

REF 2021 Import/Export documentation

Version: 2.6, December

Updates

Minor updates have been made following the publication of the submission system validation rules document. These changes are highlighted in blue.

New updates have been made to the documentation to take into account the changes to the submission system resulting to the changed timescale as a result of the COVID-19 pandemic. These changes are highlighted in green.

1. The import/export file formats have been updated bring them in-line with the submission system. Most of the changes involved the renaming of fields or values. Some new fields have been added when the implementation of the part of the system required them to be. The postal address details have been removed from the case study contacts as they are no longer required. The impact case study grants section has been redesigned due to better understanding of the requirements for this section.

The import engine will support any files using the previous format *except* for the format of the impact case studies. The changes are highlighted through the document..

Introduction

2. This document provides details of the structure of the import/export file formats, including the names of the tables and fields and details of the expected data types and field lengths. It should be read in conjunction with the 'Guidance on submissions' (REF 2019/01), hereafter 'Guidance on submissions', and 'Panel criteria and working methods' (REF 2019/02), hereafter 'Panel criteria. These are available at www.ref.ac.uk.

3. 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.

4. The case sensitivity of table and field names will follow the convention of the file format. If the file format is case sensitive then the names will follow the camel case convention which is how they appear in this document.

Free text fields

5. All free text fields included in the import/export files should not contain any formatting, and in nearly all cases there is a word limit applied to the field during validation. The submission system will allow the text to be imported in full if it does not exceed the stated character length limits.

Import/export tables

6. The import/export file formats will break down the submission data into the following tables. Some of the details of how these tables are structured depends partly on the file format.

REF form	Table	Name
	Research groups	researchGroup
REF1a	Current staff	currentStaff
REF1b	Former staff	formerStaff
	Former staff contracts	formerStaffContract
REF2	Outputs	Outputs
	Link between staff and outputs	staffOutputLink
REF3	Impact case studies	impactCaseStudy
	Impact case study grants	impactCaseStudyGrants
	Impact case study contacts	impactCaseStudyContact
REF4a	Research doctoral degrees awarded	researchDoctoralDegrees
REF4b	Research income	researchIncome
REF4c	Research income in-kind	researchIncomeInKind
REF5a	Institutional level environment statement	institutionEnvironmentStatement
REF5b	Environment statement	environmentStatement
REF6a	Requests to remove the minimum of one requirement	removeMinimumOfOneRequests
REF6b	Output reduction requests	outputReductionRequests

	Unit rationale statement	unitRationaleStatement
--	--------------------------	------------------------

Common fields

7. In some file formats these fields will appear in every table. In the hierarchical file formats like XML and JSON these may appear only once in the hierarchy.

Field name	Type	Restrictions	Comments
Ukprn	String	Must be 8 characters long	The UKPRN for the institution importing the records
unitOfAssessment	Number	Between 1 and 34	The number of the unit of assessment the records will be imported into
multipleSubmission	Character	A letter between A – Z	Only required if the institution is making more than one submission to a unit of assessment

Research groups

Field name	Type	Restrictions	Comments
Code	Character	An alpha or numeric character	
Name	String	Maximum length 128 characters	

Current staff

Field name	Type	Restrictions	Comments
hesaStaffIdentifier	String	Must be 13 characters long	
staffIdentifier	String	Maximum length 24 characters	Only required if there is no HESA staff identifier.
Surname	String	Maximum length 64 characters	

Initials	String	Maximum length 12 characters	
dateOfBirth	Date		
Orcid	String	Must be 37 characters	The ORCID should not begin with https://orcid.org/ , as the submission system will add the prefix.
contractedFTE	Decimal	2 decimal places	
researchConnection	String	Maximum length 7,500 characters	See Guidance on Submissions paragraphs 123 to 127.
reasonsForNoConnectionStatement	String	One or more of CaringResponsibilities, PersonalCircumstances, ApproachingRetirement, DisciplinePractice	See Guidance on Submissions paragraphs 123 to 127.
isEarlyCareerResearcher	Boolean		Only required for staff members without a HESA staff identifier
isOnFixedTermContract	Boolean		
contractStartDate	Date		
contractEndDate	Date		
isOnSecondment	Boolean		
secondmentStartDate	Date		
secondmentEndDate	Date		
isOnUnpaidLeave	Boolean		
unpaidLeaveStartDate	Date		
unpaidLeaveEndDate	Date		
researchGroups	Character	An alpha or numeric character	¹ Can be repeated up to 4 times.

