ICNARC Documentation

# Scripts

The scripts for cleaning the ICNARC data, creating relevant tables, building the FDM and the dashboard view are:

| **Script Name** | **Google Doc/Drive** | **BigQuery Script\*** |
| --- | --- | --- |
| icnarc\_\_00\_\_null\_the\_NULLs | [link](https://docs.google.com/document/d/187aMpwqB2ntYYoDv_ZKtycvqocDqYWAW7laKdrOF-B4/edit?usp=sharing) | [link](https://console.cloud.google.com/bigquery?sq=23854758888:ca0438309710441a87365b3cba7b4664) |
| icnarc\_\_01\_\_format\_dates | [link](https://docs.google.com/document/d/1SLMKf2Jo8skFbXq9Vw0OzeOUq7H2mgZ1ybpLnqVFiu4/edit?usp=sharing) | [link](https://console.cloud.google.com/bigquery?sq=23854758888:2fae60a20f654d39b6966423cea9917a) |
| icnarc\_\_02\_\_format\_times | [link](https://docs.google.com/document/u/3/d/1Hc2ETFaO7zPuABE7eSQe86LRtTmFB9TqPDMrm1HBcYk/edit) | [link](https://console.cloud.google.com/bigquery?sq=23854758888:0526005af82c4482b0abcc881abfccb2) |
| icnarc\_\_03\_\_format\_other\_columns | [link](https://docs.google.com/document/u/3/d/1wHwEHWvreeEuldi-UggEl8pcjEmE3XCxlHhqZuFv8Vk/edit) | [link](https://console.cloud.google.com/bigquery?sq=23854758888:147836ee57694481a80f6fe894fb19d3) |
| icnarc\_\_04\_\_estimate\_discharge | [link](https://docs.google.com/document/u/3/d/1GCHWn3f1_VH44QcQSPfEMArDSu1m_MJX-D8X9i6WkE0/edit) | [link](https://console.cloud.google.com/bigquery?sq=23854758888:58179340ba404a82bc5d30be37fae164) |
| icnarc\_\_05\_\_create\_visit\_occurrence\_table | [link](https://docs.google.com/document/u/3/d/1Bk0pInOUIbKvVk4crSmMnFGI3ynsBs7M9BIPx6EPPXQ/edit) | [link](https://console.cloud.google.com/bigquery?sq=23854758888:2b1d642b84d54fd89a1e4cae0fb231ed) |
| icnarc\_\_06\_\_rearrange\_columns | [link](https://docs.google.com/document/u/3/d/1Ei6ft21h2ugOFZnuKQCO2tzK9wHpSu1ty34eaK8zJNo/edit) | [link](https://console.cloud.google.com/bigquery?sq=23854758888:985c03742fbf44749e22a0553e4f63d5) |
| icnarc\_\_explore\_dates\*\* | [link](https://docs.google.com/document/d/1e7Ghd9eZFZVagnQyUbVK6HUR7Tw9cTr8ozyYVy4iCpY/edit?usp=sharing) | [link](https://console.cloud.google.com/bigquery?sq=23854758888:e1ef9fe3af4f425da10d3bf3bdd3cf72) |
| icnarc\_\_identify\_v2\_and\_v3\_columns\*\* | [link](https://docs.google.com/document/d/1dNjuGuzCxo5mwFYwWgkPIVbISGaKCJ-xH6ry1SQOFpI/edit?usp=sharing) | [link](https://console.cloud.google.com/bigquery?sq=23854758888:e1ef9fe3af4f425da10d3bf3bdd3cf72) |
| icnarc\_fdm\_builder.ipynb | [link](https://drive.google.com/file/d/1FFzlEiQ7BY1DcjxydGDVrkmbmVS6Xfx9/view?usp=sharing) | – |
| build\_icnarc\_dashboard\_view.ipynb | [link](https://drive.google.com/file/d/1KCAgRpr_EJcDyan3io9xx_gdmwb4ftx3/view?usp=sharing) | – |
| icnarc\_\_post-fdm\_\_finalise\_fdm | [link](https://docs.google.com/document/d/17Tsq61tdJQEw-yrb-V8qsw7Y1TsFPKa32_HfDW6bA3o/edit?usp=sharing) | [link](https://console.cloud.google.com/bigquery?sq=23854758888:3d65193bad2f4d14a87cc97921e7009e) |

\*These BigQuery scripts are set to public, so the links should work as long as the scripts themselves aren’t deleted.

\*\*These scripts are not needed to create the FDM but were used to generate information which is included in supplementary sheets within the Data Dictionary.

The scripts create tables within the dataset ‘CY\_EmW\_MYSPACE’. If this dataset is no longer available, the dataset name can be replaced in each of the scripts.

Each script begins with a brief description and set of numbered steps where appropriate. The code is also heavily annotated.

