

Data requirements for REF import files

Updated September 2013

1. This document sets out data requirements for REF import files and should be read in conjunction with the publications 'Assessment framework and guidance on submissions' (REF 02.2011, hereafter 'guidance on submissions') and 'Panel criteria and working methods' (REF 01.2012, hereafter 'panel criteria'). These are available at www.ref.ac.uk.
2. The data requirements listed show all possible data requirements, whether mandatory or optional, for the purpose of developing REF import files. Existence of a data requirement in this document does not indicate that it is a mandatory requirement for the REF.
3. When 'Date' is listed as the data type, its structure has not been included as this differs depending on the format of import file used.
4. 'Data type' information for the initials field within REF1a was updated in May 2013.
5. 'hesaStaffidentifier' field was removed from within REF1c in September 2013.

Common fields

When importing records the institution and unitOfAssessment fields must be provided.

Name	Data type	Comments	GoS reference
institution	String of 8 characters long	The institution's UKPRN. (UK Provider Reference Number)	
unitOfAssessment	Integer between 1 and 36	The unit of assessment the submission is for.	Annex D
multipleSubmission	Single letter	The multiple submission letter if more than one submission is to be made to a unit of assessment.	Paragraph 50
action	One of the values <ul style="list-style-type: none"> • Update • Overwrite • Delete 	Specifies how the importing of existing records are processed. If no action is provided then the submission system will default to Update. If the record does not exist then Update and Overwrite will insert a new record and Delete will not process the record. Update: will only change the columns included in the import file leaving all other columns with the values they contained before import.	

		<p>Overwrite : will set all the values of the columns to the values in the import file, if a column is not included then the value of the column will be set to NULL.</p> <p>Delete : Will remove the record from the database.</p>	
--	--	---	--

Research groups

Name	Data type	Comments	GoS reference
code	Single letter or digit	The code for the research group.	
name	String, up to 64 characters long	The name of the research group.	

Research staff (REF1a)

Unless otherwise stated see paragraph 84 of 'guidance on submissions'. When importing staff records one of the hesaStaffIdentifier and staffIdentifier fields must be provided.

Name	Data type	Comments	GoS reference
hesaStaffIdentifier	String, up to 13 characters long	The HESA staff identifier for the member staff.	
staffIdentifier	String, up to 24 characters long	An identifier provided by the institution for the member staff. The identifier must be unique within a submission to a unit of assessment.	
surname	String, up to 64 characters long	The last name of the staff member.	
initials	String, up to 10 characters long	The initials of the staff member.	
category	One of the values: <ul style="list-style-type: none"> • A • C 	The category of the member of staff on the census date.	Paragraphs 77-83
birthDate	Date	The date of birth of the member of staff	
contractedFte	A number to two decimal places between 0.2 and 1	The contracted FTE on the census date.	
isResearchFellow	Boolean (True / False)	A value which indicates whether the staff member is a research fellow (for HEFCW-funded institutions only)	Footnote 3
isEarlyCareerResearcher	Boolean (True / False)	A value which indicates whether the staff	Paragraphs 85-87

		member is an early career researcher.	
startDate	Date	The date of starting as academic staff at the institution, if between 1 January 2008 and 31 October 2013	Paragraph 84j
isOnFixedTermContract	Boolean (True / False)	A value which indicates whether the staff member is on a fixed term contract	
contractStartDate	Date	The date the contract started on.	
contractEndDate	Date	The date the contract ended or is due to end.	
isOnSecondment	Boolean (True / False)	A value which indicates whether the staff member is on secondment.	
secondmentStartDate	Date	The date the secondment started on.	
secondmentEndDate	Date	The date the secondment ended or is due to end.	
isOnUnpaidLeave	Boolean (True / False)	A value which indicates whether the staff member is on unpaid leave.	
unpaidLeaveStartDate	Date	The date the unpaid leave started.	
unpaidLeaveEndDate	Date	The date the unpaid leave ended or is due to end.	
isNonUKBased	Boolean (True / False)	A value which indicates whether the staff member is not UK based.	
nonUKBasedText	String	Text explaining the details of the connection between their research activity and submitted unit in the UK.	Paragraph 79d
isSensitive	Boolean (True / False)	A value indicating the staff record contains sensitive information and should be excluded from publication.	Paragraph 36
CircumstanceExplanation	String	Text explaining the staff circumstances cited by a member of staff. (REF1b field)	
ResearchGroup1	Single letter or digit	The code for the first research group the member belongs to.	

