

Project description sheet issued by a federal administration

Practical information

Name of the public institution who has issued the project: Federal public Service Justice

Name of the project: INtroduction of STatistics on Penal Sentences in Electronic Form

Timing of the scientific tasks to be accomplished: Phase I : 12 months (September 2008-September 2009), Phase II : 12 months (September 2009-September 2010)

Allocated budget to the scientific team:

The two phases of the project will be implemented by the team that will be selected in phase I. The two phases lead to two separate contracts. Phase II is conditioned to the financing of the outsourcing of the encoding of the statistics that are still in paper version.

Deliverables and valorisation activities by the scientific team (not foreseen in the project description below):

- *International seminar (beginning of the project, state of the art concerning the data/work to perform) ;*
- *report* exposing the way the data concerned by the project could make international/European comparisons or harmonisation possible by the public authority;
- *Valorisation activity by the end of the project:* the team will make a proposal in this respect in its submission form.
- Deliverables from the research contract with federal Science Policy Office: *project description* (max. 25 lines in French, Dutch, English and German for the Web site of the Science Policy Office), *summary of the project* in max. 10 pages in F/D/EN/G for the web site of the Federal Science Policy Office, *a document in max. 40 lines* destined to the AGORA Newsletter at the end of the project in F/D/EN, *interims reports* (max. 2 pages, issued every 6 months destined to the financing of the project), *reports for the users' committee* (see below) ;

All these activities are to be financed within the allocated budget.

Working of the project:

The public authority steers and manages the project and collaborates closely with the team and the federal Science Policy Office (which form the technical committee of the project). The project is followed up by a users' committee who comes together at least three times in the course of the project (at the beginning, in the middle and at the end). Its role is to give a positive input to the team and the public authority in the management of the project. It is composed by representatives of other public administrations and by other researchers. The scientific team is in charge of preparing the documents for this committee, the public institution is responsible of the agenda and the minutes of the meetings.

Date of the information session with a delegate of the public institution (inscription at least a week in advance by e-mail – naji@belspo.be . without mark of interest, the meeting will be cancelled): to be yet confirmed

The project

1. Name of the Public Institution

Federal Public Service Justice

2. Project Name

INtroduction of STatistics on Penal Sentences in Electronic Form

3. Acronym

INSTAP: INtroduction of STatistics on Penal Sentences in electronic form

4. Description of the project and context within the Public Institution

Within the framework of the long-term support programme for the development of the information society (2001-2008), the Scientific Policy FPS financed a research project with the intention of working out an *information system allowing the incorporation, distribution and critical provision of statistics on existing penal sentences*. The research led to the construction of a *prototype (including 30 years of penal statistics over the 177 years of existence of the Belgian state)*.

This tool would enable the FPS Justice to evaluate the critical mass of existing statistical data capable of providing information on the functioning of penal justice both in the past and during the current period. The tool would promote access to these statistics, currently spread among a number of completely different sources, and would allow them to be distributed and used more pertinently. A single website would allow access to all the Belgian penal statistics from one source.

Nevertheless, this usage can only be envisaged if resources are invested to make the transition from the already existing *prototype to an operational product* at the FPS Justice level, which supposes the incorporation of the structures from the remaining years (particularly the most current data) and encoding of the figures for the remaining years, among other things. The object of this Agora INSTAP project is to allow the *transition from the prototype to a product incorporating all the Belgian penal statistics into the application*. **This prototype, developed within the framework of the QUETLET-NET project, is the property of the State, according to the contractual provisions of the Federal Scientific Policy. This means that it may be supplied to the team which will be selected to implement INSTAP.**

It is not the purpose of INSTAP to support the production of penal statistics, but rather to obtain a tool which, when made available to the producers of the statistical data, will allow these statistics to be better distributed and used.

Please find below a brief introduction to the Quetelet.net prototype (see also <http://www.quetelet.net/>).

Description of Existing Prototype

Quetelet.net is a project that originated from the meeting of two areas of scientific expertise. The first area of expertise is database engineering and in particular temporal databases. The second one is the expertise accumulated concerning legal and penal statistics. The objective of the project was to create a critical system for the incorporation, digitisation and access to Belgian penal statistics.

