Project description sheet issued by a federal administration

This description sheet is issued by a federal administration. It describes the nature of the project it wants to implement, the tasks that should be attributed to a scientific team yet to be selected, the budget and timing devoted to the scientific team. It also describes the way the public institution will implement the end result of the scientific team both in terms of integration in its own functioning as in terms of dissemination of valuable information for researchers and the public at large.

Practical information

Name of the public institution who has issued the project:
Federal co-ordination committee for combating undeclared work and social fraud

Name of the project: Optimisation of the OASIS data warehouse

Timing of the scientific tasks to be accomplished: 24 months

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); *description sheet of the database* (in English, Dutch and French) ;

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 it 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):
15 January, Federal Science Policy Office, **2 p.m.**

The project

1. Name of the Public institution

Federal co-ordination committee for combating undeclared work and social fraud

2. Name of the project

Optimisation of the OASIS data warehouse

3. Acronym

OASIS

4. Description of the project and situation within the framework of the institution

The four federal social inspection services (coming under the jurisdictions of the FPS Social Security, the National Social Security Office, the FPS Employment, Labour and Social Dialogue and the National Employment Office) have a data warehouse at their disposal by the name of OASIS. This data warehouse contains administrative data of various federal bodies. Presently data are supplied by, amongst others, the following sources:

- DMFA (Multifunctional declaration) data on the workers' wages and working hours (NSSO)
- DIMONA (immediate declaration), provides the start and ending dates of any employment (NSSO)
- Register of building sites and sub-contractors (NSSO)
- Register of Employers (NSSO)
- Register of accounts and recoveries (NSSO)
- Unemployment register (NEO)
- Quarterly VAT declarations (FPS Finances)
- VAT clients and contractors listings

At this moment the social and fiscal data on all employers and their staff of, amongst others, the construction sector (description in french can be found on : http://www.ksz-bcss.fgov.be/fr/fluxdonnees/fluxdonnees%5F26.htm) are loaded (anonymously) into OASIS. Additionally to OASIS there is another file containing the results of investigations.

OASIS supports the social inspectors in their mission, permitting them to select enterprises for investigation and generating synthetic useful social information for each firm. An important asset of OASIS is to enable a rapid response to any possible case of fraud. The selection of firms is made using programmed "alarms" based on the experience of the social inspectors. Up till now no application (data mining) exists permitting to check, validate or improve these "alarms". Moreover, the data have up till now only been used by the inspection services themselves. The data warehouse has a unexploited potential to acquire, through scientific research, a better view on social fraud and to generate information for the policy makers in this domain.

The present project is aimed, on the one hand, at optimizing OASIS by introducing data mining and, on the other, at enabling a better view on the social fraud phenomenon. The data will therefore be put at the policy makers' and researchers' disposal. Both aims being complementary, dynamic modelling on the basis of data mining will permit a more refined description of the (new) fraud mechanisms, thus giving the inspection services , the policy makers and the interested researchers an improved picture of the phenomenon in our country.

The pursuit of these objectives will lead to a concrete result in the form of an interface convenient for several user communities, which are: the inspection services, the policy makers, the scientists and the interested public. A differentiated access to this interface will enable the user communities to consult and to exploit the useful data and results. The scientific team selected for implementing this project will have to able to cope with both objectives: i.e. it will need a thorough knowledge of

data mining techniques as well as a comprehensive view on the use that is made of OASIS by and for third parties. The team will e.g. have to determine which data and results meet the relevancy and validity criteria so as to be presented to the user communities.

5. Tasks to be accomplished

a. by the research team:

The scientific team can help optimize OASIS by amongst others:

- assessing and comparing the existing tool to internationally recognized tools or efforts, so as to enable comparisons and harmonisation;
- pointing out the social and economic phenomena emerging through the use of OASIS;
- investigating into the fraud mechanisms or into the risk factors occurring in one or several sectors of economy and detectable through OASIS;
- checking, with the use of data mining techniques how the "alarms" correlate and explaining how the observed behaviour of the "alarms" is to be understood;
- suggesting improvements to the "alarms" and presenting new "alarms" on the basis of data mining techniques;
- examining how , within a data mining context, the OASIS "alarms" can be improved by making use of data generated by investigations (requiring amongst others a definition of "positive results" and of "false" positive results);
- pointing out, through data mining, the inherent dynamism of the data, thus improving the knowledge of fraud mechanisms and the knowledge of the sector;
- pointing out the social scientific hypotheses validated by the data;
- co-operating to the exact interpretation of the data mining results;
- pointing out the possible useful extensions of the data warehouse in the field of variables, sectors and historical data on the inspection services;
- co-operating to assess validly and permanently the OASIS performances in terms of efficiency (user community inspection services), for instance by the use of data mining compatible cost functions;
- validating the relevant data and results generated on behalf of the user communities;
- developing an interface matching technically and intrinsically the user communities' needs;
- accompanying the institution during the interface's integration and implementation phases;
- -transferring useful knowledge to the inspection services through a close co-operation in the course of the project , an appropriate description of the final result and a thorough valorisation at the end of the project.

b. By the public institution:

1. *prior to the research project:*

    - taking the necessary steps so as to enable the selected research team to start immediately its data investigation work as from the start of the project;
    - investigating which steps will be required so as to enable the OASIS platform to cope with the predictable volume of new demands and of new users generated by the project;
    - examining how data mining techniques and a user communities convivial interface such as suggested by the research team are to be linked to OASIS.

2. *during the research project:*

- giving access to OASIS;
- support, orientation and implementation by the institution;
- guarantying contacts with the inspectors and the supervisors;
- integrating the research group into the appropriate working groups;
- integrating and implementing the research results into the OASIS platform as a form of interface for the various user communities;
- Implementing the required steps so as to enable the OASIS platform to cope with the increased volume of calls and users.

6. Final product to be delivered:

a. By the research team:

- Description of the results and documentation on the results will become available via the interface;
- Full description of the interface for the various user communities;
- Guidance during the integration and implementation phases of the interface.

b. For the public institution

- integration and implementation of the interface ;
- spreading of the required training material

7. Valorisation of the end product

a. For the public institution: integration of the end product into its own way of working

The public body will implement the interface in the course of the research project in collaboration with the research team.

b. for the 'public': access conditions and modalities for scientific users and for the general public

The researchers will have access to the data on the interface via a procedure that will therefore be set up. The project gives a new direction to the OASIS development, the means deployed for this, will be those provided for the OASIS development.

8. Timing, planning and allocated budget

a. timing of the activities:

1 September 2007 - 31 August 2009

b. planning of the activities:

1. for the research team

The research team will elaborate a planning with the advisory committee at the start of the project.

2. For the public institution

Development of the platform.
Providing a work space and accesses for the researchers

9. Special conditions

a. Suggestions for the composition of the support committee:

Experts from the four social inspection services familiar with the development and the use of OASIS
Members of the steering group OASIS (in addition to the reports of the advisory committee to the steering group)
FPS Employment, Labour and Social Dialogue
FPS Finances
FPS Economy

b. Clause of confidentiality and intellectual property:

The agreement with the research institutions will contain a data confidentiality clause.
The advisory committee of the project and the steering group OASIS will supervise the respect of data confidentiality.

c. Presence of the research team in the institution

The presence of the researchers at any relevant OASIS meeting is required
The researchers will have a working space at their disposal with direct access to the data warehouse
The issue whether the researches will be given a protected “extra muros” access to the data warehouse will be examined.