

PBRF 2012 Census Data Definition

Status	Final
Version	5.0
Date	14 May 2012

Prepared by PBRF Solution and Data Architect Team

Document References

Document	Ref #
Performance-Based Research Fund Quality Evaluation Guidelines 2012 (Referred to as PBRF Quality Evaluation Guidelines 2012)	A389550

Changes in this version 5.0

Change	Section
1. Change Field Name PBRFEligible to WillSubmitEP and Field Title to PBRF Eligible to Will Submit EP.	3.15
2. New Notes column for WillSubmitEP field	3.15
3. Clarified Notes column for Date of First Academic Appointment field	3.19
4. Clarified Notes column for Date of Employment Contract Change field	3.20

Changes in this version 4.0

Change	Section
5. Changed reference to Provider Code to Provider Number	3.1
6. Various wording changes in Notes column for census file fields	3.1 to 3.12, 3.15, 3.17, 3.20
7. Corrected field description for Gender	3.4
8. Clarified that Ethnicity and Position Title data are mandatory for all staff members.	3.10, 3.11
9. Confirmed that a staff member cannot be employed for more than 1.0 FTE	3.12
10. Clarified that Date of Employment Contract Change is required only for PBRF eligible new and emerging researchers	3.20
11. Check that if End Date is entered for a PBRF eligible researcher, it must be within 12 months prior to the Census Date	3.14
12. Clarified that if a PBRF eligible researcher has left a TEO within 12 months of the census date, then the following data fields are not mandatory: <ul style="list-style-type: none"> • Nominated Academic Unit • New or Emerging Researcher • Date of First Academic Appointment • Date of Employment Contract Change • Subject Area of Research 	3.16, 3.17, 3.19, 3.20, 3.21

Changes in version 3.0

Change	Section
1. Clarified the definition of the census data	1.1, 2

Changes in version 2.0

Change	Section
1. Staff ID data type changed to string to recognise that some Staff IDs may contain alpha characters	3.2
2. Clarified that the NSN for PBRF-eligible researchers must exist in the NSI (National Student Index) as a master record, and that NSN plus date of birth in the census data must match the NSI data	3.3
3. End Date is an optional field (not mandatory)	3.14

4. Clarified that Date of First Academic Appointment is required only for PBRF-eligible new and emerging researchers	3.19
5. Updated list of SAC funded degree delivering TEOs	Appendix A1

Table of Contents

1	INTRODUCTION.....	1
1.1	DOCUMENT PURPOSE.....	1
1.2	AUDIENCE.....	1
1.3	DOCUMENT SCOPE.....	1
1.4	TERMINOLOGY & CONVENTIONS.....	1
2	CENSUS FILE STRUCTURE OVERVIEW.....	2
3	CENSUS SPECIFICATION DETAILS.....	4
3.1	PROVIDER NUMBER.....	4
3.2	STAFF ID.....	4
3.3	NATIONAL STUDENT NUMBER.....	4
3.4	GENDER.....	5
3.5	DATE OF BIRTH.....	5
3.6	TITLE.....	5
3.7	FIRST NAME.....	5
3.8	MIDDLE NAME.....	6
3.9	FAMILY NAME.....	6
3.10	ETHNICITY.....	6
3.11	POSITION TITLE.....	6
3.12	FULL TIME EQUIVALENT.....	7
3.13	START DATE.....	7
3.14	END DATE.....	7
3.15	WILL SUBMIT EP.....	8
3.16	NOMINATED ACADEMIC UNIT.....	8
3.17	NEW OR EMERGING RESEARCHER.....	9
3.18	PREVIOUS PROVIDER ID.....	9
3.19	DATE OF FIRST ACADEMIC APPOINTMENT.....	9
3.20	DATE OF EMPLOYMENT CONTRACT CHANGE.....	10
3.21	SUBJECT AREA OF RESEARCH.....	10
	APPENDIX A.....	11

1 Introduction

1.1 Document purpose

This document describes the fields that make up the Census data schema for the PBRF 2012 Quality Evaluation.

