



Fiche de projet pour les institutions publiques

Informations pratiques

Nom de l'institution publique ayant introduit le projet: SPF Justice
http://www.just.fgov.be/index_fr.htm

Nom du projet : Introduction des STatistiques Pénales sous forme électronique (INSTAP)

Timing des tâches scientifiques : Phase I : 12 mois (septembre 2008-septembre 2009), Phase II : 12 mois (septembre 2009-septembre 2010)

Budget alloué à l'équipe scientifique :

Les deux phases de ce projet sont mises en oeuvre par la même équipe mais fera l'objet de deux contrats séparés. La phase II est conditionnée au financement de la sous-traitance en vue d'encoder les statistiques pénales se trouvant sous format papier.

Délivrables et activités de valorisation à charge de l'équipe scientifique (non compris ceux décrits dans la fiche ci-après):

- *Séminaire international* en début de projet (état de l'art) ;
- *rapport succinct* exposant la manière dont les données visées par le projet peuvent être comparées et harmonisées au niveau international/Européen;
- *activité de valorisation en fin de projet* : l'équipe fera une proposition dans son formulaire de soumission avec pour objectif de faire connaître au mieux le résultat du projet.
- *délivrables visés par le contrat de recherche* : *description du projet* (max. 25 lignes en trois langues – FR/NL/EN/), *résumé du projet en maximum 10 pages* également en quatre langues et à remettre à l'issue du projet, un *texte en FR/NL/EN destiné à la Newsletter AGORA* (max. 40 lignes), *rapports intermédiaires* (maximum 2 pages, tous les 6 mois permettant de libérer les tranches de financement de l'équipe scientifique), *rapports pour le comité d'accompagnement* (voir plus loin) ;

Ces activités sont à charge du budget de l'équipe scientifique.

Déroulement du projet :

L'institution publique gère son projet et entretient un dialogue étroit avec l'équipe et la Politique scientifique (qui forment le *comité technique*). Le projet est accompagné par un comité d'accompagnement qui se réunit à échéances régulières et au moins trois fois durant la durée du projet (au début, au milieu et à la fin du projet). L'équipe se charge de fournir les documents pour ce comité, l'institution publique gère l'agenda et les procès-verbaux.

date de la réunion d'information (inscription une semaine à l'avance par email – naji@belspo.be . A défaut d'inscription, la réunion est annulée) : **encore à confirmer**

Le Projet

1. Nom de l'Institution publique

SPF Justice

2. Nom du projet

INtroduction des STatistiques Pénales sous forme électronique

3. Acronyme

INSTAP: INtroduction des STatistiques Pénales sous forme électronique

4. Description du projet et contexte dans lequel il s'inscrit au sein de l'institution publique

Dans le cadre du programme pluriannuel de soutien au développement de la société de l'information (2001-2008), le SPF Politique scientifique a financé un projet de recherche visant l'élaboration d'un *système informatique permettant l'intégration, la diffusion et la mise à disposition critique des statistiques pénales existantes*. La recherche a abouti à la construction d'un *prototype* (incluant 30 années de statistiques pénales sur les 177 années d'existence de l'état belge).

Cet outil pourrait permettre au SPF Justice de valoriser la masse importante de données statistiques existantes susceptibles d'apporter une connaissance sur le fonctionnement de la justice pénale tant dans le passé que durant la période actuelle. L'outil favoriserait l'accès à ces statistiques qui sont actuellement totalement dispersées, et permettrait leur diffusion et leur utilisation de façon plus pertinente. Un site web unique permettrait d'accéder à l'ensemble des statistiques pénales belges en un seul et même lieu.

