Sandbox

A “Sandbox” environment has been created, with support from the Central Statistics office (CSO) of Ireland and the Irish Centre for High-End Computing (ICHEC). It provides a technical platform to load Big Data sets and tools. It gives participating statistical organisations the opportunity to:

(a) Test the feasibility of remote access and processing – Statistical organisations around the world will be able to access and analyse Big Data sets held on a central server. Could this approach be used in practice? What are the issues to be resolved?
(b) Test whether existing statistical standards / models / methods etc. can be applied to Big Data;
(c) Determine which Big Data software tools are most useful for statistical organisations;
(d) Learn more about the potential uses, advantages and disadvantages of Big Data sets – “learning by doing”;
(e) Build an international collaboration community to share ideas and experiences on the technical aspects of using Big Data.

For more information, see the presentations and conclusions from the workshop where the Sandbox was launched (Dublin, 16 April 2014)

The Sandbox environment will be available for the rest of 2014 and 2015. The use of the Sandbox is managed by a task team comprising representatives of national and international statistical organisations.

The results of the 2014 Sandbox experiments are available here. The 2015 results are available here.

For 2016 and beyond, there is a strong interest from national and international statistical organisations to keep the sandbox open on a subscription basis. Please see the Sandbox Prospectus, and the options paper produced by a “sprint” session held in Cork in June. This was subsequently endorsed by the HLG Executive Board, which oversees the Big Data project.