

Annex A: Pilot data specification

Evidence requires three data tables each covering the relevant Units of Assessment:

- Table 1 Staff specification with unique staff ID
 - This should ideally include all staff and research students associated with research outputs in all of the relevant Units of Assessment
 - This **must** include all research staff that were eligible to be submitted to the 2008 RAE (complete for Cat A and Cat B staff, and including Cat C and D staff where the data are readily available) in all of the relevant Units of Assessment.
- Table 2 Output specification with unique output ID
 - This should ideally include all research outputs in the relevant Units of Assessment for all staff included in Table 1, published between 1 Jan 2001 and 31 Dec 2007.
 - This **must** include RAE2008 submitted items plus all RAE eligible journal articles by Cat A and C staff included in the pilot (and any known RAE-eligible journal articles by Cat B and D staff) in the relevant Units of Assessment, published between 1 Jan 2001 and 31 Dec 2007.
 - This will provide a broader coverage than just Thomson articles, and ideally include conference proceedings and other non-journal outputs.
- Table 3 linking named staff to outputs

Table 1 STAFF specification

Field	Type	Expected values	Mandatory
Institution	String	Your HESA institution code	Yes
HESASTaffIdentifier	String	<13 characters>	Yes (for Cat A and B staff)
InstitutionalStaffIdentifier	String	<24 characters>	Yes
UnitOfAssessment	Integer	1-67	Yes
Unit, department, school or other location ¹	String	<Unlimited>	
Title	String	<50 characters>	
First name	String	<100 characters>	
Other first-names	String	<500 characters>	
Initials	String	<50 characters>	Yes
LastName	String	<500 characters>	Yes
Alternative or prior surname(s)	String	<500 characters>	
Alias or 'known as' for publications	String	<100 characters>	
e-mail address ²	String	320	Yes
Eligible for the 2008 RAE	Boolean	True/false	Yes

¹ This information will be used to assist in identifying papers on the commercial database(s).

² E-mail address either for the individual, or another person responsible for verifying the individual's outputs.

Field	Type	Expected values	Mandatory
Selected for the 2008 RAE	Boolean	True/false	Yes (for all RAE eligible staff)
RAE staff category	Character	A, B, C, D	Yes (for all RAE eligible staff)
Early career researcher ³	Boolean	True/false	Yes
Start date (if started after 2001-01-01)	Date	2001-01-01 to 2007-10-31	Yes (if recruit)
Prior institution (if started after 2001-01-01)	String	HESA institution code	
Leave date (if left before 2007-10-31)	Date	2001-01-01 to 2007-10-31	Yes (if leaver)
Destination institution	String	HESA institution code	

³ An early career researcher is defined as an individual of any age who first entered the academic profession on employment terms that qualified them for submission to RAE2008 as Category A staff on or after 1 August 2003

Table 2 OUTPUT specification

Field	Type	For OutputType = 'D'	Expected values	RA2 field	Mandatory
Institution	String		HESA institution code (including campus code if applicable)	Yes	Y
InstitutionalUniqueOutputId	String		<24 characters>	Yes	Y
Year	String		2001-2007	Yes	Y
OutputType	Character		A-T, a-t	Yes	Y
LongTitle	Text	Article Title	Text	Yes	Y
ShortTitle	String	Vol number / Edition	<512 characters>	Yes	Y
Pagination	String	pp	<64 characters>	Yes	Y
Publisher	String	FULL journal title	<256 characters>	Yes	Y
Editors	String		<256 characters>	Yes	
ISBN	String	ISSN	<24 characters>	Yes	
PublicationDate	Date	M/Y or D	2001-01-01 to 2007-12-31	Yes	
DOI	String		<256 characters>	Yes	
IsInterdisciplinary	Boolean	Default = no	true/false	Yes	

Field	Type	For OutputType = 'D'	Expected values	RA2 field	Mandatory
IsSensitive	Boolean	Default = no	true/false	Yes	
ListOfAllAuthors	String		<512 characters>	No	
Indexed by Thomson Reuters (ISI)*	Boolean		true/false	No	
Thomson unique identifier (UT or ISI_LOC)*	String		<15 characters>	No	
Indexed by SCOPUS*	Boolean		true/false	No	
SCOPUS author identifier*	String		N/K	No	
SCOPUS article identifier*	String		N/K	No	

* To be included if known

Table 3 STAFF-OUTPUT (author) links

Record all outputs for each staff ID, record all staff IDs for each output.

HESA institution code	HESAStaffIdentifier	InstitutionalStaffIdentifier	InstitutionalUniqueOutputId
0124	050124987654 Bloggs		20080701
0124	050124987654 Bloggs		200623782
0124	980124546789 Miggins		20080701
Etc	Etc		