Cette usage n'est toutefois envisageable que si les moyens sont investis pour passer du *prototype* déjà disponible au *à un produit opérationnel* au niveau du SPF Justice, ce qui suppose entre autres l'intégration des structures des années restantes (notamment les données plus actuelles) et l'encodage des chiffres des années restantes. L'objet de ce projet Agora INSTAP est de permettre le *passage du prototype à un produit intégrant l'ensemble des statistiques pénales belges dans l'application*. **Ce prototype, développé dans le cadre du projet QUETELET-NET, est propriété de l'Etat, selon les dispositions contractuelles de la Politique scientifique fédérale. Il peut donc être délivré à l'équipe qui sera sélectionnée pour mettre en oeuvre INSTAP.**

INSTAP ne vise pas à appuyer la production de statistiques pénales mais bien à se munir d'un outil qui, mis à la disposition des producteurs de données statistiques, en permettra une meilleure diffusion et utilisation.

Une brève présentation est faite ci-après du prototype Quetelet.net (voir aussi sur <http://www.quetelet.net/>).

Description du prototype existant

Quetelet.net est un projet né de la rencontre entre deux expertises scientifiques. D'une part, l'ingénierie des bases de données et en particulier les bases de données temporelles. D'autre part, l'expertise accumulée autour des statistiques judiciaires et pénales. L'objectif du projet

était de créer un système critique d'intégration, de numérisation et d'accès aux statistiques pénales belges.

- *Intégration et numérisation.* Les statistiques pénales belges sont, quel que soit le support (papier ou, plus récemment, électronique), présentées invariablement depuis 175 ans sous forme de tableaux préconçus. L'équipe scientifique a développé une application permettant à des encodeurs experts (pour les structures des tableaux) et non-experts (pour les chiffres) d'encoder ces tableaux dans une base de données. Au sein de celle-ci, chaque fait statistique est modélisé sous la forme d'un vecteur composé d'une valeur pour chacune des dimensions qui le décrivent, d'une période de validité et d'un chiffre. Une telle opération, qui éclate véritablement la structure des tableaux, permet de proposer à l'utilisateur final de construire des requêtes statistiques qui font appel à des chiffres issus de plusieurs tableaux. Par ailleurs, les tableaux sont également numérisés sous forme de microfilms scannés. Le projet permet d'assurer ainsi la conservation de statistiques qui, pour certaines, commencent à se dégrader sérieusement sous leur format papier. Au delà de cette fonction de conservation, la possibilité qu'aura l'utilisateur de consulter à tout moment la source originale pour chaque chiffre du tableau issu de sa requête constitue une garantie d'authenticité indispensable.
- *Accès.* Les statistiques pénales belges sont actuellement disséminées dans de nombreux lieux de conservation. Aucune institution n'est en possession de l'entièreté de la collection. Pour construire sa requête, l'utilisateur final est assisté dans sa recherche statistique par une interface de formulation de requêtes qui lui permet de déterminer la période, les dimensions et les valeurs pour lesquelles il souhaite disposer de données chiffrées. La diffusion des statistiques est donc orientée vers l'utilisateur et permet à celui-ci d'effectuer des opérations qui prendraient un temps considérable s'il lui fallait consulter les versions papier. Il en est par exemple ainsi de la sérialisation (comparaison des données dans le temps) ou de la comparaison géographique qui, grâce à l'application Quetelet.net, peuvent présenter sur un seul tableau - construit sur base d'une requête - des chiffres qui nécessiteraient la consultation de très nombreux tableaux papier.
- *Critique.* La facilitation de l'accès aux statistiques pénales pose néanmoins question. L'accès public est en effet susceptible de provoquer une augmentation de mauvais usages et interprétations. C'est pourquoi, le site web a été conçu de sorte à pouvoir fournir de nombreux contenus critiques qui permettent d'éviter ces écueils en avertissant l'internaute de la complexité des données qu'il consulte. D'une part, le site fournit des contenus critiques statiques : Que sont les statistiques pénales belges ? Que peut-on en faire ? Qui les produit ? ... D'autre part, il fournit des critiques dynamiques générées automatiquement à partir des requêtes des utilisateurs. Un analyseur de résultats repère les éventuelles évolutions sémantiques des nomenclatures au cours du temps et en avertit l'utilisateur. Ainsi, lors d'une requête de sérialisation sur les condamnés mineurs par exemple, l'évolution de l'intervalle d'âge définissant la catégorie « mineurs » est signalée automatiquement.

