NAPLAN Online – Jurisdictional Results and Reporting Data Set Technical Specs (DRAFT)

File specification for extraction of results and reporting data from the National Assessment Platform

*(Based on the version 2.057 of the Results & Reporting data)*

|  |  |
| --- | --- |
| Release: | Draft |
| Date of this version: | 24-May-17 |
| Author: | NSIP team |
| Document Version Number: | 0.93 |

Contents

[Version Control 3](#_Toc483392932)

[1. Introduction 4](#_Toc483392933)

[1.1 Purpose 4](#_Toc483392934)

[1.2 Terminology 4](#_Toc483392935)

[1.3 Additional files 4](#_Toc483392936)

[1.4 Document distribution 5](#_Toc483392937)

[1.5 Background – Assessment platform 5](#_Toc483392938)

[1.6 Operations and Timing of data extracts 5](#_Toc483392939)

[1.7 Format of data extracts 5](#_Toc483392940)

[1.8 Method of data extracts 6](#_Toc483392941)

[2. XML – SIF Objects representing the Data Set 7](#_Toc483392942)

[2.1 Data scope 10](#_Toc483392943)

[2.2 Representation of events, tests, and domains 10](#_Toc483392944)

[2.3 Cardinality of objects 10](#_Toc483392945)

[3. Interface specifications 12](#_Toc483392946)

[3.1 Export Complete Results and Reporting Data for NAPLAN participants. 12](#_Toc483392947)

[3. Data Validation 12](#_Toc483392948)

[3.1 Test Data for Export 12](#_Toc483392949)

[4. Data Transport 13](#_Toc483392950)

[5. Results and Reporting Data Set – Specifications 14](#_Toc483392951)

[5.1 Student Results Export File/s– XML: 14](#_Toc483392952)

[5.2 Other resources: 14](#_Toc483392953)

# Version Control

|  |  |  |  |
| --- | --- | --- | --- |
| Results and Reporting Data Set Specification Document Version Control | | | |
| Version | Date: | Author/Organization: | Comments |
| V0.1 | 31/8/2016 | Linda Marshall and Anthony Yaremenko /NSIP | Initial Draft |
| V0.2 | 2/9/2016 | NSIP | Edits following ESA/Janison review |
| V0.3 | 7/9/2016 | Nick Nicholas, NSIP  LM AY NSIP | Updates to data transport, referenced dataset  Updates to referenced dataset, XML and data exchange infrastructure post meeting with Janison. |
| V0.4 | 22/9/2016  27/9/2016 | NSIP Team | Updates for XML based on interactions and feedback from Janison as well as the intention to support the transfer of the Codeframe.  Updates of XML after XSD creation and validation |
| V0.5 | 29/9/2016 | NSIP Team | XML structure NAPTestItem and NAPTestlet updated after feedback |
| V0.6 | 27/10/2016 | NSIP | Updated student scope statements for inclusion in dataset  Updated method of data extract  Updated sample XML provided  Updated embedded dataset to v2.02  Adjusted advice on timing of data extracts |
| V0.7 | 15/12/2016 | NSIP Team | Updated embedded dataset to v2.051  Reviewed advice and methods on data extracts |
| V0.8 | 09/02/2016 | NSIP Team – LM | Updated embedded dataset to v2.053  Link to Sample XML provided instead of the actual data.  Other minor edits |
| V0.9 | 20/2/2017  21/2/2107 | NSIP Team | Updated Diagram from NSIP Team. ESA updates to timing of extracts.  Updated embedded DataSet spreadsheet to 2.054  Added: Representation of events, tests and domains |
| V0.91 | 14/3/2017 | NSIP Team | Refined Representation of events, tests and domains |
| V0.92 | 10/4/2017 | NSIP Team | Added Dataset scope (refer section 2.1)  Added Cardinality - Student Participation Status Table and data impacts (refer section 2.3) |
| V0.93 | 24/05/2017 | NSIP Team | Clarified “0” vs NULL in student test participation status table (pg 11)  Updated html link for sample file (pg 14)  Updated reference to R&R dataset s/sheet (now at 2.057 following update to parallel test to be optional rather than mandatory to support Writing test) |

# Introduction

## 

## 1.1 Purpose

These guidelines are based on the approved NAPLAN Results and Reporting Data Set, this document also references the approved NAPLAN Online Registration Data Set and addresses:

* Elements available for exporting from the National Assessment Platform relating to the Student, the Registration of the Student into the platform, the Student’s NAPLAN test session and results as well as school, jurisdiction and national averages
* Elements used to support the production of various reports including the SSSR and ISR (but not the reports themselves)
* All elements available to authorised jurisdictional/TAA users as a result of NAPLAN assessment online for a given testing event.