Former staff

Field name	Type	Restrictions	Comments
staffIdentifier	String	Maximum length 24 characters	
Surname	String	Maximum length 64 characters	
Initials	String	Maximum length 12 characters	
dateOfBirth	Date		
Orcid	String	Must be 37 characters	The ORCID should not begin with https://orcid.org/ , as the submission system will add the prefix.
excludeFromSubmission	Boolean		Indicates the staff should not be included in the submission. No records with this flag set should remain in the submission when submitting it to the REF 2021.

Former staff contract

8. For each former staff member this information may be repeated for each contract. For the non-hierarchical file formats the staff identifier fields from the **Former staff** table will be included on the table as well.

Field name	Type	Restrictions	Comments
hesaStaffIdentifier	String	Must be 13 characters long	
contractFTE	Decimal	2 decimal places	
researchConnection	String	Maximum length 7,500 characters	See Guidance on Submissions paragraphs 123 to 127.
reasonsForNoConnectionStatement	String	One or more of CaringResponsibilities, PersonalCircumstances,	See Guidance on Submissions paragraphs 123 to 127.

		ReducedHours, NormalDisciplinePractice	
startDate	Date		
endDate	Date		
isOnSecondment	Boolean		
secondmentStartDate	Date		
secondmentEndDate	Date		
isOnUnpaidLeave	Boolean		
unpaidLeaveStartDate	Date		
unpaidLeaveEndDate	Date		
researchGroups	Character	An alpha or numeric character	¹ Can be repeated up to 4 times.

Research outputs

9. More information for the requirements for outputs can be found in Annex K of the Guidance on Submissions on in the Output Information Requirements spreadsheet available from the REF website.

Field name	Type	Restrictions	Comments
outputIdentifier	String	Maximum length 24 characters	
webOfScienceIdentifier	String	Maximum length 20 characters	More guidance on the use of this field will be provided when the integration with the citation API has been worked out further.
outputType	Character	A letter between A – V	
Title	String	Maximum length 7,500 characters	If the output has no title, a description is required.
Place	String	Maximum length 256 characters	
Publisher	String	Maximum length 256 characters	
volumeTitle	String	Maximum length 256 characters	
Volume	String	Maximum length 16 characters	

Issue	String	Maximum length 16 characters	
firstPage	String	Maximum length 8 characters	
articleNumber	String	Maximum length 32 characters	
isbn	String	Maximum length 24 characters	
issn	String	Maximum length 24 characters	
Doi	String	Maximum length 1024 characters	
patentNumber	String	Maximum length 24 characters	
Month	String	One of 1 – 12 or January – December or Jan – Dec	Only required for outputs linked to former staff members. See Guidance on Submissions paragraph 264b.
Year	String	One of 2014, 2015, 2016, 2017, 2018, 2019, 2020	
url	String	Maximum length 1024 characters	
isPhysicalOutput	Boolean		An indication that the output will be provided in physical form.
supplementaryInformation	String	Maximum length 1024 characters	See Guidance on Submissions paragraph 264i.
numberOfAdditionalAuthors	Number	A possible integer	See Guidance on Submissions paragraphs 268 to 272.
isPendingPublication [deprecated]	Boolean		https://ref.ac.uk/media/1417/guidance-on-revisions-to-ref-2021-final.pdf paras 44-45).
pendingPublicationReserve [deprecated]	String	Maximum length 24 characters	https://ref.ac.uk/media/1417/guidance-on-revisions-to-ref-2021-final.pdf paras 44-45).
isForensicScienceOutput	Boolean		See Guidance on Submissions paragraphs 275 and 276.
isCriminologyOutput	Boolean		See Guidance on Submissions paragraphs 277 and 278.
isNonEnglishLanguage	Boolean		See Guidance on Submissions paragraphs 285 to 287.
englishAbstract	String	Maximum length 7,500 characters	

