**Transact-SQL code script name:**

SGSS\_Extract\_Data

**Author:**

Jordan Wilson, 1/2017 ([jordan.wilson@phe.gov.uk](mailto:jordan.wilson@phe.gov.uk))

**Update history:**

1/2017 – Version 1.

2/2/2017 – Corrected and extended @MicrobiologyLaboratory\_SGSSName and @SuperOutputArea specifications.

**Description:**

Based on supplied arguments, it extracts and pre-processes data from SGSS, primarily including de-duplication and the adjustment of instances where episode duration has been exceeded.

‘Duplicates’ is defined here as records for the same patient and same episode with different specimen or record codes – these are not ‘true’ (exact) duplicates which have already been addressed in SGSS.

The process has been designed to – hopefully – be as robust and flexible as possible for epidemiological colleagues who are not routine users of SQL, Microsoft SQL Server or SGSS.

**Supplementary files in the SGSS\_Extract\_Data folder:**

* *User Guide*
* *Data definition*
* *SGSS\_Extract\_Data*
* Variant - *SGSS\_Extract\_Data\_Exceedances\_Report*
* Variant – *SGSS\_Extract\_Data\_Descriptive\_Epidemiological\_Report*

**Basic instructions:**

1. Skip down the page to ‘—ENTER PASS-IN ARGUMENTS’.
2. Enter values of interest within inverted commas. It is very important for accurate data retrieval that you correspond to specification below.

Key tips -

* If you require a ‘customised’ version of the script, contact the author.

Two variants currently exist:

* *SGSS\_Extract\_Data\_Exceedances\_Report* – Retrieves the data for a selected week of the year for when data were reported by a laboratory/ies to SGSS (contrast with specimen date).
* *SGSS\_Extract\_Data\_Descriptive\_Epidemiological\_Report* – Contains a final modified block of code to rename and reorder fields of interest and remove others.
* Text entered does not have to be case-sensitive.
* Fields containing patient-identifiable data (PID) are at the end of the series of fields, and with the ‘\_PID’ prefix for easy identification and/or removal if required.
* Please use as narrow pass-in arguments as possible, especially for date ranges. SGSS has server limitations and is a widely-used resource, therefore broad arguments are likely to cause disruptions and result in extended execution time for the process.
* Exceedance of episode length is defined as 14 days for all organisms except –
* Influenza A – 42 days.
* *Salmonella* – 91 days.
* *Mycobacterium* – 182 days.
* Creutzfeldt-Jakob disease, Hepatitis B, Hepatitis B acute, Hepatitis B chronic, Hepatitis C, HIV, Human T-cell leukemia-lymphoma virus – Continuous.

*Arguments - Specification*

|  |  |
| --- | --- |
| Argument | Requirements |
| @Organism | The organism of interest.  Mandatory if the user is interested in organism/specimen or both data sets.  Optional if the user is interested in antimicrobial data set.  Organism name can be exact or approximate. Exact is recommended as a good practice.  Examples, ‘aspergillus niger’; ‘aspergillus ni’; ‘aspergillus’. |
| @Antimicrobial | The antimicrobial of interest.  Mandatory if the user is interested in antimicrobial or both data sets.  Optional if the user is interested in organism/specimen data set.  Antimicrobial name can be exact or approximate. Exact is recommended as a good practice.  Examples, ‘penicillin’; ‘pen’. |
| @StartDate | The earliest specimen date.  Mandatory.  Format: D (numeric) MMM (character) YYYY (numeric).  Example, ‘1 jan 2017’. |
| @EndDate | The final specimen date.  Mandatory.  Format: D (numeric) MMM (character) YYYY (numeric).  Example, ‘31 jan 2017’. |
| @MicrobiologyLaboratory\_SGSSName | The name used by SGSS for the (South West) microbiology laboratory of interest.  Mandatory if the user is interested in this rather than distinction by geographical area of residence.  Optional if the user is interested in microbiology laboratory rather than geographical area of residence.  Permitted values:   * ‘derriford hosp. (plymouth)’ * ‘dorchester microbiology laboratory’ * ‘exeter microbiology laboratory’ * ‘gloucestershire royal hospital’ * ‘great western hospital’ * ‘hpa south west bristol lab’ (now discontinued) * ‘north devon district hospital (Barnstaple)’ * ‘poole microbiology laboratory’ * ‘royal bournemouth hospital’ * ‘salisbury microbiology laboratory’ * ‘severn pathology’ * ‘southmead hospital (bristol)’ [now discontinued] * ‘taunton microbiology laboratory’ * ‘torbay hospital (torquay)’ * ‘truro microbiology laboratory’ * ‘weston general hospital (weston-super-mare)’ * ‘royal united hospital (bath)’   These names are correct as of 1/2017. |
| @SuperOutputArea | The geographical area of patient residence of interest.  Mandatory if the user is interested in this rather than distinction by microbiology laboratory.  Optional if the user is interested in microbiology laboratory rather than distinction by geographical area of patient residence.  Permitted value for the South West: ‘south west’.  All other permitted values for other geographical regions of residence:   * ‘s east’ * ‘east’ * ‘scotland’ * ‘wales’ * ‘london’ * ‘n ireland’ * ‘phe region not found’ * ‘e mids’ * ‘york&hum’ * ‘n west’ * ‘channel islands’ * ‘isle of man’ * ‘w mids’ * ‘n east’ |
| @DataSet | The SGSS data set of interest.  Optional.  Permitted values: ‘org’, ‘ant’, ‘both’.  ‘org’ - organism/specimen data (the ‘CDR’ data set).  ‘ant’ - antimicrobial (‘AMR’ data set)  ‘both’ – Default.  Organism/specimen and antimicrobial are interlinked data sets in SGSS, however are reported separately by laboratories. ‘Missing’ data from one data set in SGSS but not the corresponding data from the other in SGSS might be indicative of unreported data or erroneous microbiology laboratory system logic.  Example, an antimicrobial having been reported; however not the corresponding organism this was tested on suggests a gap in the SGSS data. If in doubt, use the default ‘both’ by leaving this argument blank.  Example, ‘org’. |
| @Phagetype | The phagetype of interest.  Optional. Not available for antimicrobial data set.  Phagetype name can be exact or approximate. Exact is recommended as a good practice. |
| @Serotype | The serotype of interest.  Optional. Not available for antimicrobial data set.  Serotype name can be exact or approximate. Exact is recommended as a good practice. |
| @SpecimenType | The specimen type of interest.  Optional. Not available for antimicrobial data set.  Specimen type name can be exact or approximate. Exact is recommended as a good practice. |

1. Two output tables are produced:

* Raw – The raw data for trouble-shooting or reference by the database manager or data manager. By default, ignore this.
* Pre-processed – The transformed product.

A specimen code in the SpecimenCode column with ‘(new epi.)’ following the specimen code itself signifies that this is a further row (episode) to a row (episode) already present in the Pre-processed table. In other words, the specimen date between the original row for this patient and organism and this row for the same patient and organism exceeds 14 days (as per the definition of episode for most organisms), and has therefore not been de-duplicated as it’s by definition a new episode despite being for the same patient.

**Trouble-shooting:**

In the event of the process being cancelled or disrupted manually due to unanticipated execution time or accident, then:

* Skip down the page to ‘—TROUBLE-SHOOTING’.
* Highlight this section.
* Hit ‘execute’ or F5. This resets the process to the starting position.
* An error message will appear. Ignore this.
* Execute the script as per usual, with refined arguments in the ‘—ENTER PASS\_IN ARGUMENTS’ if required.