Once the **numbered** scripts have been run, the ICNARC FDM Builder script can be run where the start date is ‘**earliest\_visit\_date**’ and the end date is ‘**latest\_visit\_date**’. Ensure that the extract date is updated when relevant. Then the script to build the dashboard should be run again (note that previous versions of the dashboard views must be deleted). The dashboard itself can be found [here](https://datastudio.google.com/reporting/8ae3146f-9e47-45fb-9c3b-600d0c862869) (anyone with the link can edit).

# Tables

Note that tables are created with the naming convention ‘icnarc\_XXy\_description’, where:

* XX is the number, e.g. '01' relating to the script number
* y is a letter, e.g. 'a' relating to the order of table creation within the script

Tables are dropped when they are no longer needed, so running the scripts in succession will leave you with:

* tbl\_icnarc\_V2\_copy (preserved for posterity)
* icnarc\_05i\_care\_site\_table (not currently created - waiting for care site lookup table - see [Adding Care Sites](#_d9pyrwph1xs8))
* icnarc\_05l\_concept\_id\_table
* icnarc\_05n\_visit\_occurrence\_table
* icnarc\_06a\_data

# Admission Software Versions

There looks to be at least two versions of admission software used in the ICNARC data - version 2 and presumably version 3. I believe these are linked with the versions of the ICNARC ‘[Case Mix Programme](https://www.icnarc.org/Our-Audit/Audits/Cmp/About)’ where ICUs participate in a regular audit, thereby entering data into a standardised set of fields upon each admission.

Within the data there is a field called ‘adm\_is\_ver\_2\_adm’ (Yes/No). As can be seen in the ICNARC Data Dictionary tab ‘Admssn Ver Fields; Missingness’, there is a rough picture that different sets of columns are used when the admission is a version 2 admission or not.

I believe that when ‘adm\_is\_ver\_2\_adm’ is ‘No’, the admissions are version 3. This is because the fields that are only used in this case match with the fields in the Case Mix Programme version 3.1 [documentation](https://drive.google.com/file/d/1mzr4Pp50Y4D9LlqTi0OT_-t1TEDZtQ4A/view?usp=sharing) (See ICNARC Resources). I have not been able to find version 2 of this document (ICNARC did not reply to me when I emailed them directly), but have found a list of [now-deprecated ICNARC fields](https://drive.google.com/file/d/1JzxuXGFyqGQJcRvyUaFDcJ3dsENgNmfI/view?usp=sharing) in an appendix of a published paper (See ICNARC Resources).

From what I understand from Tom Lawton (Critical Care Consultant) and Carolyn Cartright (ICU Data Administrator), a company called Mela Solutions create admission software for ICUs called [MedICUs](https://mela.co.uk/medicus-critical-care-2/), which contains the required fields from ICNARC,

“MedICUs is a clinical Management System for Intensive Care Units, the System enables units to collect and store demographic and clinical patient information efficiently, calculating a number of different patient severity and mortality scores including APACHE II, SAPS II, PRISM (Paediatrics) POSSUM and IMP. It conforms to the ICNARC dataset specification and includes a checking routine to ensure the integrity of the data for export. In addition MedICUs automatically gathers the CCMDS dataset.” - [Mela Solutions](https://mela.co.uk/products/) (accessed 30/08/22).

I hear that the BRI ICU have recently switched to MedICUs version 4, so the ‘versions’ may not be the version of the ICNARC Case Mix Programme but the version of MedICUs software. Either way, be mindful that some columns are only populated for certain versions (see column ‘Only Recorded in Admission Software Version’ in the ICNARC Data Dictionary and the tab ‘Admssn Ver Fields; Missingness’).

# Adding Care Sites

I have not yet created the Care Site table because I am waiting to receive a care site lookup table. I have written a section of code in script 05 (steps 8 and 9) that will create the care site table and add the care sites to the visit occurrence table. Once a lookup table is created, this should be uploaded manually as ‘CY\_MYSPACE\_EmW.icnarc\_care\_site\_id\_upload’, and the script can be edited to comment out my temporary code and uncomment the care site table creation code. There is also code to copy this table to the FDM dataset in script ‘icnarc\_\_post-fdm\_\_finalise\_fdm’ - make sure to uncomment this code.

# 

# Reinstating Removed Columns

Near the end of the cleaning process, I decided to remove the following columns, in addition to ones I had deemed meaningless early on in cleaning (see post-scripts at the end of scripts 01-03 for details). I had been speaking with Tom Lawton (Critical Care Consultant) and Carolyn Cartright (ICU Data Administrator) during the cleaning progress and found that they were also unsure of what these columns meant.