isInterdisciplinary	Boolean		See Guidance on Submissions paragraphs 273 and 274.
proposeDoubleWeighting	Boolean		See Guidance on Submissions paragraphs 279 to 283.
doubleWeightingStatement	String	Maximum length 7,500 characters	
doubleWeightingReserve	String	Maximum length 24 characters	The output identifier for the reserve for the pending publication. See Guidance on Submissions paragraphs 279 to 283.
conflictedPanelMembers	String	Maximum length 512 characters	See Guidance on Submissions paragraphs 261 to 263.
crossReferToUoa	Number	Between 1 and 34	See Panel criteria paragraphs 399 to 404.
additionalInformation	String	Maximum length 7,500 characters	See Guidance on Submissions paragraphs 284.
isDelayedByCovid19	Boolean		https://ref.ac.uk/media/1417/guidance-on-revisions-to-ref-2021-final.pdf paras 28-40
covid19Statement	String	Maximum length 7,500 characters	https://ref.ac.uk/media/1417/guidance-on-revisions-to-ref-2021-final.pdf paras 28-40
doesIncludeSignificantMaterialBefore2014	boolean		Indicates the additional information statement includes a statement about significant material in common with an output submitted to REF 2014.
doesIncludeResearchProcess	boolean		Indicates the additional information statement includes information about the research process and/or content.
doesIncludeFactualInformationAboutSignificance	boolean		Indicates the additional information statement includes factual information about the significance of the research.
researchGroups	Character	An alpha or numeric character	

openAccessStatus	String	One of Compliant, NotCompliant, DepositException, AccessException, TechnicalException, OtherException, OutOfScope, ExceptionWithin3MonthsOfPublication	See Guidance on Submission paragraphs 223 to 255.
outputAllocation1	String	Maximum length 128 characters	This is required for UOAs 7, 10,11, 12, 26, 27, 28, 29, 33 and 34. See output allocation guidance at http://www.ref.ac.uk/guidance/additional-guidance/ for more information.
outputAllocation2	String	Maximum length 128 characters	This is required for UOA 26 and optional for UOA10. As above see output allocation guidance at http://www.ref.ac.uk/guidance/additional-guidance/ for more information.
outputAllocation3	String	Maximum length 128 characters	This is required for UOA 12. As above see output allocation guidance at http://www.ref.ac.uk/guidance/additional-guidance/ for more information.
outputSubProfileCategory	String	Maximum length 128 characters	Specifies the output sub-profile category for UOAs 3 and 12. See panel criteria and working methods paragraphs 181 and 183.
requiresAuthorContributionStatement	Boolean		This flag is to enable the submission system to track the author contribution statements to aid institutions in developing their submissions.

isSensitive	Boolean		Indicates the output record contains sensitive information and should be excluded from publication.
excludeFromSubmission	Boolean		Indicates that the output record should be excluded from submission. No records with this flag set should remain in the submission when submitting it to the REF 2021.
outputPdfRequired	Boolean	Export only	Will identify journal articles which the REF team have not been able to retrieve from publishers
outputPdf	² Binary		The PDF of the full text of the output when submitting the output electronically. See Guidance on Submission Annex K.
mediaOfOutput	Boolean	Must not exceed 264 characters in length	Must be used to describe the version of electronic output being returned where not possible to submit the final version in electronic form. E.g. "Proof", "Author Accepted Manuscript". See updated invitation to submit to REF 2021 as PDF at: https://ref.ac.uk/publications/updatedinvitation-tosubmit-to-ref2021/ for more information.

Link between staff and outputs

10. This table links staff to outputs, so the submission system can check the numbers of output submitted per staff member.

Field name	Type	Restrictions	Comments
hesaStaffIdentifier	String	Must be 13 characters long	

staffIdentifier	String	Maximum length 24 characters	
outputIdentifier	String	Maximum length 24 characters	
authorContributionStatement	String	Maximum length 7,500 characters	
isAdditionalAttributedStaffMember	Boolean		A value indicating whether this staff member is an additional attributed staff member for a double weighted output or an output submitted to main panel D.