En accord avec les objectifs du projet Quetelet.Net, l'architecture du système se compose de quatre sous-systèmes fonctionnels et techniques (Figure 1). La *base de données* stockant l'ensemble des informations des tableaux statistiques constitue le cœur du système. Le *sous-système de collecte* permet d'extraire les données des tableaux sous forme papier pour ensuite les intégrer dans la base de données. Le *site web* est composé notamment d'une interface de formulation de requêtes permettant à l'utilisateur de spécifier sa demande. Enfin, le *sous-système de requêtes* constitue le lien entre le site web et la base de données. Un *traducteur de requêtes* traduit chaque requête formulée par l'utilisateur dans le langage de définition de données (ici SQL) du système de gestion de base de données (SGBD). L'exécution de la requête fournit un résultat transmis à un *analyseur*. Ce dernier l'examine et génère automatiquement, le cas échéant, des commentaires dynamiques destinés à aider l'utilisateur dans son interprétation des données. Le résultat et les commentaires sont alors fournis à l'utilisateur par l'intermédiaire du site web.

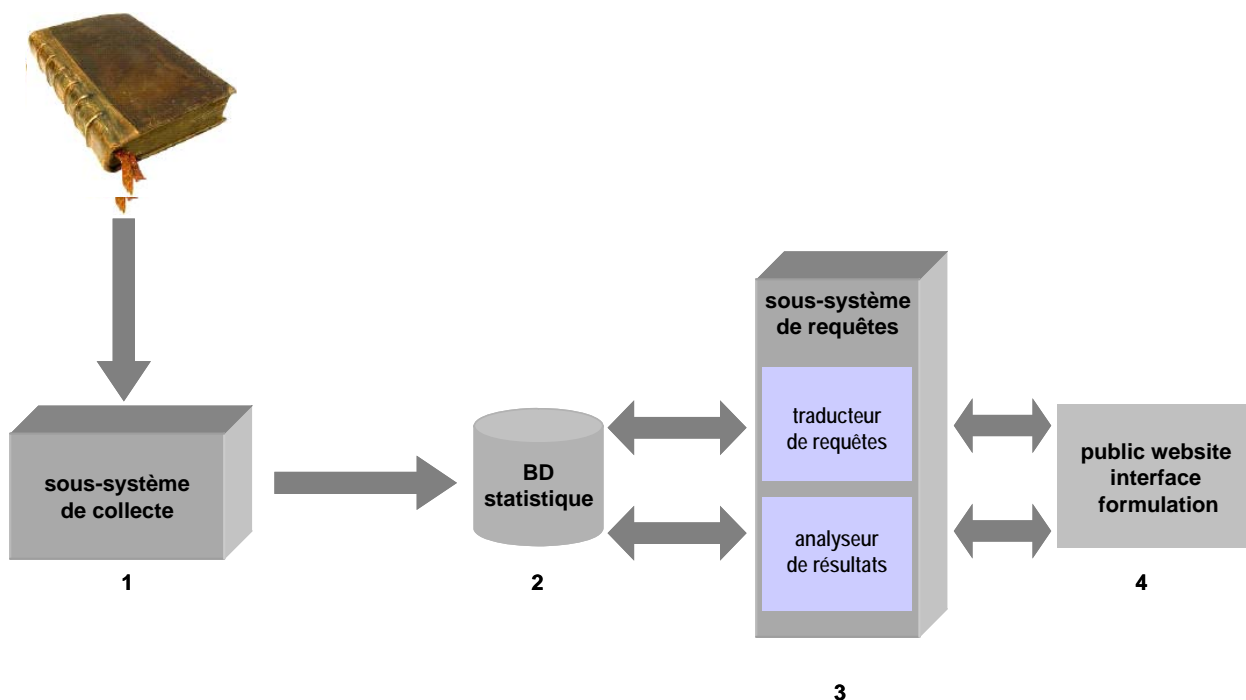


Figure 1 : Architecture du système Quetelet.net

De 2003 à 2007, le projet Quetelet.net a bénéficié d'un financement de la Politique scientifique fédérale à travers son programme "Société de l'information". Il est le fruit d'une collaboration entre plusieurs institutions :