The Census submitted by tertiary education organisations (TEOs) must include details of all staff members employed or contracted for services (both those PBRF eligible and those not eligible) between 15 June 2011 and 14 June 2012.

1.2 Audience

The audience for this document includes:

- TEO developers involved in the development of systems to supply Census to the TEC.
- TEO business users who need to understand the data items to export using human resource systems.

While this document has been written with a non-technical audience in mind it does assume a basic understanding of comma separated value format.

1.3 Document scope

This document describes the data elements contained within a census, their meaning, data types, and basic validation rules.

1.4 Terminology & conventions

This document describes data to be supplied in a comma separated value (CSV) format. Some basic terminology and conventions used in this document are:

- Data Field – an item of data to be supplied is known as a “field” e.g. LastName and FirstName are examples of fields.
- Comma & Quote delimited – Field values are normally enclosed within an open and close double quote (e.g. “field value”) and each field is separated by a comma (“field1”, “field2”, “field3”).

Using quote marks allows the use of more complex data values where commas are part of the data value and not to be confused with the comma to separate fields (e.g. “Field1”, “field, 2, value”, “field3”). The value for the second field in the previous example is “field, 2, value” and will be interpreted as a single value. Empty quotes (e.g. “”, “field2”) will be interpreted as a null or an empty or not supplied value.

- Data Types – This document lists the type or format of data required for a given data field (e.g. string, integer, date etc.).
- Cardinality – Describes how many instances of an element must be supplied
 - 0 – Optional, can only supply 1
 - 1 – Mandatory, must supply a single value

2 Census File Structure Overview

The Census file structure is designed to allow a TEO to submit information about all staff members employed or contracted for services (both those PBRF-eligible and those not eligible) between 15 June 2011 and 14 June 2012. Information about researchers is required by the TEC as part of the PBRF process while information about other staff members is captured for use by the Ministry of Education.

The main reasons for the Census data to be CSV are as follows:

- Census data was provided in the last round of assessment in this format
- Census data can be easily extracted from HR systems and exported to flat file structure formats like comma separated value (CSV).

The design of the Census structure and processing assumes that the Census file:

- Is for a single TEO only
- Must contain at least one staff member but may contain many
- Will not contain a header row and will provide the fields in the order listed in the table below.

The table below summarises the basic structure of a Census file:

Field Position	Field Name	Field Title	Section
1	ProviderNumber	Provider Number	3.1
2	StaffID	Staff ID	3.2
3	NSN	National Student Number	3.3
4	Gender	Gender	3.4
5	DOB	Date of Birth	3.5
6	Title	Title	3.6
7	FirstName	First Name	3.7
8	MiddleName	Middle Names	3.8
9	LastName	Family Name	3.9
10	Ethnicity	Ethnicity	3.10
11	PositionTitle	Position Title	3.11
12	FTE	Full Time Equivalent	3.12
13	StartDate	Start Date	3.13
14	EndDate	End Date	3.14
15	WillSubmitEP	Will Submit EP	3.15
16	NominatedAcademicUnit	Nominated Academic Unit	3.16
17	NewEmergResearcher	New or Emerging Researcher	3.17

Field Position	Field Name	Field Title	Section
18	PrevInstitID	Previous Provider ID	3.18
19	AcademicAppointDate	Date of First Academic Appointment	3.19
20	DofEmpContractChange	Date of Employment Contract Change	3.20
21	SubjectArea	Subject Area of Research	3.21

The following string is an example of how data may be supplied given the CSV format and field positions described above.

The data is completely arbitrary and is only used to demonstrate an example of a comma quote separated value in relation to the field names described above.