If you would like to reinstate any of these columns, search in the scripts for five dashes (-----). These are placed nearby any line of code that affects one of these columns. You can likely uncomment related lines of code to reinstate the columns. Note that you will have to add them to script 06 which is hard-coded to rearrange the columns.

| # | Column | Rationale | Example values |
| --- | --- | --- | --- |
| 1 | event\_date | We do not know what event relates to. Most of the time it is the same as the unit admission date but sometimes up to 60 days later. | 4881 unique Values - Examples: 2013-03-07, 2005-09-22, 2017-02-20, 2014-01-27, 2012-03-15 |
| 2 | event\_time | As above. | 979 unique Values - Examples: 16:40:00, 18:15:00, 10:53:00, 17:00:00, 10:43:00 |
| 3 | dob\_estimated | This is only populated when the patient’s DOB is 01/01/yyyy, but to maintain anonymity, patient’s DOBs are changed to 15/mm/yyyy, so this would wrongly tell people the day of birth. | 2 unique Values: NO, YES |
| 4 | other\_intermediate\_care\_area | Presumably only filled in when the patient came from or was discharged to ‘other intermediate care area’. It’s unclear whether this is to or from and is only ever null or PCU. | 1 unique Values: PCU |
| 5 | signature\_list | We do not know what this means. | 2 unique Values: No, Yes |
| 6 | acp\_main\_hosp\_speciality | Tom Lawton said that they don’t have ACPs so this might be being used for something else. | 24 unique Values - Examples: ENT, General Medicine, Gastroenterology, General Surgery, Infectious Diseases |
| 7 | episode\_id | This is only present for 3 people. | 3 unique Values: 5.000000000, 1.000000000, 0E-9 |
| 8 | outreach\_visited\_prior\_to\_adm | This is only 0 or null. | 1 unique Values: 0E-9 |
| 9 | transfers\_information | This is a unique value for each admission so potentially could be identifying if someone had access to the admission system. | 16503 unique Values - Examples: 33110857.000000000, 33104528.000000000, 33114833.000000000, 33111696.000000000, 33109865.000000000 |
| 10 | unit\_admission\_number | We do not know what this relates to. | 385 unique Values - Examples: 20060103.000000000, 20040316.000000000, 20040588.000000000, 20180074.000000000, 20050178.000000000 |
| 11 | diagnosis\_information | This is a unique value for each admission so potentially could be identifying if someone had access to the admission system. | 16503 unique Values - Examples: 33110761.000000000, 33103663.000000000, 33114921.000000000, 33111708.000000000, 33109667.000000000 |
| 12 | mpm\_information | This is a unique value for each admission so potentially could be identifying if someone had access to the admission system. | 16503 unique Values - Examples: 33110822.000000000, 33103661.000000000, 33115022.000000000, 33111790.000000000, 33109691.000000000 |
| 13 | number\_of\_staffed\_beds | The only values are either 0 (appears 11954 times) or 12 (appears 4552 times). |  |

# FDM Problem Records

At the time of writing, the current behaviour of the FDM Builder script is to remove ‘problem records’ (e.g. FDM start date is after a person's birth date) from the FDM’d source data table and omit them from the person and observation period tables. I have been asked to return the problem records back to the FDM’d source data table, but keep them omitted from the person and observation period tables. Note that my visit occurrence table still contains the problem records - I asked whether this was an issue and was told it is not.

# 

# ICNARC Resources

The following documents are available in the Drive which will be useful to researchers using the ICNARC data.

**I strongly recommend that researchers are provided with the PDFs of these documents along with their data extract. Potentially** [**this link**](https://drive.google.com/drive/folders/1ImfRMqm0IRk3RHQc9x0aQxfxi4eMV-7K?usp=sharing) **to the containing folder could be sent.**

| **Resource** | **Description** |
| --- | --- |
| Data Collection Manual | Manual for entering data as part of the ICNARC audit. Provides a page of information for each field that has come from Version 3. Note that the fields in this manual are in alphabetical order rather than the order of collection. Also note that the Connected Bradford team did not receive all of these fields/columns. Page numbers for each field are provided in the Data Dictionary. |
| Data Flow | A chronological flow of when each Version 3 field is collected, along with a lettered code for each (provided in the data dictionary). The collection of some fields are conditional on others, e.g. questions about pregnancy only asked about women - this conditionality is documented in the Data Dictionary. Note that the Connected Bradford team did not receive all of these fields/columns. |
| Data Collection Form | A chronological flow of when each Version 3 field is collected, along with the possible values for categorical fields. Note that the Connected Bradford team did not receive all of these fields/columns. |
| List of Columns from Previous ICNARC Version | An excerpt of an appendix to a publication (2005) on the use of routine data. Outlines the ICNARC audit and lists the fields recorded in this early version of ICNARC. This was helpful in determining the meaning and order of the fields. Note that the Connected Bradford team did not receive all of these fields/columns. |

Other resources can be found here: <https://www.icnarc.org/Our-Audit/Audits/Cmp/Resources>

– Emily Williams, 23-30 Aug 2022