ResearchGroup2	Single letter or digit	The code for the second research group the member belongs to.	
ResearchGroup3	Single letter or digit	The code for the third research group the member belongs to.	
ResearchGroup4	Single letter or digit	The code for the fourth research group the member belongs to.	

Staff circumstances (REF1b)

The staff circumstance requirements are explained in paragraphs 63 to 91 of 'panel criteria and working methods'.

Name	Data type	Comments	GoS reference
hesaStaffIdentifier	String, up to 13 characters long	The HESA staff identifier for the member staff. When using XML this field is not required due to the document's hierarchical structure.	
staffIdentifier	String, up to 24 characters long	An identifier provided by the institution for the member staff. The identifier must be unique within a submission to a unit of assessment. When using XML this field is not required due to the document's hierarchical structure.	
circumstanceIdentifier	One of the values: <ul style="list-style-type: none"> • 1 • 2 • 3 • 4 • 5 • 6 • 7 	<p>The type of circumstances cited for the staff member.</p> <p>1 : Early career researcher</p> <p>2 : Part time, career break or secondment</p> <p>3 : Qualifying period of maternity, paternity or adoption leave</p> <p>4 : Period of additional paternity or adoption leave under four months</p> <p>5 : Category A junior clinical academic</p> <p>6 : Category C clinical, health or veterinary professional</p>	

		7 : Complex circumstances	
earlyCareerStartDate	Date	The date the staff member first met the definition of an early career researcher.	Paragraphs 85-86
totalPeriodOfAbsence	A number to 2 decimal places	The number of months within the assessment period that the staff member has been absent.	
numberOfQualifyingPeriods	Integer	The number of qualifying periods of maternity, paternity or adoption leave.	
complexOutputReduction	Integer	The number of outputs the user wishes to reduce without penalty for the member of staff citing complex circumstances.	

Category C circumstances (REF1c: Category C staff details)

The category C circumstance requirements are explained in paragraphs 101 to 104 of 'guidance on submissions'.

Name	Data type	Comments	GoS reference
staffIdentifier	String, up to 24 characters long	An identifier provided by the institution for the member staff. The identifier must be unique within a submission to a unit of assessment. When using XML this field is not required due to the document's hierarchical structure.	
employingOrganisation	String, up to 256 characters long	The name of the organisation that the staff member is employed by.	
jobTitle	String, up to 64 characters long	The job title for the staff member at the organisation the member is employed by	
explanatoryText	String	Text explaining their research responsibilities and how their research is focused in the submitting unit.	

Research outputs (REF2)

Unless otherwise stated see paragraph 118 of 'guidance on submissions'. When importing output records either the outputIdentifier fields or one of the staff identifier fields and the outputNumber field is required.

Name	Data type	Comments	GoS reference
hesaStaffIdentifier	String, up to 13 characters long	The HESA staff identifier for the member staff.	

staffIdentifier	String, up to 24 characters long	An identifier provided by the institution for the member staff. The identifier must be unique within a submission to a unit of assessment.	
outputNumber	Number between 1 and 4	The number of the output for the staff member.	
outputIdentifier	String, up to 24 characters long	An identifier provided by the institution for the output. The identifier must be unique within a submission to a unit of assessment.	
outputType	A letter between A and U	The type of output.	
title	String	The title of the output.	
place	String, up to 256 characters long		
publisher	String, up to 256 characters long		
volumeTitle	String, up to 256 characters long		
volume	String, up to 16 characters		
issue	String, up to 16 characters		
firstPage	String, up to 8 characters long		
articleNumber	String, up to 32 characters long		
isbn	String, up to 24 characters long		
issn	String, up to 24 characters long		
doi	String, up to 256 characters long		
patentNumber	String, up to 24 characters long		
year	One of the values : <ul style="list-style-type: none"> • 2007 • 2008 • 2009 • 2010 • 2011 • 2012 • 2013 	The year the output was published (first entered the public domain, or for confidential reports, was lodged with the relevant body).	
url	String, up to 1024 characters long		
mediaOfOutput	String, up to 24 characters long		
numberOfAdditionalAuthors	An integer greater than -1	The number of additional co-authors.	