## 1.2 Terminology

Table 1.1: Terminology

|  |  |
| --- | --- |
| Term | Definition |
| ACARA | Australian Curriculum and Assessment Authority |
| ESA | Education Services Australia (tasked with developing the assessment platform) |
| ISR | Individual Student Report (rocket-ship report) |
| NSIP | National Schools Interoperability Program |
| OAWG | On line Assessment Working Group |
| SIF | Systems Interoperability Framework |
| SRM | Student Registration Management System |
| SSA | School Sector Authority (examples include the Department of Education Victoria) |
| SSSR | School Student Summary Report |
| TAA | Test Administration Authority (examples include the VCAA, BoSTES, QCAA) |
| XML | eXtensible Markup Language |

## 1.3 Additional files

This document is complemented by a number of files to assist with integration development.

Files include:

|  |  |
| --- | --- |
| Filename | Description |
| Online NAPLAN Reporting DataSet 2.056.xlsx | Reference excel spreadsheet, adapted from previous versions of the Results & Reporting dataset used in NSIP’s 2016 Data Systems Readiness (DSR) consultations. |
| XSD/XML | Link for Sample XML can be found in section 6.1 |

## 1.4 Document distribution

This document will be made available via NSIP’s Github page: *https://github.com/nsip/naplan-results-reporting*

## 1.5 Background – Assessment platform

The Australian Government Department of Education is funding Education Services Australia (ESA) to develop the Online National Assessment Platform (‘the Assessment Platform’) as part of a broader program of work that supports a range of assessment types including national online assessment. One of the components of the Assessment Platform is an online-based assessment delivery system. This system will allow for test creation and management using an existing item authoring system, user management and student registration, test administration, delivery, access and presentation of tests to eligible students, as well as scoring of these assessments and reporting for assessment events. Offline and low-bandwidth delivery solutions are also included as part of the overall assessment delivery system (outside the scope of this document).

## 1.6 Operations and Timing of data extracts

The timing of jurisdictional results and reporting extracts is dependent on the platform vendor’s implementation. The generation of a complete dataset may take significant processing time. For 2017:

**For Test Administration Authorities and Jurisdictional organisations acting as a TAA in a given State/Territory:**

Authorised jurisdictional users are able to extract data from the platform once all schools have moved into the ‘Results’ phase. i.e. extraction of data requires all school principal checklists to be completed. This is of critical importance and jurisdictions extracting data will need to take this into account. Extracts can take up to 8 hours to generate, and are available once in the “Results” phase for the authorised user.

Operational guidelines regarding actual real-world timing of extracts are as follows:

* TAA accesses the Results & Reporting Dataset generation & retrieval page via the TAA dashboard
* Dataset will be accessible after all schools within the TAA’s Jurisdiction have entered the “Results” phase
* TAA generates the Results & Reporting Dataset, generating the R&R Dataset for the sector specified (within the users’ scope)
* Platform displays confirmation of generation and the password required to access the extract on screen
* TAA receives email notification upon completion of Results & Reporting Dataset generation
* TAA accesses the Results & Reporting Dataset page via the TAA dashboard and downloads the generated data extract, accepting a privacy agreement in the process
* TAA opens data extract, using the password initially provided upon generation
* If TAA wishes to re-download or re-generate the data extract at any point after this, they can do so by following the steps above. The most recently generated extract will always be available for download after its initial generation.

## 1.7 Format of data extracts

The format of the data will be XML (eXtensible Markup Language). XML is able to model the complex, repeating nested data structures produced by the assessment platform, something which flat text files (such as CSV – comma separated value files), cannot readily accomplish. XML is both human and machine readable and will allow for validation of the file contents via XSD (XML Schema Definition). Student writing script responses will be in HTML.

An open source tool, NIAS will be available to assist with conversion of XML to CSV where required. Please contact NSIP via [info@nsip.edu.au](mailto:info@nsip.edu.au) for further details on NIAS.

## 1.8 Method of data extracts

Jurisdictions will be able to request results and reporting data from the platform initially using a web based GUI (this will expand to include a REST based API in the future). Requests for results and reporting data will be able to be made for:

* All schools a jurisdiction is authorised to view, (for example all Victorian schools)
* All schools for a particular sector (school system)
* A single school
* Multiple schools (for example all schools that are part of the calibration sample)
* NOTE: If at point of generation, extracts can be generated by discipline (assessment event) and year level. Further refinement down to an individual school level at the generation point is not possible.

# XML – SIF Objects representing the Data Set



#### XML – SIF Objects representing the entities associated with NAPLAN Assessment

The data available for export via the Jurisdictional Results & Reporting data set is represented as follows:

**1. Student Personal**: Contains all of the student demographic data and identifiers related to the student. Given schools can add/modify details of students on the assessment platform, the original input registration data will most certainly be different to the final student details recorded in the platform.

