



# NHWDS

## Dataset build and quality statement

### National Health Workforce Dataset overview:

The Australian Health Practitioner Regulation Agency (AHPRA), in conjunction with the national boards, is responsible for the national registration process for 14 health professions. The data from this annual registration process, together with data from a workforce survey that is voluntarily completed at the time of registration, forms the National Health Workforce Data Set (NHWDS). Data in the NHWDS includes demographic and employment information for registered health professionals.

A very small proportion of practitioners complete their workforce surveys on paper rather than online. AHPRA provides the completed paper survey forms to the Department for scanning, cleansing and merging the paper-based survey data into the online registration and survey dataset.

### Survey response rates for registered and employed health professionals:

In 2015 the overall response rate for *all professions* was **92.8%**.

The 2015 survey response rates by profession are as follows:

Aboriginal and Torres Strait Islander health practice	73.8%
Chiropractic	94.9%
Chinese medicine	95.0%
Dental	94.5%
Medical	94.3%
Medical radiation practice	95.6%
Nursing and Midwifery	91.4%
Occupational therapy	93.3%
Optometry	95.2%
Osteopathy	93.7%
Pharmacy	96.5%
Physiotherapy	93.5%
Podiatry	93.1%
Psychology	97.2%

### Changes in NHWDS custodianship:

AHPRA was established in 2010 under the National Registration and Accreditation Scheme (NRAS). At the commencement of the NRAS, the Australian Institute of Health and Welfare (AIHW) was engaged as data custodian of the NHWDS. The AIHW's contract expired on 30 June 2016, and the Department assumed custodianship of the NHWDS on 1 July 2016. The Department has produced revised NHWDS releases from 2013 onwards. Both the original AIHW data and the Department's enhanced data from 2013 onwards are publicly available on the Health Workforce data tool.



These revised NHWDS releases vary somewhat from the AIHW's original versions due to minor differences in the method of imputation for survey non-response and to the Department's enhanced geocoding methods, which support a more granular analysis of the geographic distribution of the health workforce.

### **Collecting the data:**

AHPRA is responsible for the management of the NRAS and as such, they host and maintain the database that manages the scheme. They also take carriage of administering the workforce survey as part of the registration renewal processes. AHPRA creates database extracts from both the registration database and the workforce survey database to build the NHWDS.

### **Building the NHWDS:**

The Department has developed a streamlined and substantially automated process for the production of the NHWDS. User intervention is required when the data contains exceptions to define rules and to proceed through each documented step. The process consists of Extract, Transform and Load processes, geocoding of address variables, an imputation method for non-response, metadata being applied, and quality assurance checks on the produced dataset.

### **Imputation:**

Imputation is a process for filling in values missing due to non-response in order to create a complete dataset that can be used for workforce analysis and planning purposes. Imputation replaces missing values with plausible values based on other available information.

### **Geocoding:**

A large subcomponent of the cleaning process is consumed in dealing with the location data. There are multiple locations for each health practitioner and varying levels of accuracy in the raw data. The location cleaning process uses files from both AusPost (Australia Post) and the G-NAF (Geographical National Address file) to validate locations. An enhanced approach to geocoding supports more accurate analysis and modelling of the geographic distribution of the health workforce.

### **Metadata:**

An automated process was developed to focus on the collection, storage and use of metadata. Metadata for all variables in the NHWDS are stored in tables and used during the build process. These tables store variable attributes such as labels, descriptions, lengths and any codes and descriptions related to that variable.

Using the metadata tables during the build process ensures consistent metadata across datasets and years. Consistent metadata across annual releases supports improved time-series analysis, and an audit trail of changes to all values throughout



the process is maintained for transparency. This is very important for data analysis, reporting, and integration to business intelligence. It also allows data dictionaries to be generated automatically using the metadata.

**Quality assurance:**

The quality assurance of the NHWDS is detailed and contains many processes both automatic and Ad-Hoc. These processes compare annual datasets and flag any unexpected results.

**Administration:**

The cleansing, imputing, production and release of NHWDS data is an ongoing process that follows AHPRA's registration renewal cycle:

- Nurses and Midwives data – January / February
- Medical Practitioners data – February / March
- Allied Health data – March / April.

**Technical documentation:**

Detailed technical documentation on the development and production of the NHWDS is available on request.

Please email the Department via [healthworkforcedata@health.gov.au](mailto:healthworkforcedata@health.gov.au).