UK Research and Innovation

INFORMATION GOVERNANCE

Data Protection Impact Assessment (DPIA)

Template Published: 28 September 2018 Revision: 1.0

(Please save this template and rename to the appropriate system / process)

Name of the processing activity	
REF 2021 Submissions System	
DPIA reference	**Starting date of the DPIA**
CONFIRM WITH DAVID HYATT	11/04/2019
Owner of the process	**Date of process go-live**
REF Director (Kim Hackett)	11/04/2019

DPIA version control

Version	Date Issued	Author	Comments
FINAL 1.0	01/04/2019	Andy Hepburn	Approved version for release
Draft 1.1	05/09/2019	Andy Hepburn	Updated to include contract as a lawful basis. Clarify A9.2(j) as basis for processing special category data. Data sharing across UK FBs. REF4a data sharing with HEIs.
Draft 1.2	10/09/2019	Andy Hepburn	Final drafting
RELEASE 2.0	12/09/2019	Andy Hepburn	Approved version for release

This form only needs completing if the processing involves 'personal data': Any information relating to an identified or identifiable natural person.

This form is split into 2 sections:

Section 1 DPIA pre-assessment – determines whether a full DPIA (completion of Section 2) is required

Section 2 Completing the DPIA – full DPIA

At the back of this template, there are **Guidance notes.**

Pre-populated text in the form << text >> should be over-written and the table boxes will grow to accommodate text, so do not feel restricted.

Add links to supporting documents where possible to save replication or duplication.

Section 1. DPIA pre-assessment

The pre-assessment will determine whether a full DPIA is required.

Explain the need to collect and process personal data

Business Need:

This processing is required to facilitate the peer-review process for the REF, and to ensure adherence to the rules of the REF.

The Purpose:

The personal data will be used to verify that outputs have been correctly attributed to the correct research staff, and that the number of outputs submitted is correct, based on number (FTE) of staff returned.

Sensitive personal data will be used to reduce the number of outputs submitted, and to exempt specific staff from being submitted with a minimum of one output, as per the policy on Staff Circumstances as set out in the Guidance on Submissions for REF2021.

Describe the personal data involved

How many individuals are affected?

85,000

List the different categories of Data Subjects involved (e.g. Research and Innovation Community, UK Research and Innovation Board, UKRI Council or Peer Review Boards and Committees, Peer Reviewer, Current or past employee, visiting worker, student or third-party, agency worker, member of the public, a supplier, children and young people, part of a research study)?

Researchers from HEIs

People from HEIs who are administering submissions (often, but not exclusively, people working in research offices)

Members of peer-review panels and sub-panels

Describe the categories of personal data involved (e.g. work contact details, personal contact details, bank accounts, financial, health, and trades unions etc.)? Note: you don't have to list each field, just different groups of personal data.

Names, roles and employers

Special category data, where submitted (we expect the numbers to be relatively small)

What is the lawful basis for processing the personal data?

Put an X in all relevant boxes

Public Task	Contractual	Legal obligation	Legitimate interest	Vital interest	Consent
X	X	X			

If using **legitimate interest** or **consent** basis, please enter why?

n/a

Qu	Criteria to determine if the processing is 'likely to result in a high risk'	Y / N
1	Will the processing create an evaluation or score? No (the scoring will be done in the Assessment System, which is a separate system)	N
2	Will the processing result in an automated decision having a legal or similar significant effect? No – it is all peer-reviewed	N
3	Does the processing involve systematic monitoring? No	N
4	Does it involve special category data, or highly personal data? Yes, although most of the detail of the special category data remains at the HEIs and is not returned to us. However, some Staff Circumstances submissions will contain narrative data of a sensitive personal nature.	Y
5	Does it involve large scale processing?	N
6	Have datasets been matched or combined? Yes – datasets are matched for audit reasons, and also to run analysis of the data.	Y
7	Are vulnerable data subjects involved? No	N
8	Is Innovative technology being used, and/or current technology being used in an innovative way? Not for personal data (possibly not at all)	N
9	Does the processing prevent data subjects from exercising a right, using a service or a contract? No	N

1.1. Pre-assessment decision

The more criteria met by the processing, the more likely it is to present a high risk to individuals, but in general, if at least two criteria are matched, a DPIA will be required.