- ***Incorporation and Digitisation.*** The Belgian penal statistics, whatever the media use (paper or, more recently, electronic), have been presented invariably in the form of recognised tables for 175 years. The scientific team has developed an application allowing expert encoders (for the table structures) and non-experts (for the figures) to encode these

tables into a database. Within this application, each statistical fact is modelled in the form of a vector made up of a value for each of the dimensions that describe it, a valid time and a figure. Such an operation, which really gives rise to the structure of the tables, can prompt the end user to construct statistical queries that call up figures from more than one table. Moreover, the tables are also digitised in the form of scanned microfilms. The project thus allows statistics to be preserved, some of which are starting to degrade seriously in their paper form. Beyond that preservation function, the possibility for the user to consult the original source for each figure in the table produced by his query, at any time, is an indispensable guarantee of authenticity.

- **Access.** The Belgian penal statistics are currently spread among numerous preservation sites. No institution has the whole of the collection in its possession. To construct his or her query, the end user is assisted in his statistical research by a query builder interface allowing him to determine the time, size and values for the figures he wants to generate. The distribution of statistics is thus user-oriented and allows him or her to undertake operations which would take considerable time if the paper versions had to be consulted. This is the case, for example, for serialisation (comparison of data over time) or geographic comparison which, thanks to the Quetelet.net application, can be presented in a single table - constructed on the basis of a query - of figures which would require the consultation of a large number of paper tables.
- **Critique.** The ease of access to penal statistics, nevertheless raises a question. Public access is, in fact, capable of causing an increase in misuse and misinterpretations. That is why the website has been designed in such a way as to be able to provide numerous critique contents allowing these pitfalls to be avoided, while warning the visitor of the complexity of the data he or she is consulting. On the one hand, the site provides static critique contents: What are the Belgian penal statistics? What can you do with them? Who produces them? ... On the other hand, it provides dynamic critiques generated automatically from the user queries. A results analyser detects any semantic trends of the nomenclatures over time and brings them to the attention of the user. Thus, in the case of a serialisation query on youth offenders, for example, the trend in the age interval defining the category "youth" is automatically flagged.

In accordance with the objectives of the Quetelet.Net project, the system architecture is made up of four functional and technical sub-systems (Figure 1). The *database* storing all the information from the statistical tables constitutes the core of the system. The *data gathering sub-system* allows data to be extracted from the paper tables so they can then be incorporated into the database. The *website* is made up in particular of a query builder interface allowing the user to specify his request. Finally, the *query sub-system* forms a link between the website and the database. A *query translator* translates each query built by the user into the data definition language (here SQL) of the database management system (DBMS). Running the query provides a result sent to an *analyser*. The latter examines it and automatically generates any dynamic comments to help the user interpret the data. The results and comments are then provided to the user through the website.