Example:

```
"7001","STFHV1","1234567890","M","19600115","Prof","Alex","Anthony Bat", "Ballock", "100",
"Director of Materials Research" ,"0.50", "20080810" ,"20090810" ,"Y" ,"APMSC" ,"N" ,"7005"
,"19900901" ,"20010901" ,"210"
```

The following example of data is where:

- the LastName field contains a single quote character "O'Reilly"
- the PositionTitle field value contains a comma character *"Director of Materials, Research"*

Example:

```
"7001","STFHV1","1234567890","M","19600115","Prof","Alex","Anthony Bat", "O'Reilly", "100",
"Director of Materials, Research" ,"0.50", "20080810" ,"20090810" ,"Y" ,"APMSC" ,"N" ,"7005"
,"19900901" ,"20010901" ,"210"
```

3 Census Specification Details

3.1 Provider Number

Provider Number is a unique number that identifies a TEO. This is also known as the EDUMIS number.

Field Name	Cardinality	Data Type	Notes
ProviderNumber	1	Number 4 characters	The field is used to identify TEOs. An empty value will result in an error and the record will not be processed. See Appendix A1 for a list of valid codes.

3.2 Staff ID

This is the identifier used by the TEO to reference the staff member. This may be an employee number or staff ID.

Field Name	Cardinality	Data Type	Notes
StaffID	0	string 10 characters	The field is used by the TEO to identify staff members. This field will be used as a reference field only. It may be useful in the event of a processing query for the TEC to be able to quote this when talking to a TEO.

3.3 National Student Number

A unique number issued by the National Student Index (NSI) that identifies a researcher who is shown as submitting an EP.

Field Name	Cardinality	Data Type	Notes
NSN	0	number 10 characters	The field is used by the PBRF IT system to identify staff members. Must be provided if WillSubmitEP = "Y" as the NSN will be used as the primary mechanism to match researchers in the Census to those who have Evidence Portfolios. NSNs are not required for staff who are not submitting Evidence Portfolios. The TEC will use this number to search for the researcher in the National Student Index and check that the date of birth matches and that the NSI record found is a master record.

3.4 Gender

A code identifies a staff member's gender.

Field Name	Cardinality	Data Type	Notes
Gender	1	string 1 character	<p>The field is used to identify gender of staff members.</p> <p>When analysing staff data the TEC will be able to produce summaries broken down by gender, e.g. staff in various subject areas by gender.</p> <p>Any value other than M or F will default to U (Unknown).</p> <p>See Appendix A2 for a list of valid codes.</p>

3.5 Date of Birth

The day, month, and year of birth of the staff member.

Field Name	Cardinality	Data Type	Notes
DOB	1	date 8 character	<p>The field is used in conjunction with the NSN to identify staff members.</p> <p>The field will be used in combination with the NSN to check that the NSN supplied matches the correct record on the National Student Index (NSI).</p> <p>The field will also be used by the TEC to report by age, e.g. staff in various subject areas by age. The age of a staff member is calculated as at 31 July of the return year.</p> <p>See Appendix A6 for valid format.</p>

3.6 Title

Title of the staff member.

Field Name	Cardinality	Data Type	Notes
Title	0	string 255 characters	e.g. Mr., Ms., Prof., Dr., etc.

3.7 First Name

The first or forename for the staff member.

Field Name	Cardinality	Data Type	Notes
FirstName	1	string 255 characters	<p>The field is used to identify staff members.</p> <p>An empty value will result in an error and the record will not be processed.</p>

3.8 Middle Name

This field consists of the staff member's middle, second forename or alternate name.

Field Name	Cardinality	Data Type	Notes
MiddleName	0	string 255 characters	The field is used to identify staff members.

3.9 Family Name

This field consists of the staff member's surname (also known as last name or family name)

Field Name	Cardinality	Data Type	Notes
LastName	1	string 255 characters	The field is used to identify staff members.

3.10 Ethnicity

This field allows the TEC to monitor staff members from different ethnic groups in New Zealand.

Field Name	Cardinality	Data Type	Notes
Ethnicity	1	string 3 characters	<p>Mandatory for all staff.</p> <p>The codes identify the ethnic group or groups to which a staff member belongs.</p> <p>From the ethnic groups specified, a single ethnicity is determined for each staff member by a system of priority recording based on the Statistics New Zealand standard classification of ethnicity.</p> <p>See Appendix A3 for a list of valid codes.</p>

3.11 Position Title

This field consists of the employment position held by the staff member at the TEO.