- Etablissements scientifiques de l'Etat : Institut National de Criminalistique et de Criminologie (INCC, sous tutelle du SPF Justice), Archives générales du Royaume et Archives de l'Etat dans les Provinces (AGR, Institution à gestion séparée ressortant de la Politique scientifique fédérale)
- Administrations : Service de la Politique criminelle, Direction générale des Statistiques et de l'Information économique du SPF Affaires économiques
- Centres de recherches universitaires : Centre d'Histoire du Droit et de la Justice (UCL), Laboratoire d'Ingénierie des Bases de Données (FUNDP), Cellule Interfacultaire de Technology Assessment (FUNDP), Vakgroep Nieuwste Geschiedenis (Ugent)

L'objectif, à l'issue des quatre années de financement du projet, était d'obtenir un *prototype*. Trente années de statistiques pénales ont été introduites dans l'application, échantillonnées de manière à avoir un panel aussi varié que possible.

5. Réalisation à mettre en œuvre

Ce projet se déroulera en deux phases:

- Phase I : durant la période allant de septembre 2008 à septembre 2009, le prototype sera alimenté par les données statistiques sous format électronique (disponibles à partir de 1990) ;
- Phase II : de septembre 2009 à septembre 2010, le projet couvrira les statistiques restantes encore sous format papier.

Cette seconde phase est conditionnée au financement de la sous-traitance nécessaire pour encoder les données sous format papier. Le SPF Justice se chargera de rechercher le financement requis.

5.1. Réalisation à mettre en oeuvre par l'équipe scientifique :

phase I

L'introduction des statistiques pénales belges existant sous format informatique dans l'application. Cette tâche peut se concevoir en quatre modules avec, comme préalable, la collecte des données statistiques.

Le premier module consiste en l'encodage des structures des années concernées en ce compris les années actuelles qui ont vu se développer de nouvelles statistiques relatives au fonctionnement de la justice pénale (statistique des parquets correctionnels au niveau du Collège des procureurs généraux, statistique des condamnations au niveau du point d'appui du service de la Politique criminelle, statistiques des juridictions de la jeunesse au niveau de l'Institut National de Criminalistique et de Criminologie, ...).

Il est le point départ du second module, l'encodage des chiffres des années concernées.

Module 1 : Evolution du module d'extraction et d'intégration des données

- Evolution du module d'extraction et d'intégration. Les performances du prototype d'extraction et d'intégration des données doivent être analysées. Au terme de cette étude, des modifications éventuelles seront apportées au prototype afin d'obtenir un outil permettant une gestion optimale d'importants volumes de données.

Module 2 : Encodage des structures des années concernées

- Analyse conceptuelle des tableaux. La construction de la base de données nécessite une bonne compréhension des concepts présents dans les tableaux. La plupart d'entre eux n'ont pas été décrits lors du développement du prototype, construits sur base de données plus anciennes.

Module 3 : Encodage des chiffres des années concernées

- Développement d'une application permettant l'intégration des données électroniques dans la base de données. Pour les années les plus récentes, de nombreux tableaux statistiques ont été publiés sous forme électronique (PDF, XLS, ...). Le développement d'un système permettant d'importer les chiffres de ces tableaux dans la base de données sans passer par un encodage manuel doit permettre d'éviter des coûts de sous-traitance non seulement pour les années récentes mais également pour les années à venir.

Module 4 : Evolution du module d'accès critique

- Evolution de la typologie. La typologie, qui permet d'alimenter les commentaires critiques dynamiques, doit évoluer pour pouvoir intégrer les nomenclatures de l'ensemble des années restantes.
- Intégration de la typologie dans la base de données. Les évolutions de la typologie doivent pouvoir être implémentées dans la base de données.