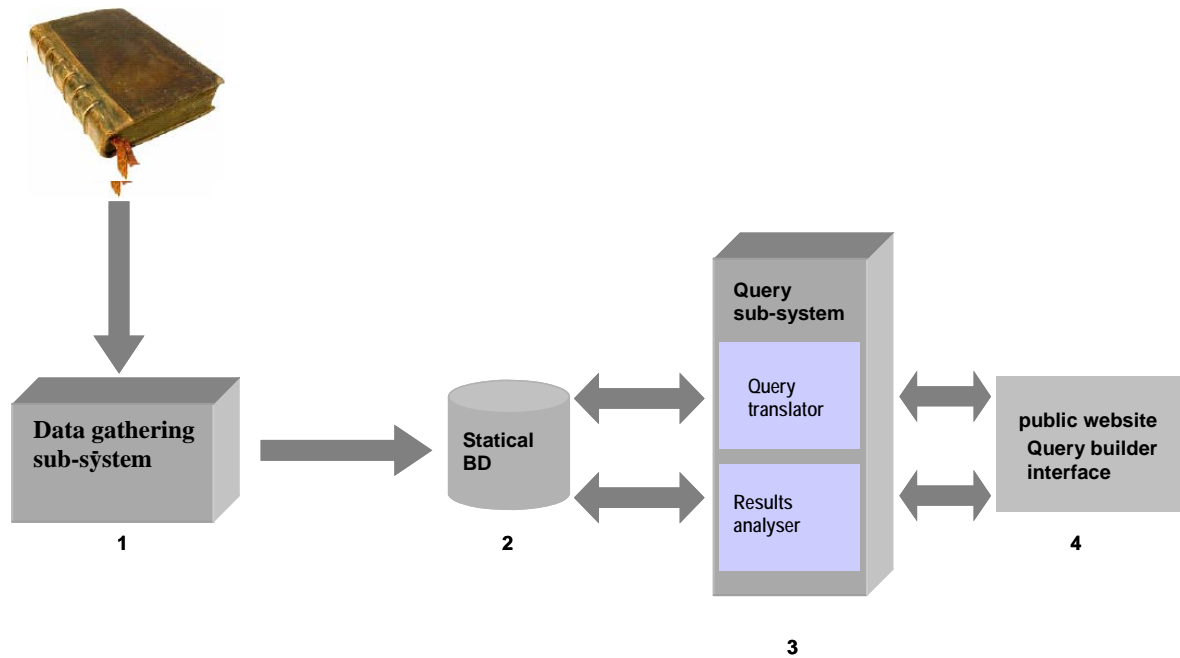


Figure 1: Architecture of the Quetelet.net System

From 2003 to 2007, the Quetelet.net project was financed by Federal Scientific Policy through its "Information Society" programme. It is the culmination of a collaboration between several institutions:

- State scientific establishments: National Institute of Forensic Science and Criminology (INCC/NICC, under the auspices of FPS Justice), National Archives and State Archives in the Provinces (institution under separate management under the Federal Scientific Policy)
- Administrative bodies: Criminal Policy Department, General Directorate of Statistics and Economic Information of the FPS Economic Affairs
- University research centres: Centre of the History of Law and Justice (Université Catholique de Louvain), Database Engineering Laboratory (FUNDP), Interfaculty Unit of Technology Assessment (FUNDP), Modern History Study Group (Ugent)

The objective, at the end of the four years of funding, was to obtain a *prototype*. Thirty years of penal statistics were entered into the application, sampled in such a way as to have as diverse a sample group as possible.

5. Tasks to be accomplished

This project will run in two phases:

- Phase I: during the period from September 2008 to September 2009, the prototype will be supplied with statistical data in electronic form (available from 1990 onwards);
- Phase II: from September 2009 to September 2010, the project will cover the remaining statistics still in paper form.

This second phase is conditional upon financing the sub-contract needed to encode the data in paper form. The FPS Justice will be responsible for putting together the required finance.

5.1. Tasks for the scientific team:

Phase I

Entering Belgian penal statistics already in computerised form into the application. This task can be divided into four modules once the statistical data have been gathered.

The first module consists of encoding the structures of the relevant years. This includes the current years which have seen new statistics develop in relation to the operation of penal justice (statistics on criminal prosecutions at the College of Principal Public Prosecutors, statistics on sentences at the base of the Criminal Policy Department, statistics on youth judgements at the National Institute of Forensic Science and Criminology, etc.).

The encoding of the figures of the relevant years is the starting point of the second module.

Module 1: Modification of the data extraction and incorporation module

- **Modification of the extraction and incorporation module.** The performances of the data extraction and incorporation module must be analysed. After being studied, any necessary modifications will be made to the prototype in order to obtain a tool allowing the optimum management of significant volumes of data.

Module 2: Encoding the structures of the relevant years

- **Conceptual analysis of the tables.** The construction of the database needs a good understanding of the concepts in the tables. Most of them have not been described during the development of the prototype, constructed on older databases.

Module 3: Encoding the figures of the relevant years

- **Development of an application allowing electronic data to be incorporated into the database.** For the most recent years, numerous statistical tables have been published in electronic form (PDF, XLS, etc.). The development of a system allowing the figures from these tables to be imported into the database without going through manual encoding should make it possible to avoid sub-contracting costs, not only for recent years but also for the years to come.

Module 4: Modification of the critique access module

- **Modification of classification.** The classification, which allows dynamic critique comments to be supplied, must be modified to allow the nomenclatures of all the remaining years to be incorporated.
- **Incorporation of the classification into the database.** It must be possible to implement the modifications of the classification in the database.
- **Modification of the results analyser.** The system which analyses the results of the user query must be modified to allow the incorporation of the changes resulting from the entry of the figures of the remaining years.