isPendingPublication	Boolean (True / False)	A value which indicates whether the output is to be published in December 2013.	Paragraph 111b
isDuplicateOutput	Boolean (True / False)	A value which indicates whether the output has been listed against another member of staff in the submission.	See references in 'Panel criteria', Annex A
isNonEnglishOutput	Boolean (True / False)	A value which indicates whether the output has been published in a language other than English.	Paragraphs 128 – 130
isInterdisciplinary	Boolean (True / False)	A value which indicates whether the output has arisen from interdisciplinary research	Paragraph 119
proposeDoubleWeighting	Boolean (True / False)	A value which indicates whether the output is proposed for double weighting.	See references in 'Panel criteria', Annex A
doubleWeightingStatement	String	A statement justifying the proposal for double weighting.	See references in 'Panel criteria', Annex A
reserveOutput	An integer between 1 and 4	Identifies an output that will not be assessed if this output is accepted as double weighted.	See references in 'Panel criteria', Annex A
*hasConflictsOfInterests	Boolean (True / False)	A value which indicates that named panel members have conflicts of interest with the output.	
conflictedPanelMembers	String	The name(s) of the panel member(s) which may have conflicts of interest for commercial reasons.	Paragraphs 115 – 117
*isOutputCrossReferred	Boolean (True / False)	A value which indicates whether the output is proposed to be cross referred to another panel.	Paragraphs 75d and 119. 'Panel criteria' Part 1 paragraphs 96-100.
crossReferToUoa	Integer between 1 and 36	The panel to cross refer the output to.	As above
additionalInformation	String	Additional information as requested by panels.	See references

			in 'Panel criteria', Annex A
englishAbstract	String	A short abstract in English describing the content and nature of the work, for outputs not written in English.	Paragraphs 128 – 130
researchGroup	Single letter or digit	The code for the research group associated with this output.	Paragraph 119
isSensitive	Boolean (True / False)	A value indicating whether the output record contains sensitive information and should be excluded from publication.	Paragraph 36
excludeFromSubmission	Boolean (True / False)	A value indicating whether the output record should be excluded from submission	
scopusIdentifier	String 20 characters long	The identifier of the journal article or conference proceeding in the Scopus database (export only and only for journal articles/conference proceedings for panels using citation data after the output has been matched with Scopus)	
citedByCount	Integer	The number of journal articles or conference proceedings citing the output (export only and only for journal articles/conference proceedings for panels using citation data after the output has been matched with Scopus)	

*This column has been removed and is no longer required during import.

Impact template (REF3a)

Details of requirements for the impact template are in 'guidance on submissions', paragraphs 149-155 and Annex F; and in the 'panel criteria', relevant sections of Part2 and Annex B.

Name	Data type	Comments	GoS reference
requiresRedaction	Boolean (True / False)	A value which indicates the template requires redaction before	Paragraph 36

		publication.	
statement	Binary	The contents of the PDF file which contains the impact template. When using a text-based import format the binary data should be BASE64 encoded. If uploading using MS Access, see notes at the end of this document.	
redactedStatement	Binary	The contents of the PDF file which contains the redacted impact template. When using a text based import format the binary data should be BASE64 encoded. If uploading using MS Access, see notes at the end of this document.	

Impact case studies (REF3b)

Details of the requirements for impact case studies can be found in paragraphs 156-164 and Annexes F-G of 'guidance on submissions' and in the relevant section of the 'panel criteria'.

Name	Data type	Comments	GoS reference
caseStudyIdentifier	String, up to 24 characters long	An identifier provided by the institution for the case study. The identifier must be unique within a submission to a unit of assessment.	
title	String, up to 256 characters long	A title for the case study	
redactionStatus	One of the values: <ul style="list-style-type: none"> • NotRedacted • RequiresRedaction • NotForPublication 	<p>The redaction status of the case study.</p> <p>NotRedacted : The case study can be published without redaction.</p> <p>RequiresRedaction : The case study needs to be redacted prior to publication.</p> <p>NotForPublication : The case study should not be published at all.</p>	

conflictedPanelMembers	String	The name(s) of the panel member(s) which may have conflicts of interest for commercial reasons.	
caseStudy	Binary	The contents of the PDF file which contains the impact case study. When using a text based import format the binary data should be BASE64 encoded. If uploading using MS Access, see notes at the end of this document.	
redactedCaseStudy	Binary	The contents of the PDF file which contains the redacted impact case study. When using a text based import format the binary data should be BASE64 encoded. If uploading using MS Access, see notes at the end of this document.	
*isCaseStudyCrossReferred	Boolean (True / False)	A value which indicates whether the case study is proposed to be cross referred to another panel.	
crossReferToUoa	Integer between 1 and 36	The panel to cross refer the output to.	