- Evolution de l'analyseur de résultats. Le système qui analyse les résultats de la requête de l'utilisateur doit évoluer pour intégrer les changements dus à l'introduction des années restantes.

L'équipe scientifique se chargera de la mise en oeuvre de ces modules et de la coordination scientifique des différents aspects du projet.

Phase II

L'introduction de l'ensemble des statistiques pénales belges publiées dans l'application. Cette tâche peut se concevoir en quatre modules avec, comme préalable, la collecte des recueils statistiques et leur microfilmage / numérisation.

Le premier module consiste en l'encodage des structures des années restantes (existant sur format papier). Il est le point départ du second module, l'encodage des chiffres des années restantes.

Préalable : Microfilmage / Numérisation

- Collecte des recueils statistiques
- Microfilmage / Numérisation. Cette opération sera sous-traitée et ne fait donc pas partie de la présente demande de financement.

Module 1 : Evolution du module d'extraction et d'intégration des données

- Evolution du module d'extraction et d'intégration. Les performances du prototype d'extraction et d'intégration des données doivent être analysées. Au terme de cette étude, des modifications éventuelles seront apportées au prototype afin d'obtenir un outil permettant une gestion optimale d'importants volumes de données.

Module 2 : Encodage des structures des années restantes

- Formation des encodeurs. L'encodage des structures des tableaux demande une certaine expertise en statistiques pénales. C'est pourquoi il faudra former les jobistes qui seront engagés pour cette tâche.
- Encodage des structures. L'encodage des structures des tableaux dans l'application sera sous-traité à des jobistes. Il ne fait donc pas partie de la présente demande de financement.
- Suivi de l'encodage. Pour que l'encodage des structures soit parfaitement réalisé, il est nécessaire d'affecter un expert en statistiques pénales à la surveillance du travail des encodeurs.
- Analyse conceptuelle des tableaux. La construction de la base de données nécessite une bonne compréhension des concepts présents dans les tableaux. Si la plupart d'entre eux ont déjà été décrits lors du développement du prototype, d'autres ne manqueront pas de faire leur apparition lors de l'analyse des tableaux des années restantes.

Module 3 : Encodage des chiffres des années restantes

- Encodage. L'encodage des chiffres proprement dit ne nécessite pas d'expertise en statistiques pénales. Il sera sous-traité à une société d'encodage et ne fait donc pas partie de la présente demande de financement.
- Suivi de l'encodage. L'encodage des chiffres implique des contacts réguliers avec la société sous-traitante et la résolution des éventuels problèmes liés à l'application d'encodage.
- Examen des chiffres illisibles. Certains chiffres ne sont pas lisibles sur les photographies des tableaux utilisées par la société d'encodage. Ces chiffres devront

être examinés à la source originelle (les recueils) et il faudra dégager une attitude à adopter en cas d'illisible à cette source.

- Exécution des scripts de remplissage de la base de données. Une fois l'encodage effectué par la société d'encodage via Internet, il faut encore exécuter les scripts qui permettent à l'application d'intégrer les chiffres dans la base de données.

Module 4 : Evolution du module d'accès critique

- Evolution de la typologie. La typologie, qui permet d'alimenter les commentaires critiques dynamiques, doit évoluer pour pouvoir intégrer les nomenclatures de l'ensemble des années restantes.
- Intégration de la typologie dans la base de données. Les évolutions de la typologie doivent pouvoir être implémentées dans la base de données.
- Evolution de l'analyseur de résultats. Le système qui analyse les résultats de la requête de l'utilisateur doit évoluer pour intégrer les changements dus à l'introduction des années restantes.