The scientific team will be responsible for implementing these modules and the scientific coordination of various aspects of the project.

Phase II

Entering all published Belgian penal statistics into the application. This task can be divided up into four modules once the statistical compendiums have been collected and microfilmed / digitised.

The first module consists of encoding the structures of the remaining years (existing in paper form). The encoding of the figures of the remaining years is the starting point of the second module.

Prerequisite: Microfilming / Digitisation

- **Collection of statistical compendiums**

- Microfilming / Digitisation. This operation will be sub-contracted and is therefore not part of the present request for finance.

Module 1: Modification of the data extraction and incorporation module

- **Modification of the extraction and incorporation module.** The performances of the data extraction and incorporation prototype must be analysed. At the end of this study, any necessary modifications will be made to the prototype in order to obtain a tool allowing the optimum management of significant volumes of data.

Module 2: Encoding the structures of the remaining years.

- **Training Encoders** Encoding table structures requires a certain expertise in penal statistics. This is why the casual workers appointed for this task will have to be trained.
- Encoding structures. The encoding of table structures in the application will be subcontracted to casual workers. It is therefore not part of the present request for finance.
- **Supervision of encoding.** To ensure that the structures are encoded accurately, it is necessary to appoint an expert in penal statistics to supervise the work of the encoders.
- **Conceptual analysis of the tables.** The construction of the database needs a good understanding of the concepts in the tables. Although the majority of them have already been described during the development of the prototype, others are guaranteed to appear during the analysis of the tables of the remaining years.

Module 3: Encoding the figures of the remaining years.

- Encoding. The encoding of the figures per se does not need expertise in penal statistics. It will be sub-contracted to an encoding company and is therefore not part of the present request for finance.
- **Supervision of encoding.** The encoding of figures involves regular contact with the subcontracting company and the solving of any problems associated with the encoding application.
- Examination of illegible figures. Some of the figures are not legible on the photographs of the tables used by the encoding company. These figures will have to be examined from the original source (the compendiums) and it will be necessary to define the approach to be adopted in the event of the illegibility of that source.
- **Running scripts and filling in the database.** Once the encoding has been done by the encoding company through the Internet, it is still necessary to run the scripts allowing the application to incorporate the figures into the database.

Module 4: Modification of the critique access module

- **Modification of classification.** The classification, which allows dynamic critique comments to be supplied, must be modified to allow the nomenclatures of all the remaining years to be incorporated.
- Incorporation of the classification into the database. it must be possible to implement the modifications of the classification in the database.
- **Modification of the results analyser.** The system analysing the results of the user query must adapt to be able to incorporate the changes resulting from the entry of the figures of the remaining years.

Scientific Coordination

- Scientific coordination of the various aspects of the project

5.2. Tasks for the Public institution:

Phase I

- Provision of data needed for the project
- Relay to the various departments producing statistical data in cooperation with the National Institute of Forensic Science and Criminology.

- Development of useful resources so that the tool best meets the expectations of the various producers of statistics in the area of penal justice.
- Provision of existing documents which may supply the apparatus criticus included in the tool (in cooperation with the INCC/NICC).
- Organisation of supervisory committees.
- Putting together the resources needed to ensure the completion of the project in a second phase.

Phase II

- Provision of the data needed for Phase II of the project
- Relay to the various departments producing statistical data in cooperation with the National Institute of Forensic Science and Criminology.
- Development of useful resources so that the tool best meets the expectations of the various producers of statistics in the area of penal justice.
- Provision of existing documents which can supply the apparatus criticus included in the tool (in cooperation with the INCC/NICC).
- Organisation of supervisory committees.

6. Expected final Product:

6.1. Expected final product from the scientific team

Following Phase I and, as the case may be, Phase II, the expected end product of the scientific team consists of the construction of a critical system for incorporating, digitising and accessing Belgian penal statistics as described above, consisting of, initially, Belgian penal statistics existing in electronic form. A second phase will set out to incorporate all the existing penal statistics.

6.2. final product expected for the public institution

The end product of the scientific team is integrated into the public institution and becomes a critical system for the incorporation, digitisation and accessing of Belgian penal statistics as described above, consisting of Belgian penal statistics existing in electronic form in an initial phase and all the existing penal statistics in a second phase.

