

# Sandbox Prospectus

## Introduction

The "Sandbox" was created in 2014 in the context of the Big Data project overseen by the High-Level Group for the Modernisation of Official Statistics (HLG-MOS). It provides a shared platform for statistical organisations to collaborate on evaluating and testing new tools, techniques and sources which could be useful for modern statistical analysis. The initial focus was to understand the importance of "Big Data" for official statistics, but it soon became obvious that the sandbox has many more uses.

This prospectus aims to inform you about an initiative to extend access to the sandbox beyond the current project. It outlines how your organisation can benefit from this shared asset of the international statistical community.

## What is the sandbox?

The sandbox is a state-of-the-art shared computing environment. It provides the sort of computing power that few statistical organisations can provide for themselves. It was developed in partnership with the Irish Central Statistics Office and the Irish Centre for High-End Computing (ICHEC), which hosts the sandbox.

The sandbox is a unique platform where participating organisations can engage in collaborative research activities. It is open to all producers of official statistics. The experiences and lessons learned from the sandbox will be shared with the broader official statistics community.

## What can I use it for?

The sandbox is a very flexible environment supporting various use cases, including:

- 1. Running experiments and pilots** - The sandbox can be used for experiments involving creating and evaluating new software programmes, developing new methodologies and exploring the potential of new data sources. This use case extends the current role of the sandbox beyond Big Data, and encompasses all types of data sources.
- 2. Testing** - Setting up and testing of statistical pre-production purposes is also possible in the sandbox, including simulating complete workflows and process interactions.
- 3. Training** - The sandbox can be used as a platform for supporting training courses. It can run special software for high performance computing which cannot be installed or run on standard computers. Non-confidential demonstration datasets can be uploaded and shared, facilitating shared training activities across organisations. The sandbox environment also allows statisticians opportunities for self-learning, e-learning and learning by doing.

**4. Supporting the implementation of the Common Statistical Production Architecture (CSPA)** - The sandbox can be used as a statistical laboratory where researchers can jointly develop and test new CSPA-compliant software.

**5. Data hub** - The sandbox also provides a shared data repository, subject to confidentiality constraints. It can be used to share non-confidential data sets that cover multiple countries, as well as public-use micro-data sets.

#### **What are the finance and governance arrangements?**

Based on strong interest from a number of statistical organisations, ICHEC is willing to continue to provide the sandbox after the end of the current project (December 2015), as a service to the international statistical community, on a non-profit basis. However, users will be required to pay an annual subscription to cover the costs of technical support, hardware upgrades and installation of software.

To subscribe, please send an expression of interest (a simple e-mail is sufficient) to [support.stat@unece.org](mailto:support.stat@unece.org). You will be asked to agree to standard terms of use, and you will receive an invoice for the annual subscription fee (currently 10,000 Euros) from the National University of Ireland, Galway (which provides financial administration services to ICHEC).

Each subscriber will be given a seat on the Strategic Advisory Board, a new group which will oversee sandbox operations, and collectively decide, in consultation with ICHEC, on priorities for expenditure on software and hardware. The Strategic Advisory Board will also set future subscription levels based on the number of subscribers and the desired level of service.

Any organisation producing official statistics can subscribe to the sandbox. Other organisations may be considered on a case-by-case basis subject to the approval of the Strategic Advisory Board.

#### **What will I get for my money?**

- Access to the sandbox, shared tools, datasets and other resources for your staff
- Access to international collaboration projects and opportunities
- Technical support to ensure the sandbox infrastructure is kept up to date and relevant to your needs

#### **Where can I find more information?**

- <http://www1.unece.org/stat/platform/display/bigdata/Sandbox>

We hope you will be able to join this exciting new joint venture for the international statistical community.