Coordination scientifique

- Coordination scientifique des différents aspects du projet

5.2. Réalisation à mettre en oeuvre par l'institution publique:

Phase I

- Mise à disposition des données nécessaires au projet
- Relais vers les différents services producteurs de données statistiques en collaboration avec l'Institut National de Criminalistique et de Criminologie.
- Développement des moyens utiles pour que l'outil réponde au mieux aux attentes des différents producteurs de statistiques dans le domaine de la justice pénale.
- Mise à disposition de documents existants pouvant alimenter l'apparat critique inclus dans l'outil (en collaboration avec l'INCC).
- Organisation des comités d'accompagnement.
- Recherche des moyens nécessaires pour assurer l'achèvement du projet dans une deuxième phase.

Phase II

- Mise à disposition des données nécessaires à la phase II du projet
- Relais vers les différents services producteurs de données statistiques en collaboration avec l'Institut National de Criminalistique et de Criminologie.
- Développement des moyens utiles pour que l'outil réponde au mieux aux attentes des différents producteurs de statistiques dans le domaine de la justice pénale.
- Mise à disposition de documents existants pouvant alimenter l'apparat critique inclus dans l'outil (en collaboration avec l'INCC).
- Organisation des comités d'accompagnement.

6. Produit final attendu:

6.1. Produit final attendu de l'équipe scientifique

A l'issue de la phase I et, le cas échéant, de la phase II, le produit final attendu de l'équipe scientifique consiste en la construction d'un système critique d'intégration, de numérisation et d'accès aux statistiques pénales belges tel que décrit plus haut, comprenant dans une première phase les statistiques pénales belges existant sous forme électronique. Une seconde phase visera l'intégration de l'ensemble des statistiques pénales existantes.

6.2. Produit final attendu pour l'institution publique

Le produit final de l'équipe scientifique est intégré au sein de l'institution publique et devient un système critique d'intégration, de numérisation et d'accès aux statistiques pénales belges tel que décrit plus haut, comprenant dans une première phase les statistiques pénales belges existant sous forme électronique et dans seconde phase l'ensemble des statistiques pénales existantes.

(Rappel: l'outil ne vise en rien la production de données statistiques).

7. Valorisation du produit final:

7.1. pour l'institution publique: intégration du produit final dans son fonctionnement propre

Le prototype devenu un système opérationnel à l'issue du projet sera implémenté et fera l'objet de son maintien et de son actualisation.

Le scénario actuellement envisagé est une collaboration

- entre deux SPF : dans la deuxième phase (ultérieure) une collaboration avec le SPF Economie - Direction de la statistique pour ce qui concerne les aspects liés à l'hébergement du site.
- Avec l'Institut National de Criminalistique et de Criminologie, Etablissement scientifique de l'Etat sous la tutelle du Ministre de la Justice, pour ce qui concerne la prise en charge du *suivi scientifique de l'apparat critique* intégré dans le système.

Le produit final devrait permettre, que ce soit à l'issue de la phase I ou de la phase II, la diffusion via un site unique de l'ensemble des statistiques pénales belges disponibles, ceci avec toutes les plus-values propres à l'outil telles que précédemment décrites. Dans la première le produit vise les données les plus récentes (à partir des années nonante), et dans une phase ultérieure l'ensemble des données existantes (actuellement sous forme papier).

7.2. pour le 'public': modalités et conditions d'accès pour le monde scientifique et le grand public

Le produit final représenterait un apport fondamental pour la recherche scientifique. En permettant aux utilisateurs de formuler leurs propres requêtes pouvant porter sur plusieurs tableaux, elle leur permettrait d'économiser de très nombreuses heures de recherche et de calculs, ainsi que de mettre le doigt sur d'éventuelles évolutions sémantiques des nomenclatures.

À terme, le public non-expert aurait également accès au système. Les risques liés à la facilitation de l'accès public seront contrebalancés par l'apparat critique développé tout au long du projet.