Field Name	Cardinality	Data Type	Notes
PositionTitle	1	string 50 characters	Mandatory for all staff. This field is used to capture the staff member's employment or contracted position in their organisation.

3.12 Full Time Equivalent

This field consists of the full-time equivalent (FTE) that the staff member is employed in the position at the TEO.

Field Name	Cardinality	Data Type	Notes
FTE	1	decimal 4 characters	<p>The FTE status for part-time staff is the percentage of full-time employment (to two decimal places). For example a part-time role of 10 hours per week would be 25% (or 0.25 FTE) if the full-time role is 40 hours per week.</p> <p>The decimal point must be supplied. 10% would be 0.10 and 100% would be 1.00</p> <p>A staff member cannot be employed for more than 1.00 FTE.</p>

3.13 Start Date

This field consists of the date on which employment of the staff member commenced at the TEO.

Field Name	Cardinality	Data Type	Notes
StartDate	1	date 8 characters	<p>If the date of commencing employment (or contract) is in the last year this may affect the level of PBRF funding that can be obtained.</p> <p>See Appendix A6 for valid format.</p>

3.14 End Date

This field consists of the employment end date of the staff member at the TEO.

Field Name	Cardinality	Data Type	Notes
EndDate	0	date 8 characters	<p>If a PBRF eligible staff member leaves within the year prior to the PBRF Census, the TEO may be able to obtain some benefit from the period of service that was conducted in the year. See Chapter 2 Section B of <i>PBRF Quality Evaluation Guidelines 2012</i> for more information.</p> <p>In this situation, if an End Date is entered (it must fall within the 12 month period prior to the census date) and if WillSubmitEP = "Y", the following fields do not need to be entered:</p> <ul style="list-style-type: none"> • Nominated Academic Unit • New or Emerging Researcher • Date of First Academic Appointment • Date of Employment Contract Change • Subject Area of Research <p>See Appendix A6 for valid format.</p>

3.15 Will Submit EP

This field consists of a flag to identify staff for which the TEO may submit an Evidence Portfolio.

Field Name	Cardinality	Data Type	Notes
WillSubmitEP	1	string 1 character	<p>TEOs must determine which staff, including contractors, are eligible to participate in the 2012 Quality Evaluation (PBRF-eligible) and then submits those Evidence Portfolios they have determined could fulfil the requirements for a funded Quality Category.</p> <p>The TEC PBRF staff-eligibility criteria are detailed in Chapter 2 Section B of <i>PBRF Quality Evaluation Guidelines 2012</i>.</p> <p>This flag identifies all the staff for which the TEO may submit an Evidence Portfolio.</p> <p>See Appendix A4 for a list of valid codes.</p>

3.16 Nominated Academic Unit

This field consists of the employment position held by the PBRF eligible staff member at the TEO.

Field Name	Cardinality	Data Type	Notes
NominatedAcademicUnit	0	string 50 characters	<p>Mandatory if WillSubmitEP = "Y" and EndDate is not entered, otherwise optional.</p> <p>This field will be used by the TEC for reporting of the 2012 PBRF Evidence Portfolio Peer Review Panel quality categories. This is one of the four levels of quality reporting that the TEC will provide both to TEOs and publicly. It will be used to aggregate the results of PBRF eligible staff members within the same nominated academic unit at a TEO, including those not submitted for external panel evaluation.</p> <p>Actual choice of unit name is at the discretion of the TEO but should not contain the TEO name itself. Because it is used to aggregate results it is essential the TEOs understand the need for consistent use of the same name for any one Nominated Academic Unit – similar names or instances of wording reversals will be reported separately. Where a staff member is working jointly in more than one Nominated Academic Unit, the TEO should choose only one – the one where most time is spent – or which is most representative of the research undertaken. Wherever possible a single unit should not represent less than five staff members as this places the confidentiality of the staff members' quality categories at risk.</p>
NominatedAcademicUnit			<p>Data in this field could be Faculty, School or Dept Name e.g.</p> <ul style="list-style-type: none"> • Department of History • School of Music • Applied Mathematics Research Unit

3.17 New or Emerging Researcher

This field consists of a flag to identify staff who meet the criteria as a 'New and Emerging Researcher' for PBRF consideration.

