Sandbox task team

A small sub-group has worked on defining the details of the second work package in the project proposal, namely the 'sandbox' or practical application.

The proposal outlines the work package as follows:

> ***Work Package 2: Shared computing environment ('sandbox') and practical application***
>
> This work package will form the practical element of the project. A web-accessible environment for the storage and analysis of large-scale datasets will be created and used as a 'sandbox' for collaboration across participating institutions. One or more free or low-cost, internationally-relevant datasets will be obtained and installed in this environment, with the goal of exploring the tools and methods needed for statistical production and the feasibility of producing Big Data-derived statistics and replicating outputs across countries. Simple configurations with tools and data will, whenever possible, be released in 'virtual machines' that partners will be able to download in order to test them within their own technical environments.
>
> The sandbox will be used as a platform for proving concepts in two related strands:
>
> (a) Statistics:
>
> - the possibility of producing valid and reliable statistics from novel sources, including the ability to produce statistics which correspond in a predictable and systematic way with existing 'mainstream' products, such as price statistics
> - the cross-country applicability of new analytical techniques and sources, such as the analysis of data from social networking websites. This will be done by attempting to reproduce the results of a national project in other countries
>
> (b) Tools:
>
> - the efficiency of various software tools for large-scale processing and analysis
> - the applicability of the Common Statistical Production Architecture (CSPA – under development) to the production of statistics using Big Data sources.

Members of the group working on defining the sandbox:

Name	Country/organization	email
Tomaž Špeh	Slovenia	tomaz.speh@gov.si
Carlo Vaccari	Italy	vaccari@istat.it
Martin Ralphs	UK	martin.ralphs@ons.gsi.gov.uk
Tanya Yarmola	UNECE	tanya.yarmola@gmail.com
Andrew Murray	Ireland	andrew.murray@cso.ie
Pilar Rey del Castillo Vittorio Perduca	Eurostat	pilar.rey-del-castillo@ec.europa.eu vittorio.perduca@ec.europa.eu
Marco Puts	Netherlands	m.puts@cbs.nl
Brian Studman	Australia	brian.studman@abs.gov.au

Goals of this group:

1. Define precisely what concept(s) is/are to be proved, and outline how this exercise will prove or disprove them
2. Clarify the value of the exercise in terms of the HLG and international work (i.e. define exactly how this exercise will add more than can already be provided by the projects of individual countries working alone)
3. **Explore alternative scenarios regarding choice of**
    a. **tools, environment**

    b. **datasets**
    c. **statistics to be produced**
    d. **methods of producing them**
4. Outline costings, timeline for work during 2014, etc.
5. Obtain concrete offers of funding, in-kind infrastructural support and expertise for project during 2014.
6. Draft an annex to the project proposal that summarises the work in each of the above goals.

Timeline and methods of work:

- Starting late September 2013
- Approx 6 weeks: probably weekly calls **(set a call at the same time each week if possible?)**
- Webex calls and offline commitment: approx 2-3 hours per week
- Final deliverables: a detailed specification to present as an annex to the project proposal to the HLG at their November meeting.