*This column has been removed and is no longer required during import.

Impact case study contacts

Each impact case study can name up to 5 contacts as sources of corroboration for the case study, see 'guidance on submissions' Annex G.

Name	Data type	Comments	GoS reference
caseStudyIdentifier	String with a maximum length of 24 characters	An identifier provided by the institution for the case study. The identifier must be unique	

		within a submission to a unit of assessment.	
number	Integer between 1 and 5	The number of the contact.	
contactType	One of the values: <ul style="list-style-type: none"> • ContactDetails • FactualStatement 	A value which indicates whether the corroboration may be provided through the individual's contact details, or a factual statement already made by them.	
Name	String, up to 64 characters long	The name of the individual for corroboration.	
jobTitle	String, up to 64 characters long	The job title of the individual.	
emailAddress	String, up to 128 characters long	The email address of the contact for corroboration. Not required for a factual statement.	
alternateEmailAddress	String, up to 128 characters long	The second email address of the contact for corroboration. Not required for a factual statement.	
phone	String, up to 24 characters long	The phone number of the contact for corroboration. Not required for a factual statement.	
organisation	String, up to 128 characters long	The name of the organisation the individual works for.	
addressLine1	String, up to 64 characters long	The first line of the address for the contract for corroboration. Not required for a factual statement.	
addressLine2	String, up to 64 characters long	The second line of the address for the contract for corroboration. Not required for a factual statement.	
addressLine3	String, up to 64 characters long	The third line of the address for the contract for corroboration. Not required for a factual statement.	

addressLine4	String, up to 64 characters long	The fourth line of the address for the contract for corroboration. Not required for a factual statement.	
addressLine5	String, up to 64 characters long	The fifth line of the address for the contract for corroboration. Not required for a factual statement.	
postcode	String, up to 10 characters long	The post code of the address for the contact for corroboration. Not required for a factual statement.	
country	String, up to 64 characters long	The country of the address for the contact for corroboration. Not required for a factual statement.	
corroborateText	String, up to 512 characters long	Text describing what aspects of the case study the contact or factual statement can corroborate.	

Research doctoral degrees awarded (REF4a)

The requirements for data on research doctoral degrees awarded are at paragraphs 166 – 170 of ‘guidance on submissions’.

Name	Data type	Comments	GoS reference
year	One of the values: <ul style="list-style-type: none"> • 2008 • 2009 • 2010 • 2011 • 2012 	The academic year the research doctoral degree was awarded in.	
degreesAwarded	A positive number to two decimal places.	The number of research doctoral degrees awarded.	

Research income and research income in kind (REF4b/c)

The requirements for the research income and research income-in-kind data are at paragraphs 171 – 182 of ‘guidance on submissions’.

Name	Data type	Comments	GoS reference
source	Integer between 1 and 15	The source of the income.	

		<p>1 : BIS Research Councils, Royal Society, British Academy and Royal Society of Edinburgh</p> <p>2 : UK-based charities (open competitive process)</p> <p>3: UK-based charities (other)</p> <p>4 : UK central government bodies, local authorities, health and hospital authorities</p> <p>5 : UK industry, commerce and public corporations</p> <p>6 : EU government bodies</p> <p>7 : EU-based charities (open competitive process)</p> <p>8 : EU industry, commerce and public corporations</p> <p>9 : EU other</p> <p>10 : Non-EU based charities (open competitive process)</p> <p>11: Non-EU industry, commerce and public corporations</p> <p>12 : Non-EU other</p> <p>13: Other sources</p> <p>14 : Income from specific bodies that fund health research (see GoS paragraph 172)</p> <p>15: BIS Research Councils (for income-in-kind)</p> <p>(For REF4b, sources 1-14 apply; and REF4c, sources 14 and 15 apply)</p>	
Income2008	Integer	The income for the year 2008-09.	
Income2009	Integer	The income for the year 2009-10.	
Income2010	Integer	The income for the year 2010-11.	
Income2011	Integer	The income for the year 2011-12.	
Income2012	Integer	The income for the year 2012-13.	