Field Name	Cardinality	Data Type	Notes
NewEmergResearcher	0	string 1 character	<p>Mandatory if WillSubmitEP = "Y" and EndDate is not entered, otherwise optional.</p> <p>This flag denotes that a PBRF-eligible staff member meets the criteria as a 'New and Emerging Researcher' for the purposes of PBRF assessment. See Chapter 3 Section E of <i>PBRF Quality Evaluation Guidelines 2012</i> for more information about new and emerging researchers.</p> <p>See Appendix A4 for a list of valid codes.</p>

3.18 Previous Provider ID

A four-digit code assigned by the Ministry of Education to uniquely identify a TEO which has merged with another TEO.

Field Name	Cardinality	Data Type	Notes
PrevInstitID	0	string 4 characters	<p>The code is used for reporting purposes within the PBRF process, to identify significant organisational divisions within a TEO which have resulted from TEO merger.</p> <p>See Appendix A1 for a list of valid codes.</p>

3.19 Date of First Academic Appointment

The day, month, and year of first academic or research appointment, worldwide on or after 1 January 2006.

Field Name	Cardinality	Data Type	Notes
AcademicAppointDate	0	date 8 characters	<p>Mandatory if WillSubmitEP = "Y" and NewEmergResearcher = "Y" and EndDate is not entered, otherwise optional.</p> <p>Date must be on or after 1 January 2006.</p> <p>Required by TEC to confirm PBRF eligibility as a New and Emerging Researcher.</p> <p>See Appendix A6 for valid formats.</p>

3.20 Date of Employment Contract Change

The day, month, and year the staff member's employment contract change to include a requirement to undertake research or degree level teaching on or after 1 January 2006.

Field Name	Cardinality	Data Type	Notes
DofEmpContractChange	0	date 8 characters	<p>Mandatory if WillSubmitEP = "Y" and NewEmergResearcher = "Y" and EndDate is not entered, otherwise optional.</p> <p>Date must be on or after 1 January 2006.</p> <p>If the researcher has been determined to be New and Emerging because they were first appointed to a PBRF-eligible position or equivalent position on or after 1 January 2006, TEOs should enter the same date entered in the AcademicAppointDate field.</p> <p>Required by TEC to confirm PBRF eligibility as a New and Emerging Researcher.</p> <p>See Appendix A6 for valid formats.</p>

3.21 Subject Area of Research

This field consists of the staff member's subject area of research.

Field Name	Cardinality	Data Type	Notes
SubjectArea	0	string 3 characters	<p>Mandatory if WillSubmitEP = "Y" and EndDate is not entered, otherwise optional.</p> <p>See Appendix A5 for a list of valid codes.</p>

Appendix A

A.1 SAC funded degree delivering TEOs

SAC funded degree delivering TEO Codes	
Code	Description
Polytechnics	
6003	Bay of Plenty Polytechnic
6006	Christchurch Polytechnic Institute of Technology
6007	Eastern Institute of Technology
6010	Manukau Institute of Technology
6011	Nelson Marlborough Institute of Technology
6012	Northland Polytechnic
6022	Open Polytechnic of New Zealand
6013	Otago Polytechnic
6015	Southern Institute of Technology
6004	Unitec New Zealand
6009	Universal College of Learning
6018	Waiariki Institute of Technology
6019	Waikato Institute of Technology
6008	Wellington Institute of Technology
6017	Western Institute of Technology Taranaki
6014	Whitireia Community Polytechnic
Universities	
7008	AUT University
7006	Lincoln University
7003	Massey University
7001	University of Auckland
7005	University of Canterbury
7007	University of Otago
7002	University of Waikato
7004	Victoria University of Wellington
Wananga	
8630	Te Wananga O Aotearoa
9241	Te Wananga O Raukawa
9386	Te Whare Wananga O Awanuiarangi
Private Training Establishments	
8530	AIS St Helens
9242	Anamata
8694	Bethlehem Institute of Education
8979	Carey Baptist College
8717	Good Shepherd College – Te Heparā Pai
8550	International Pacific College, NZ
8563	Laidlaw College Inc