**2. NAP Event Student Link:** This object represents a student’s registration for at a NAPLAN assessment. It includes key student identifiers, school identifiers, details of the test sat (including test name, level, domain), participation and exemption details, adjustments and PNPs, device details and the date of the test.

**3. NAP Test:** This object represents an assessment assigned in the context of NAPLAN, targeted at a particular domain and year level.

**4. NAP Testlet:** This object represents a testlet assigned in the context of NAPLAN, consisting of a number of items, and subject to branching rules according to adaptive testing.

**5. NAP Student Response Set:** Captures the student’s response to NAPLAN Test Items in the context of a NAPLAN test. Results are captured for each testlet, and each item within each testlet, that the student responds to. Data elements include Report exclusion flags, calibration and equating samples, domain scores, plausible scaled values for domain scores, student band, testlet details, item responses, script images where appropriate, and links to test items.

**6. NAP Test Item:** Contains the item identifier, domain, subdomain, descriptor, released status, item difficulty, proficiency band, curriculum content descriptions, stimuli and writing rubrics. The writing rubrics in turn consist of scores, and score descriptions within each score. Where the item has not been published, the object refers to the URL of the item exemplar instead. Substitute test items point back to the original item(s) that they substitute for, along with an indication of the PNP codes applicable to the substitution.

**7. School Info:** This object represents each school associated with students who are registered for NAPLAN assessment and for which the system has recorded NAPLAN test scores. It may be required to relate a Student to a school or tenancy.

**8. NAP Test Score Summary:** For a given school this object details the aggregate scores for a NAPLAN test, including national, school, and jurisdictional averages for the same test.

**9. Test Structure - CodeFrame (derived from NAP Test Item):** The full structure of a test, (known as the NAPLAN codeframe), is included in the export to jurisdictions. ACARA typically provide codeframes for the following year’s tests to jurisdictions once the tests have been finalised. For 2017, the 2017 codeframes will be established by ACARA when the tests are finalised (in early 2017). This structure is conveyed through additional fields in the NAP Test Item object, describing the testlet(s) and test(s) they belong to.

Underlying Conceptual Model is described below. The codeframe objects are given in blue.

The NAPWritingRubric objects are in a many:many relationship with NAPTestItem objects. In the implementation XML model, the Rubrics and Stimuli are contained within NAP Test Item. The DomainScores exist within the NAP Student Response Set.

*ii) Underlying conceptual model for the entities associated with NAPLAN Assessment*

## 2.1 Data scope

The dataset extract’s scope is the ‘Main test’ and does not cover the ‘Survey & Practice questions’ which are not part of the ‘Main test’.

## 2.2 Representation of events, tests, and domains

The representation of events and tests specified here is intended for consumption in the context of reporting. This representation may be at odds with the representation of the same for the purposes of test administration.

In particular, test administration of NAPLAN makes the following conflations:

|  |  |  |
| --- | --- | --- |
| **Domain** | **Event** | **Test** |
| Writing | Writing | Writing |
| Reading | Reading | Reading and  Conventions of Language |
| Spelling | Conventions of Language |
| Grammar & Punctuation |
| Numeracy | Numeracy | Numeracy |

That is, the domains of Spelling and Grammar & Punctuation are conflated in the single test of Conventions of Language; and the tests of Reading and Conventions of Language are administered within a single event (as the branching of Conventions of Language depends on the result of Reading).

For the purposes of reporting test results, however, these distinctions are irrelevant. All result objects are keyed to a single domain. The NAP Event Student Link and NAPTest objects are populated redundantly: one identical event is created for each of Reading, Spelling, and Grammar & Punctuation, and one identical test for each of Spelling and Grammar & Punctuation.

## 2.3 Cardinality of objects

Every student registered in the National Assessment Platform will have a NAP Event Student Link object generated for each NAPLAN test they are eligible for. The object is necessary for consumers to have access to, in order to indicate the status of a student in that test, whether or not they sat the test (as detailed via their student test participation status).

If a student has sat the test, and abandons the test halfway, partial results may still be published for the student with a Sanctioned Abandonment status depending on the treatment of the student as per protocols and business processes. Please refer to the below table for student participation status impacts on the data for that test:

