APPENDIX A

Customer Project Specification

1. Background

The Department for Business, Energy and Industrial Strategy (BEIS) is responsible for a wide range of policy areas in the United Kingdom, including: business; industrial strategy; labour markets; science, research and innovation; energy and clean growth; climate change.

BEIS has already done considerable work using online job vacancy data in order to monitor UK labour markets over the past year. It is seeking to acquire a new licence for a micro-level dataset of UK online job vacancies, in order to develop evidence in a number of policy areas of interest, including innovation, labour markets and economic growth. The dataset will be used both for research projects and for monitoring of labour markets.

The licence should be as broad as possible, with as few limitations as possible regarding the number and scope of its projects, as well as publication of its findings.

2. Aims and Objectives of the Project

The Department for Business, Energy and Industrial Strategy is looking to acquire a licence for a micro-level dataset of UK online job vacancies to answer a range of analytical questions.

The Department is particularly interested in the following variables (many of which will need to have been pre-extracted from the full text of the job description):

- Full text of job description;
- Job title;
- Employer name (and/or ideally a unique identifier that would allow matching the employer to other datasets);
- Salary (value or range);
- Geographical location (e.g. longitude, latitude, city, county, LEP, region, travel to work area);
- SIC code
- SOC code
- Skills requirements;
- Educational requirements;
- Timestamp;
- Flexible working arrangements / remote working;
- Contract type (full-time, part-time…)

Additionally, some extra variables may also be of value to the Department, such as data about job applicants, or coverage beyond the UK, although they are less important.
The dataset may be obtained from a single job portal or by aggregating data from several job portals (for instance, through web-scraping). The Department will place high value on the quality of the data, as per the following criteria:

- Number of observations;
- Variables of interest from the list above, taking into account the proportion of missing values;
- Representativeness of the data; although this is difficult to assess, the dataset should be as representative as possible of the UK vacancies market across a range of variables such as geography, sector and occupation;
- Time coverage, so that the dataset goes back as many years as possible, while remaining relatively consistent and comparable across time (improvements in web-