Impact case studies

Field name	Type	Restrictions	Comments
caseStudyIdentifier	String	Maximum length 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.
Title	String	Maximum length 256 characters	
redactionStatus	String	One of NotRedacted, RequiresRedaction, NotForPublication	
conflictedPanelMembers	String	Maximum length 512 characters	The name(s) of the panel member(s) who may have conflicts of interest for commercial reasons.
caseStudyPdf	² Binary		
redactedCaseStudyPdf	² Binary		
caseStudyDocument	² Binary		
crossReferToUoa	Number	Between 1 and 34	
corroboratingEvidence	² Binary		

IsCovid19StatementNotForPublication	Boolean		https://ref.ac.uk/media/1417/guidance-on-revisions-to-ref-2021-final.pdf paras 53-62
covid19Statement	String	Maximum length 7,500 characters	https://ref.ac.uk/media/1417/guidance-on-revisions-to-ref-2021-final.pdf paras 53-62

Impact case study grants

Field name		Type	Restrictions	Comments
grantsFunding	number	String	Maximum length 256 characters	In non-hierarchical files repeat these columns at the end of the file. See the Excel template for an example.
	amount	Number	Positive integer	
nameOfFunders		String	Maximum length 256 characters	¹ Should be repeated for multiple funders
globalResearchIdentifiers		String	Maximum length 256 characters	¹ Should be repeated for multiple identifiers
fundingProgrammes		String	Maximum length 256 characters	¹ Should be repeated for multiple funding programmes
researcherOrcids		String	Must be 37 characters	The ORCID should not begin with https://orcid.org/ . ¹ Should be repeated for multiple researchers
formalPartners		String	Maximum length 256 characters	¹ Should be repeated for multiple partners
Countries		String	Maximum length 256 characters	¹ Should be repeated for multiple countries

Impact case study contacts

11. For each impact case study this information may be repeated for each contact. For the non-hierarchical file formats the case study identifier field from the **Impact case study** table will be included on the table as well.

Field name	Type	Restrictions	Comments
Number	Number	Between 1 and 5	
Name	String	Maximum length 64 characters	
jobTitle	String	Maximum length 64 characters	
emailAddress	String	Maximum length 128 characters	
alternateEmailAddress	String	Maximum length 128 characters	
Phone	String	Maximum length 24 characters	
Organisation	String	Maximum length 128 characters	

Research doctoral degrees awarded

Field name	Type	Restrictions	Comments
Year	String	One of 2013, 2014, 2015, 2016, 2017, 2018, 2019	
degreesAwarded	Decimal	2 decimal places	

Research income

A list of the income sources and how they map to the HESA sources by year can be found in Annex A.

Field name	Type	Restrictions	Comments
Source	Number	Between 1 and 15	
income2013	Integer		
income2014	Integer		
income2015	Integer		
income2016	Integer		
income2017	Integer		
income2018	Integer		
income2019	Integer		

Research income in kind

A list of the income sources can be found in Annex A.

Field name	Type	Restrictions	Comments
Source	Number	16 and 17.	
income2013	Integer		
income2014	Integer		
income2015	Integer		
income2016	Integer		
income2017	Integer		
income2018	Integer		
income2019	Integer		

Institution environment statement

12. Unlike all the other tables listed the institution environment statement will not include the **unitOfAssessment** or **multipleSubmission** fields.

Environment statement

Field name	Type	Restrictions	Comments
requiresRedaction	Boolean		
Statement	² Binary		
statementDocument	Binary		
redactedStatement	² Binary		
covid19Statement	String		
redactedCovid19Statement	String		

Requests to remove the minimum of one requirement

13. See Guidance on Submissions paragraphs 178 to 183.

Field name	Type	Restrictions	Comments
hesaStaffIdentifier	String	Must be 13 characters long	
staffIdentifier	String	Maximum length 24 characters	Only required if there is no HESA staff identifier.
Circumstances	String	One of ECR, SecondmentsOrCareerBreaks, FamilyRelatedLeave, JuniorClinicalAcademic, RequiringJudgement	¹ Should be repeated for each circumstance which applies. See Guidance on Submissions paragraphs 179 and 180.
supportingInformation	String	Maximum length 7,500 characters	See Guidance on Submissions paragraphs 182.

Output reduction requests

Field name	Type	Restrictions	Comments
hesaStaffIdentifier	String	Must be 13 characters long	
staffIdentifier	String	Maximum length 24 characters	Only required if there is no HESA staff identifier.
typeOfCircumstance	String	One of ECR, SecondmentsOrCareerBreaks, FamilyRelatedLeave, JuniorClinicalAcademic, RequiringJudgement	See Guidance on Submissions paragraphs 160 to 162.
tariffBand	Number	Between 0 and 3	Should map to the rows of Table 1 or Table 2 in the annex L of the Guidance on Submissions for the circumstance being claimed.
supportingInformation	String	Maximum length 7,500 characters	See Guidance on Submissions paragraph 193.

