose un défi d'avenir : comment rapprocher plus étroitement et opérationnellement le droit du code informatique, en particulier lorsque - comme c'est le cas pour Admission Post-Bac - le code évolue constamment et rapidement afin de refléter les conditions même d'application de la loi et les choix politiques qui les sous-tendent ?

Le présent rapport a été élaboré par la mission Etalab, et notamment Christophe Benz, Simon Chignard, Laurent Joubert, Laure Lucchesi, Thomas Menant et Jill-Jênn Vie.

¹⁴ L'article L 612-3 du Code de l'Éducation est rédigé ainsi: "lorsque l'effectif des candidatures excède les capacités d'accueil d'un établissement, constatées par l'autorité administrative, les inscriptions sont prononcées, après avis du président de cet établissement, par le recteur chancelier, selon la réglementation établie par le ministre chargé de l'enseignement supérieur, en fonction du domicile, de la situation de famille du candidat et des préférences exprimées par celui-ci."

Glossaire

Algorithme

De manière simplifiée, un algorithme est une méthode, une façon systématique de procéder pour faire quelque chose : trier des objets, situer des villes sur une carte, multiplier deux nombres, extraire une racine carrée, chercher un mot dans le dictionnaire... (Source : Interstices, INRIA)

Données ouvertes

Données qu'un organisme met à la disposition de tous, sous forme de fichiers numériques, dans un format exploitable informatiquement (voir la définition), gratuitement, afin de permettre leur réutilisation. Les données ouvertes n'ont généralement pas de caractère personnel. Elles sont les plus granulaires possibles, mais ne permettent pas l'identification ou la ré-identification d'une personne. La réutilisation des données ouvertes n'est généralement pas soumise à conditions, sauf lorsque l'organisme opte pour des licences impliquant certaines règles de repartage (exemple licence ODBL).

Plus d'information sur : <http://www.data.gouv.fr/fr/faq> (Source : <http://opendatahandbook.org/glossary/en/>)

Forge

En informatique, une forge est un système de gestion de développement collaboratif de logiciels.

(Source : https://fr.wikipedia.org/wiki/Forge_%28informatique%29)

Hackathon

Événement réunissant des développeurs et des innovateurs afin de faire de la programmation informatique collaborative, sur un ou plusieurs jours. (Source : Conseil national du Numérique)

Open source

Un logiciel libre (ou "open source") est un logiciel dont l'utilisation, l'étude, la modification et la duplication en vue de sa diffusion sont permises, techniquement et légalement. Ceci afin de garantir certaines libertés induites, dont le contrôle du programme par l'utilisateur et la possibilité de partage entre individus. (Source : Conseil national du Numérique)

Annexe 1 - Lettre de mission



MINISTÈRE DE L'ÉDUCATION NATIONALE,
DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

Le Secrétaire d'Etat chargé de l'enseignement supérieur et de la recherche

Paris, le 1^{er} février 2017

Madame la Directrice,

Le logiciel libre s'impose de plus en plus comme la norme dans toutes les administrations nationales, en témoignent les mesures très volontaristes prises par la Bulgarie ou l'Estonie par exemple, en cohérence avec une dynamique plus globale de transparence de l'action publique à l'heure du numérique, qui a cours à l'échelle européenne et internationale.

La France s'est quant à elle montrée proactive en la matière, avec la promulgation de la loi pour une République numérique (Loi n° 2016-1321) par le Président de la République le 7 octobre dernier, qui consacre notamment l'ouverture des algorithmes publics.

Je souhaite que l'Enseignement supérieur et la Recherche français poursuive lui aussi les efforts déjà largement initiés en matière d'ouverture des données et de promotion du logiciel libre, car ils constituent à mes yeux de vraies avancées en faveur de la réappropriation de la décision et du service publics par les citoyens.