Environment template/statement (REF5)

Details of requirements for the environment template are in 'guidance on submissions', paragraphs 183-186 and Annex F; and in the 'panel criteria', relevant sections of Part 2 and Annex C.

Name	Data type	Comments	GoS reference
requiresRedaction	Boolean (True / False)	A value which indicates the template requires redaction before publication.	Paragraph 36
statement	Binary	The contents of the PDF file which contains the environment template. When using a text-based import format the binary data should be BASE64 encoded. If uploading using MS Access, see notes at the end of this document.	
redactedStatement	Binary	The contents of the PDF file which contains the redacted environment template. When using a text-based import format the binary data should be BASE64 encoded. If uploading using MS Access, see notes at the end of this document.	

User

Name	Data type	Comments
emailAddress	String, up to 128 characters long	The e-mail address of the user; this has to be unique to the entire submission system. This does not have to be a valid e-mail address, but if it is not the user will be unable to receive new passwords.
name	String, up to 64 characters long	The name of the user.
password	String, up to 64 characters long	A password for the user. The password will be set to expire on the day before the user record was imported forcing the user to change their password when they next log in. When adding a new user a value must be supplied.

accountExpires	Date	The date after which the user will no longer be able to log into the submission system.
accountType	<p>One of the values:</p> <ul style="list-style-type: none"> • Administrator • User • Automated 	<p>The type of user account.</p> <p>Administrator : Has full access to all parts of the Submission system for the HEI.</p> <p>User : Can log into the Submission system and has the permission given to them by an administrator.</p> <p>Automated : This use can only use the Submission system API.</p>
accountRestriction	<p>One of the values:</p> <ul style="list-style-type: none"> • None • ReadOnly • NoAccess 	<p>The state of the user account.</p> <p>None : The user's normal permissions apply.</p> <p>ReadOnly : A user's write access to replaced by read only.</p> <p>NoAccess : The user has not access to the Submission system.</p>

FormPermissions

Name	Data type	Comments
emailAddress	String, up to 128 characters long	The e-mail address of the user.
unitOfAssessment	Integer between 1 and 36	The unit of assessment the permission is for.
multipleSubmission	Single letter	The multiple submission letter.
form	<p>One of the values:</p> <ul style="list-style-type: none"> • REF1a • REF1b • REF1c • REF2 • REF3a • REF3b • REF4a • REF4b • REF4c • REF5 	The form the permission is for.

	<ul style="list-style-type: none"> • RG 	
permission	One of the values: <ul style="list-style-type: none"> • None • Read • ReducedRead • Write 	The permission to give to the user. None : The user has no permission to view or edit the data. Read : The user is able to view the data. ReducedRead : The user is able to view only the names and identifiers staff (applies to REF1a only). Write : The user can view and edit the data.

Function permissions

Name	Data type	Comments
emailAddress	String, up to 128 characters long	The e-mail address of the user.
function	One of the values: <ul style="list-style-type: none"> • Delete • Import • Export • Submission • News • Locks • CrossRefLookup • ScopusMatching 	The function to the permission applies to.
permission	Boolean (True / False)	If set to true the user is given permission to use the specified function and when set to false the user will have permission to use the function removed.

Notes about MS Access

The access templates use the data type *OLE Object* for the columns which contain PDF files for the REF3a, REF3b and REF5 forms. Microsoft Access stores binary data in the OLE Object data type as OLE Objects (a technology Microsoft in general no longer supports) however the submission system cannot read or write data stored in these OLE Objects. The submission system can read and write files stored in the columns by assuming they are equivalent to the *VARBINARY(MAX)* data type. For an example of how to use access to store files in columns as the submission system requires, see the Microsoft knowledge base article *ACC: Reading, Storing, & Writing Binary Large Objects (BLOBs)* located at: <http://support.microsoft.com/default.aspx?scid=kb;en-us;103257&Product=acc>

Note: Some knowledge of Visual Basic is required in order to achieve this.