



To respond to questions asked by bidders

Question 6

We are preparing a project proposal following your recent call for proposals. Would it possible to have more documentation about the available databanks (Status Vector Codebook, ROE, etc.)? That would be very useful to us.

Answer 6

The documentation is not currently available.

Question 7

We would be interested in submitting a proposal as part of the Labor Market Programs Evaluation and Research Project. The form that the demand must take, however, remains rather vague for me. Could you tell me what the form and length of the bid should be and what does it have to contain?

Answer 7

You can refer to the various pages of the statement of work, respectively to pages 3, 4, 5 and 22.

Question 8

The documentation says the income tax information (T1 and T4s) is available from 1996 to 2015. We wanted to clarify whether that was irregardless of the time the EI claimant is recorded first. That is to say if somebody appears in the data in 2002, will that person's income tax information be available all the way back to 1996 (i.e. prior to being recorded as receiving EI) and all the way to 2015 (i.e. more than 5 years after first EI appearance and training)?

Answer 8

Data on T1 tax returns goes back to 1990 and for T4 supplementary records to 1999. This means that we have all the T1 tax records for every individual, **regardless of when he/she was first observed in any of the other administrative data sources**, all the way back to 1990 (or whenever the person first filed a tax return since 1990), and all the way forward to 2015 (or whenever the person last filed a tax return). The same is true of T4 supplementary records, except the earliest year for which we have records in the database is 1999. This means that the earliest year for which we will have job-related information is 1999.



Question 9

We are thinking about proposing creating a sophisticated predictive model based on machine learning and other econometric methods, which would require substantial computational power. Would it be possible to get an idea of what will be available in terms of computational power (e.g. # of cores, RAM, software available - any open source like R)?

Answer 9

We are not able to provide this information at this time. We are currently working with our colleagues at ESDC to find the best option in terms of infrastructure to store, access, process and analyse the available data. One of the options for instance, was to use a special high-powered secured server to provide access to the various elements of the data platform. We are currently thinking in terms of database management software (e.g. SQL server) as well as commonly used statistical software, such as SAS and STATA. We will be able to confirm the final possible and practical option for access the data by end of June 2018.

All other terms and conditions remain unchanged.