SAC funded degree delivering TEO Codes	
Code	Description
8396	New Zealand College of Chiropractic
8619	New Zealand Tertiary College
8457	Pacific International Hotel Management School
8502	Te Kura Toi Whakaari O Aotearoa - New Zealand Drama School
8182	Te Whare Wananga O Te Pihopatanga O Aotearoa
8509	Whitecliffe College of Arts and Design

A.2 Gender Codes

PBRF Gender Codes	
Code	Description
M	Male
F	Female
U	Unknown

A.3 Ethnicity Codes

PBRF Ethnicity Codes	
Code	Description
100	European nfd
111	New Zealand European
121	British and Irish
122	Dutch
123	Greek
124	Polish
125	South Slav
126	Italian
127	German
128	Australian
129	Other European
211	Māori
300	Pacific Peoples nfd
311	Samoan
321	Cook Islands Maori
331	Tongan
341	Niuean
351	Tokelauan
361	Fijian
371	Other Pacific Peoples
400	Asian nfd
410	Southeast Asian nfd
411	Filipino
412	Cambodian
413	Vietnamese

PBRF Ethnicity Codes	
Code	Description
414	Other Southeast Asian
421	Chinese
431	Indian
441	Sri Lankan
442	Japanese
443	Korean
444	Other Asian
511	Middle Eastern
521	Latin American
531	African
611	Other Ethnicity
944	Don't Know
955	Refused to Answer
966	Repeated Value
977	Response Unidentifiable
988	Response Outside Scope
999	Not Stated

A.4 Simple Decision Codes

Simple Decision Codes	
Code	Description
Y	Yes
N	No

A.5 Subject Area Codes

Subject Area Codes	
Code	Description
10	Māori Knowledge and Development
20	Law
30	History, History of Art, Classics and Curatorial Studies
40	English Language and Literature
50	Foreign Languages and Linguistics
60	Philosophy
70	Religious Studies and Theology
80	Political Science, International Relations and Public Policy
90	Psychology
100	Human Geography
110	Sociology, Social Policy, Social Work, Criminology & Gender Studies
120	Anthropology and Archaeology
130	Communications, Journalism and Media Studies
140	Education
150	Chemistry

Subject Area Codes	
Code	Description
160	Physics
170	Earth Sciences
180	Molecular, Cellular and Whole Organism Biology
190	Ecology, Evolution and Behaviour
200	Agriculture and Other Applied Biological Sciences
210	Pure and Applied Mathematics
220	Statistics
230	Computer Science, Information Technology, Information Sciences
240	Engineering and Technology
250	Architecture, Design, Planning, Surveying
260	Biomedical
270	Clinical Medicine
280	Public Health
290	Nursing
300	Dentistry
320	Veterinary Studies and Large Animal Science
330	Sport and Exercise Science
340	Other Health Studies (including Rehabilitation Therapies)
350	Management, Human Resources, Industrial Relations and Other Businesses
360	Accounting and Finance
370	Marketing and Tourism
380	Economics
390	Music, Literary Arts and Other Arts
400	Visual Arts and Crafts
410	Theatre and Dance, Film, Television and Multimedia
420	Design
430	Pharmacy

A.6 Typical Date Format

Typical Date Format	
The format of all dates requested will be: CCYYMMDD	
Example: 12 March 1965 will be formatted "19650312"	
Code	Description
CC	Century Ranges from 19 to 21
YY	Year ranges from 00 to 99. Please note that single digit MUST be prefixed with 0. e.g. 01 not just 1
MM	Month ranges from 01-12. Please note that single digits MUST be pre-fixed with a 0. e.g. 01 not just 1
DD	Day ranges from 01 to 31. Please note that single digits MUST be pre-fixed with a 0. e.g. 01 not just 1.