A cet égard et compte tenu de son caractère essentiel pour les candidats et leur famille, je considère que le portail « Admission Post-Bac (APB) » est un cas d'école de la nécessité de donner à voir aux usagers le contenu des logiciels qu'ils utilisent, *a fortiori* lorsqu'ils fournissent des informations cruciales d'une part, qu'ils sont mis à disposition par la puissance publique d'autre part. Mais il illustre aussi la complexité à transiter de façon rapide d'un système hérité et historiquement fermé vers un système ouvert, car ce sont aussi bien des questions juridiques, techniques que politiques qui se posent alors.

Madame Laure LUCCHESI
Directrice de la mission Etalab
Services du Premier ministre
Secrétariat général pour la modernisation de
l'action publique
39-43 Quai André Citroën
75015 PARIS

21 rue Descartes - 75231 Paris cedex 05 - Tél. : 01 55 55 90 90

- 2 -

Compte tenu de son expertise dans ce domaine et après l'expérience menée avec succès sur le calculateur des impôts, je souhaiterais confier à Etalab une mission d'étude des conditions d'ouverture d'APB, qui inclurait les volets suivants :

- un plan d'action à mettre en œuvre pour ouvrir le code et autant que faire se peut l'explicitation associée (variables, fonctions, etc.), ce plan d'action pourra comporter plusieurs scénarios d'ouverture ;
- un accompagnement sur le chantier déjà initié de l'ouverture des données collectées par APB et autant que faire se peut l'explicitation associée (structure de la base de données, etc.) ;
- une étude d'impact juridique ;
- une évaluation des risques techniques, notamment de sécurité ;
- l'organisation d'un *hackathon* autour de nouveaux services aux usagers, avec les parties prenantes intéressées.

Dans le cadre de votre mission, vous travaillerez autant que possible sur la version la plus récente d'APB pour ce qui concerne le code – pour être au plus près de celle qui sera effective pour la rentrée 2017, et sur les données collectées à l'occasion de la rentrée 2016.

Vous travaillerez également de concert avec les interlocuteurs compétents de la Direction Générale de l'Enseignement Supérieur et de l'Insertion Professionnelles (DGESIP), de la Direction Générale de la Recherche et l'Innovation (DGRI) et de la Direction des Affaires Juridiques (DAJ) du ministère.

Des expertises externes pourront selon votre jugement contribuer à ces travaux, notamment des chercheurs ou encore l'Agence nationale de la sécurité des systèmes d'information (ANSSI) pour ce qui concerne les questions de sécurité informatique.

Les résultats de vos travaux devront être remis au plus tard fin mars 2017.

Je vous prie d'agréer, Madame la Directrice, l'expression de ma considération distinguée.



Thierry MANDON

Annexe 2 - Liste des personnes et organismes auditionnés

Ministère de l'Enseignement supérieur et de la recherche

Richard Audebrand

Ceren Inan

Emmanuel Weisenburger

Yann Caradec

INP Toulouse - service APB

Bernard Koehret

Sonia Bonnafé

Serge Richard

Mickaël T.

Frateli Labs

Muriel Ekovich

Sophie Pelisson

Université Paris Est Créteil

Pierre Valarcher

Fondation Internet nouvelle génération

Hubert Guillaud

Thierry Marcou

Annexe 3 - Proposition de DCO developer certificate of origin

Developer Certificate of Origin
Admission Post-Bac
Version 0

Proposé par le :

- Ministère chargé de l'Education nationale, de l'enseignement supérieur et de la recherche, 1 rue Descartes – 75 231 Paris cedex 05
 - L'institut national polytechnique de Toulouse, 6 allée Emile Monso – BP 34038 – Toulouse cedex 4
 - L'association pour la promotion des concours communs polytechniques (APCCP)
- qui mettent à disposition le code source sous la licence CeCILL-B v1.0

[Developer's Certificate of Origin 1.0]

En participant à ce projet, je m'engage / j'assume que :

(a) Ma contribution sera redistribuée publiquement et gratuitement sous la forme d'un code source et la licence open-source d'origine : CeCILL-B sera appliquée ;

(b) Ma contribution a été réalisée par mes soins et que je suis habilité [librement choisi] à la placer sous la licence open-source indiquée dans le code [fichier] ;

(c) Ma contribution s'appuie sur le code source développé dans le cadre du projet, à ma connaissance, ce code est couvert par une licence open-source (CeciLL) et que j'avais l'autorisation d'y apporter des modifications ;

Ou bien

(d) La contribution m'a été offerte par un tiers qui a certifié que les points (a), (b) et (c) ont été respectés et que je n'ai pas modifié cette contribution.