(Reminder: it is never the objective of the tool to produce statistical data).

7. valorisation of the final Product:

7.1. For the Public institution: implementation within its own working

The prototype as an operational system at the end of the project will be implemented and will be maintained and updated.

The current envisaged scenario is a collaboration

- Between two FPSs: in the second phase (ultimate) a collaboration with the FPS Economics - Directorate of Statistics on aspects associated with accommodation of the site.
- With the National Institute of Forensic Science and Criminology, Scientific Establishment of the State under the Minister of Justice, with regard to taking responsibility for the *scientific monitoring of the apparatus criticus* incorporated into the system.

The end product should allow, either at the end of Phase I or Phase II, the distribution through a single site of all the Belgian penal statistics available, this with all the benefits inherent to the tool as described above. In the first, the product is geared towards the most recent data (from the nineties), and in a later phase all the existing data (currently in paper form).

7.2. For the 'public': conditions and means of access for the scientific world and the general public

The end product would make a fundamental contribution to scientific research. Allowing users to formulate their own queries which may span over several tables, would allow them to save very many hours of research and calculations, and allow them to detect any semantic trends in the nomenclatures.

Eventually, the non-expert public would also have access to the system. The risks associated with making access to the public easier will be offset by the apparatus criticus developed throughout the project.

8. Timing, Planning and allocated Budget

a. timing

Phase I: 1 September 2008 to 31 August 2009

Phase II: 1 September 2009 to 31 August 2010

Phase II is conditional on finance for the encoding of data on paper to be incorporated into the system.

b. Planning

b.1. For the scientific team

Phase I

The plan given below shows indicatively the number of men/month (M/m) to be appointed to the various tasks to be implemented by the scientific team. The scientific team is still free to suggest a different phasing of the activities in its submission form.

Module 1: Development and application of a module allowing the statistics published electronically to be incorporated into our database (5M/m)

- Development of the module: 3 M/m
- Application of the module and supervision of the incorporation of statistics into the database: 1 M/m
- Running scripts and filling in the database: 1 M/m

Module 2: Modification of the module for extracting and incorporating data (1M/m)

Module 3: Encoding structures (5 M/m)

- Training of encoders: 1 M/m
- Supervision of encoding: 3 M/m
- Conceptual analysis of the tables: 1 M/m

Module 4: Modification of the critique access module (4 M/m)

- Modification of classification: 2 M/m
- Incorporation of the classification into the database: 1 M/m
- Modification of the results analyser: 1 M/m

Scientific Coordination (2M/m)

Scientific coordination of the various aspects of the project: 2 M/m

Gantt Chart

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
Development and application of the module allowing the incorporation of statistics published electronically												
Development of the module												
Application of the module and supervision of the incorporation of statistics												
Running scripts and filling in the database												
Modification of the module for extraction and incorporation												
Encoding structures												
Training Encoders												
Encoding structures												
Supervision of encoding												
Conceptual analysis of the tables												
Modification of the critique access module												
Modification of classification												
Incorporation of the classification into the DB												
Modification of the results analyser												
Scientific Coordination												

Phase II

For Phase II the task schedule and plan will be fine-tuned with the selected scientific team.

b.2. For the Public institution:

The plan for Phase I and II is modelled on the progress of the work of the scientific team and contains the following key moments:

- Organisation of the supervisory committees depending on the phasing of the research.
- Findings with the departments producing data in line with the progress of the work.
- Preparation of the subsequent supervision at the level of the apparatus criticus in line with the progress of the work.

9. Special Conditions

a) Suggestions for the composition of the supervisory committee

- Statistical unit of the FPS Justice, National Institute of Forensic Science and Criminology (Department of Criminology), Statistical Analysts from the College of Principal Public

Prosecutors, Criminal Policy Department, FPS Economics (Directorate of statistics and information), National Archives, etc.

- Proposals for the research and scientific policy team

b) Any confidentiality clause

The penal statistics which are the subject of this project are not confidential and do not require the conclusion of a specific confidentiality contract.

c) Any presence of the team within the institution

Will be organised in accordance with the progress of the research

d) Others

The implementation of such a project by a scientific team relies on combining two skills: penal statistics and database engineering.