Unit rationale statement

Field name	Type	Restrictions	Comments
unitRationaleStatement	String	Maximum length 7,500 characters	See Guidance on Submissions paragraph 177.

Annex A – Income sources

Source		Column numbers by year as in HESA templates					
		2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
1	BEIS Research Councils, The Royal Society, British Academy and The Royal Society of Edinburgh	C1	C1	C1i	C1i	C1i	C1i
2	UK-based charities (open competitive process)	C2	C2	C2	C2	C2	C2
3	UK-based charities (other)	C3	C3	C3	C3	C3	C3
4	UK central government bodies/local authorities, health and hospital authorities	C4	C4	C4	C4	C4	C4
5	UK central government tax credits for research and development expenditure		C5	C5	C5	C5	C5

6	UK industry, commerce and public corporations	C5	C6	C6	C6	C6	C6
7	UK other sources	C13	C14	C7	C7	C7	C7
8	EU government bodies	C6	C7	C8	C8	C8	C8
9	EU-based charities (open competitive process)	C7	C8	C9	C9	C9	C9
10	EU industry, commerce and public corporations	C8	C9	C10	C10	C10	C10
11	EU (excluding UK) other	C9	C10	C11	C11	C11	C11
12	Non-EU-based charities (open competitive process)	C10	C11	C12	C12	C12	C12
13	Non-EU industry commerce and public corporations	C11	C12	C13	C13	C13	C13
14	Non-EU other	C12	C13	C14	C14	C14	C14
15	Health research funding bodies						

16	Research councils income-in-kind						
17	Health research funding bodies income-in-kind						

Annex B – Summary of changes to the file formats

The import engine will support the importing of the original names along side the updated names, and any field the import engine does not recognise is ignored. Therefore with the exception of the changes to the impact case study grants section all changes are backwardly compatible.

Form	Field	Summary of changes
Research group	name	Increased the maximum length from 64 characters to 128 characters.
Outputs (REF2)	supplementaryInformation	Renamed the field from supplementaryInformationDOI.
	doesIncludeSignificantMaterialBefore2014	Field added, to enable the system to work out the word count for additional information.
	doesIncludeResearchProcess	Field added, to enable the system to work out the word count for additional information.
	doesIncludeFactualInformationAboutSignificance	Field added, to enable the system to work out the word count for additional information.
	openAccessStatus	The OtherFurtherException status has been renamed OtherException and the ExceptionWith3MonthsOfPublication has been renamed ExceptionWithin3MonthsOfPublication.
	outputAllocation1	Renamed the field from outputAllocation
	outputAllocation2	Field added.
Staff/Output links (REF2)	isAdditionalAttributedStaffMember	Field added, to record whether this staff member is an additional attributed staff member for a double weighted output or an output submitted to main panel D.
Impact case studies (REF3)	redactedCaseStudyPdf	Field added.
	corroboratingEvidence	Field added.
Impact case studies grants (REF3)		This section of the import file has been reworked completely due to a better understanding of the requirements. NOTE: Old versions of this section are not supported by the import engine.
Impact case studies contacts (REF3)	contactType, addressLine1, addressLine2, addressLine3, addressLine4, addressLine5, postcode, country, corroborateText	These fields have been removed as they are no longer required.

Requests to remove the minimum of one (REF6a)	circumstances	Renamed the RequiresJudgement circumstance to RequiringJudgement.
	supportingInformation	Renamed the field from supportingStatement
Output reduction requests (REF6b)		Section renamed from unitCircumstancesStaffList
	typeOfCircumstance	Renamed the RequiresJudgment circumstance to RequiringJudgement.
	supportingInformation	Renamed the field from supportingStatement.
Unit rationale statement (REF6b)	unitRationaleStatement	Renamed the field from supportingStatement.

¹ In hierarchical file formats these items can just be repeated in the file, for other formats a semi-colon delimited list should be provided in the single field.

² Fields of type binary will only be supported in some of the file formats. Text based file formats (XML and JSON) for example will require the binary data to be BASE64 encoded.