8. Agenda et planning des tâches et budget alloué

a. calendrier des activités

Phase I : 1^{er} septembre 2008 - jusqu'au 31 août 2009

Phase II : 1^{er} septembre 2009 jusqu'au 31 août 2010

La phase II est conditionnée au financement de l'encodage des données sous papier papier à intégrer au système.

b. planification des activités

b.1. Pour l'équipe scientifique

Phase I

La planification indiquée ci-dessous donne à titre indicatif le nombre d'homme/mois (H/M) à affecter aux différentes tâches à mettre en oeuvre par l'équipe scientifique. L'équipe scientifique reste libre de proposer un phasage différent des activités dans son formulaire de soumission.

Module 1 : Développement et application d'un module permettant d'intégrer les statistiques publiées électroniquement dans notre base de données (5H/M)

- Développement du module : 3 H/M
- Application du module et suivi de l'intégration des statistiques dans la base de données : 1 H/M
- Exécution des scripts de remplissage de la base de données : 1 H/M

Module 2 : Evolution du module d'extraction et d'intégration des données (1 H/M)

Module 3 : Encodage des structures (5 H/M)

- Formation des encodeurs : 1 H/M
- Suivi de l'encodage : 3 H/M
- Analyse conceptuelle des tableaux : 1 H/M

Module 4 : Evolution du module d'accès critique (4 H/M)

- Evolution de la typologie : 2 H/M
- Intégration de la typologie dans la base de données : 1 H/M
- Evolution de l'analyseur de résultats : 1 H/M

Coordination scientifique (2H/M)

Coordination scientifique des différents aspects du projet : 2 H/M

Diagramme de Gantt

	Mois											
	1	2	3	4	5	6	7	8	9	10	11	12
Développement et application du module permettant l'intégration des statistiques publiées électroniquement												
Développement du module	■	■	■	■	■							
Application du module et suivi de l'intégration des statistiques						■	■	■	■	■	■	■
Exécution des scripts de remplissage de la base de données						■	■	■	■	■	■	■
Evolution du module d'extraction et d'intégration	■	■	■	■	■	■						
Encodage des structures												
Formation des encodeurs	■	■										
Encodage des structures		■	■	■	■	■	■					
Suivi de l'encodage		■	■	■	■	■	■					
Analyse conceptuelle des tableaux	■	■	■	■	■	■	■					
Evolution du module d'accès critique												
Evolution de la typologie	■	■	■	■	■	■	■	■				
Intégration de la typologie dans la DB									■	■	■	■
Evolution de l'analyseur de résultats	■	■	■	■	■	■	■	■	■	■	■	■
Coordination scientifique	■	■	■	■	■	■	■	■	■	■	■	■

Phase II

Pour la phase II, le planning et le timing des tâches sera mis au point avec l'équipe scientifique sélectionnée.

b.2. pour l'institution publique

Le planning de la phase I et II se calque sur l'avancement des travaux de l'équipe scientifique et contient comme moments importants :

- Organisation des comités d'accompagnements en fonction du phasage de la recherche.
- Constats avec les services producteurs de données en fonction des phases d'avancement du travail.
- Préparation du suivi ultérieur au niveau de l'apparat critique en fonction des phases d'avancement du travail.

9. Conditions particulières

a) suggestions de composition du comité d'accompagnement

- Cellule statistique du SPF Justice, Institut National de Criminalistique et de Criminologie (Département de criminologie), Analystes statistiques du collège des Procureurs généraux, Service de la Politique criminelle, SPF Economie (Direction statistique et information), Archives générales du Royaume ...
- Propositions de l'équipe de recherche et de la politique scientifique

b) clause de confidentialité éventuelle

Les statistiques pénales dont objet dans ce projet ne sont pas confidentielles et ne nécessitent pas la passation d'un contrat spécifique de confidentialité.

c) présence éventuelle de l'équipe au sein de l'institution

Sera prévu en fonction des phases d'avancement de la recherche

d) autres

La mise en oeuvre d'un tel projet par une équipe scientifique repose sur la conjonction de deux compétences : dans les statistiques pénales d'une part et dans l'ingénierie des bases de données d'autre part.