(d) Je comprends et certifie participer librement et volontairement à ce projet et à y verser ma contribution. J'ai été informé et je consens à ce que ma contribution (comportant également l'ensemble des informations à caractère personnel me concernant et notamment ma signature soit attachées indéfiniment, à ce projet et qu'ils peuvent librement être rediffusés à des tiers dans le cadre du présent projet.

Signé par: Jean Dupont <jean.dupont@education.gouv.fr>

Annexe 4 - Cinématique de l'algorithme

1. Initialisation de la base de données¹⁵, des packages¹⁶, vues¹⁷ et des fonctions¹⁸ PL/SQL nécessaires à son bon fonctionnement

2. Chargement des données établissements

1ère ETAPE: L'inscription par internet (**du 20 janvier au 20 mars**)

3. Chargement des données étudiants

2ème ETAPE: Confirmation des dossiers de candidatures (**jusqu'au 2 avril**)

4. Chargement des candidatures des étudiants
5. Exécution du script PL/SQL d'affectation GP_L1 à chaque candidature en filière non sélective
6. Chargement des groupes et classements des filières sélectives - **27 mai**
 - a. Création des groupes par l'établissement
 - b. Pour chaque groupe, création d'un classement
 - c. Pour chaque groupe, création des données d'appel
7. Chargement de l'ordre définitif des vœux des étudiants - **31 mai**

3ème ETAPE: Résultats et réponses (**du 1er juin au 19 juillet**)

8. Exécution du script PL/SQL Generation_classement_licence.sql - **1er juin**

EXECUTION D'UNE PHASE (les dates indiquées correspondent à la 1ère phase)

9. BOUCLE ÉTABLISSEMENT **quotidienne du 2 juin au 7 juin**

- a. Exécution du script PL/SQL RAZ simulation qui démarre une simulation
- b. Exécution des programmes Java ThreadChargement.java, ThreadAlgorithme.java et ThreadAffectation.java
- c. Exécution du programme Java ThreadEcritureResultatsBatch.java
- d. Exécution du programme Java vérif_Intégrité
- e. Consultation des résultats de la simulation par les établissements
- f. Chargement de la mise à jour des données d'appel, pour chaque groupe de chaque établissement

PROPOSITION D'ADMISSION

10. Exécution du script PL/SQL Validation_Appel et affichage des résultats aux candidats
11. Chargement des réponses des candidats sur la meilleure proposition (oui / oui mais / non) - **13 juin** puis retour à l'étape 9 si ce n'est pas la 3ème phase.

Pour référence, les dates des boucles quotidiennes établissement sont:

- Pour la 2ème phase du 14 au 25 juin, les candidats devant répondre du 26 juin au 1er juillet
- Pour la 3ème phase du 2 au 13 juillet, les candidats devant répondre du 14 au 19 juillet

¹⁵ MPD_diffusion_code.sql

¹⁶ pk_generation_class_licences_body.sql, pk_generation_class_licences_head.sql, pk_verif_classement_body.sql, pk_verif_classement_head.sql

¹⁷ fonctions_pg_affectation.sql, vues_pg_affectations.sql

¹⁸ six_voeu_L1.sql, voeu_complet.sql, voeu_groupe_L1_relatif.sql, voeu_L1_relatif.sql

Toute absence de réponse est considérée comme un NON et libère la place dans l'établissement pour la prochaine phase.