If a student has not sat the test (they do not have a Participation Code of P: Present or S: Sanctioned Abandonment), they will not have any NAPLAN results to report for a test. In that case, the Platform will still generate an associated NAP Student Response Set object, populated with dummy data (e.g. no date, no ParallelTest pathway through nodes). The participation code in the NAP Event Student Link should be used by data consumers to indicate whether the NAP Student Response Set object can be used meaningfully or not. In NAP cycles after 2017, we expect that the dummy response sets will no longer be generated for students who have not sat the test.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Participation Code** | **Score present in R&R dataset? refer DomainScore/RawScore**  **DomainScore/ScaledScoreValue** | **SSSR impacts** | **ISR impacts** | **Jurisdictional R&R extract** |
| P: Present | Yes | Displayed on report | Available | All objects |
| A: Absent | No (NULL) | Blank result | No dot for student score | NAPStudent Response set object is dummy object, ignore |
| C: Cancelled | No (NULL) | Blank result | Not in extract if status against all student tests, in extract if affecting individual tests | NAPStudent Response set object is dummy object, ignore |
| E: Exempt | No (NULL) | Blank result | No dot for student score | NAPStudent Response set object is dummy object, ignore |
| W: Withdrawn | No (NULL) | Blank result | No dot for student score | NAPStudent Response set object is dummy object, ignore |
| S: Sanctioned Abandonment | No (NULL) | Blank result | No dot for student score | All objects, subject to the student providing at least one response otherwise NAPStudent Response set object is dummy object, ignore |
| R: Refused | Zero “0” score | Raw score of zero | Score of zero | NAPStudent Response set object is dummy object, ignore |
| X: No Longer Enrolled | No (NULL) | Student does not appear in the SSSR | Not in extract | NAPStudent Response set object is dummy object, ignore |

Note: “Blank result” for the SSSR indicates that the student’s name will show on the report but with no result.

# 3. Interface specifications

## 3.1 Export Complete Results and Reporting Data for NAPLAN participants.

|  |  |
| --- | --- |
| Business Functions Supported | Export complete set of NAPLAN Results & Reporting data for one, many or all schools. |
| Problem being solved | What is a complete record of the NAPLAN online assessment for all students across all year levels for a given year? |
| Description | See Diagram *(i)* above of all the objects involved in the exchange of Results and Reporting Data.  This describes the exporting of:   * All of the relevant, Student and NAPLAN Results information and their relationship/s. * The export will initially be a:   + manual process initiated by TAA or other authorised users via the assessment platform (XML files) |
| Process or Function Pre-conditions | 1. Data about students registering for the NAPLAN Online assessment cycle has been imported into the Assessment Delivery System from the SRM 2. The students in scope are registered to complete assessments on the platform 3. The export will be available in XML:    * 1. The XML file naming convention will NAPResultsReporting.xml      2. A file extension of .XML is required.      3. The XML will comply with the SIF AU (3.4.1) standard available [here.](http://specification.sifassociation.org/Implementation/AU/3.4.1/)      4. Students will be represented via the StudentPersonal object, and have any Participation Status from a linked NAPSession Object; Results will be in the NAPStudentResponseSet object linked with the NAPTestItem object. The NAPTestScoreSummary, representing cohort comparison scores, is linked to the SchoolInfo object, which is linked to the NAPSession. |
| Process or Function Post-conditions | Consumers of the results and reporting data will perform any necessary transformations on the data as required |
| Business rules | Refer section 1. 6 above – Operations and Timing |

# Data Validation

## Test Data for Export

Test data for testing the bulk export of NAPLAN Online Jurisdictional Results and Reporting data can be accessed via the *NSIP Hits Testing Service* from November 2016.

# Data Transport

The data format of exports of student results shall be SIF/XML. This will ensure that the complexity of results data can be captured properly, without devolving into a huge number of linked files for each result set, and that appropriate typing and validation can be built into data ingestion.

In the long term, it is highly desirable that the National Assessment Platform support the full SIF protocol, including SIF 3.x Infrastructure. However, in the short to medium term, this is not realistic—such infrastructure support will take time to put in place. Nor is it necessary, so long as results data will be available only to TAAs, who themselves do not have as yet, any established SIF capability.

For these reasons an interim measure is proposed, it is expected that the extracts will initially be made available via a manual selection interface in the National Assessment Platform, followed by the deployment of a REST based API. A list of schools that the TAA is responsible for will be presented and the TAA can select; a) one school; b) many schools; or c) all schools.

* The extract will deliver all the objects/entities via XML.
* All results objects will be wrapped within a single <NAPResultsReporting> wrapper tag.
* The results data may incorporate binary data, such as script images or exemplar items. Rather than include such data as blobs in the result data XML, it is preferable for the XML to reference such data as external files, through a URL or through a filename reference to a separately distributed file archive.
* NSIP is prepared to assist TAAs in consuming SIF formatted files into their existing systems, through its NIAS toolset.

# Results and Reporting Data Set – Specifications

Refer to *“Online NAPLAN Reporting DataSet 2.057.xlsx”* located here:

<https://github.com/nsip/naplan-results-reporting>

## Student Results Export File/s– XML:

Up to date Sample XML exists in GitHub at this location and is available for download here:

<https://github.com/nsip/naplan-results-reporting/blob/master/nap-samplefile.xml>

## Other resources:

XSD file:

<https://github.com/nsip/naplan-results-reporting/blob/master/SIF_MessageWithNAPWrapper.xsd>

NSIP NIAS tool for Results and Reporting:

https://github.com/nsip/nias2