Project N°: 262608

ACRONYM: Data without Boundaries

DELIVERABLE D11.7

Final versions of synthetic data tools, CTA, ECTA and cell suppression tools

WORK PACKAGE 11

Improved Methodologies for Managing Risks of Access to Detailed OS Data

PROJECT START DATE:	**1st May 2011**	**DURATION: 48 Months**
DATE OF ISSUE OF DELIVERABLE:	**April 2015**	
DOCUMENT PREPARED BY:	**3, 10, 20, 29**	**URV, UPC, ULL, UoMan**

Combination of CP & CSA project funded by the European Community
Under the programme "FP7 - SP4 Capacities"
Priority 1.1.3: European Social Science Data Archives and remote access to Official Statistics

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 262608

PREFACE

This document contains a list of software tools developed under WP11. For each software package a short description is given, along with information where the sources can be found. User manuals, full descriptions can be found as (part of) other WP11 deliverables or as published articles.

LIST OF SOFTWARE TOOLS

Synthetic data tools (URV)

These are the final versions of synthetic data tools described in Deliverable D11.3[1]: MicroHybrid and Data Shuffling, with the MDAV_ID, MDAV_SWAP and IR_SWAP algorithms.
A demo of the MicroHybrid algorithm is included as well.

The software consists of a graphical user interface (GUI) programmed in JAVA. The GUI allows the user to choose the parameter K (in case of the microaggregation algorithms) and also the different kinds of variables (e.g. quasi-identifiers in MDAV_ID or MDAV_SWAP or confidentials in IR_SWAP). The process is shown in the Logger on the right side of the interface.
Apart from this graphical program, two R-packages have also been implemented.

On the DwB website, a zip-file containing the sources of the software, along with documentation on the use of the software will soon be made downloadable on the DwB project website (check Section "Data Access"). The sources of the R-routines can also be found on CRAN.

CTA (UPC)

These are the final versions of the CTA tools, described in Deliverables D11.4[2] and D11.6[3].

The CTA software is available in binary format as this is considered to be sufficient for most National Statistical Institute and/or researcher. It can be downloaded at: http://www-eio.upc.es/~jcastro/tmp/DwB/D11.4/CTAWindowsBinaries.zip

Those justifying the need of the CTA source code may obtain it upon request to UPC. Source code cannot be redistributed to third parties.

ECTA and Cell Suppression FOS software (ULL)

These are the final versions of the ECTA and Free Open Solver Cell Suppression tools, described in Deliverables D11.4[2], D11.5[4] and D11.6[3].

A compressed file containing all source codes and documentation is available as https://www.webs.ull.es/users/jjsalaza/public/ECTA&CSP.zip
It can be used in public and private organizations, even distributed, but it cannot be commercialized by any organization without previously a written agreement with the owner.
It is subject to the ZIB Academic License (http://scip.zib.de/academic.txt), like SCIP.

[1] Progress & Intermediate Report on New Masking Techniques, Record Linkage for Output Check. Available here
[2] Software Code RCTA and ECTA Implementation. Available here
[3] Reports on new synthetic data generation and CTA. Available here
[4] Software for New Masking Techniques & for New Cell Suppression Method. Available here

Record linkage (UoMan)

This is software for Record linkage using similarity scores (RL) and software for Secure string pseudonymisation (SSP). A more thorough description of these methods will be part of Deliverable D11.8[5].

The sources, along with a description on the use of the software, can be found at
https://github.com/DuncanSmith147/record-linkage
and
https://github.com/DuncanSmith147/pseudonymization
respectively.

[5] D11.8 (Final reports on synthetic data CTA, ECTA, Cell suppression and Guidelines for output checking) will be published here as soon as it is ready and available.