Is a DPIA required?	Y / N
Yes	

1.2. Pre-assessment sign-off

Data Protection Lead	Council
Andy Hepburn	Research England

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY

Section 2. Completing the DPIA

2.1. The Processing

Describe the nature of the processing:

You might find it useful to refer to a flow diagram (see DPIA-Dataflow-template) or another way of describing data flows.

How will you collect, use, store and delete data?

Data will be submitted via an online portal, into which HEIs will submit information about staff that are in-scope for the REF. This will include identifying information including surname, initials and date-of-birth. It will also include the HESA staff identifier, which is a unique ID assigned to each staff member in data returns that are made to the Higher Education Statistics Agency (HESA).

For small numbers of staff, HEIs will submit coded and narrative information about circumstances that may have restricted their research output – this information will include sensitive personal information.

Staff data will be matched to HESA data using the HESA staff identifier for audit and statistical analysis purposes.

Additional personal information about staff, including sensitive personal information, will be sent to the audit system via a separate online portal. This information will be deleted as soon as the audit issue to which it relates has been closed.

Data will be stored in Microsoft Azure, in a datacentre located in the UK.

Data will be deleted in December 2021.

In certain circumstances, data collected by UKRI (of which Research England is a part) may be shared with one or more of the other Higher Education Funding Bodies (the Higher Education Funding Council for Wales, the Scottish Funding Council and the Department for the Economy, Northern Ireland). This can happen if the data were submitted by an HEI that is funded or regulated by that funding body. Where the data have been submitted within a joint submission by HEIs that are funded or regulated by different funding bodies, this can also lead to data sharing with those relevant funding bodies. In circumstances where data need to be considered by the REF Steering Group, membership of which comprises representatives from all four funding bodies, the data will necessarily be shared will all the representatives and therefore with all four funding bodies.

For Research doctoral degrees awarded (REF4a), HESA data for academic years 2013–14, 2014–15, 2015–16, 2016–17, 2017–18 and 2018–19 will be shared with the submitting HEIs. Normally the data will only be shared with the HEI that originally submitted it to HESA. However, in circumstances where a research student studied at two HEIs (either concurrently or sequentially) we will return that student's data to both the HEI that returned the data to HESA and the other HEI.

We have considered the data fields that will be shared with the non-HESA-submitting HEI, and have concluded that we will not be sharing with them any data that they would not already have in their student record system, apart from the study end date and the reason study ended (which would either be successful completion or death). The scope of GDPR only covers living individuals. We have considered whether it is reasonable to share with the other HEI the data disclosing the fact that that a student successfully completed their research doctoral degree, and have concluded that doing so would not compromise the rights and freedoms of the data subject.

Where has the data originated from (e.g. data subject, system)?

All the personal data will originate from HEIs. The specific systems will depend on HEIs' own data processes and practices.

List organisations the personal data is externally shared with:

Some personal data – but not sensitive personal data – may be shared with the Copyright Licensing Agency (CLA), for the purpose of auditing use of licensed copyright material such as books and journal articles.

For this reason, we have included 'contractual' as one of the lawful bases for processing, as these personal data would be shared with the CLA to satisfy the requirement of our licensing agreement with them. The lawful basis 'contractual' will only apply to users of the REF submissions system – it will not apply to staff submitted to the REF.

Describe the system(s) used to manage the data

Include: Is it a software system? Is the system a purchased commercial system or an in-house built software system? A collection of files/documents? Is it a novel technology in any way? Include any DPIA references, where appropriate.

The system is an in-house build web application. It uses C#, HTML and Javascript (Angular) to render the user interface. Data storage is in a SQL database.

There is a file store for submission upload files, which will be in XML, JSON and Excel (xlsx) format. These files will contain personal and sensitive personal data. They will be retained for two weeks and then automatically deleted.

There is also a file store for digital copies of research outputs such as books and journal articles.

Describe the context of the processing:

Would the personal data be expected to be used in this way?

Yes – the way the personal data will be used is clear from the published Guidance on Submissions.

Are there prior concerns over this type of processing or security flaws?

No.

What is the current state of technology in this area?

Mature – all the technologies are typical of this type of application.

Are there any current issues of public concern that you should factor in?

Only that there is some interest in releasing the individual scores for the research outputs (the aggregated scores are published as "output sub-profiles" anyway). This is a question of policy rather than technology.

Are you signed up to any approved code of conduct or certification scheme (once any have been approved)?

No

Any other context to be noted?

There will be further processing in the "Assessment System", which is a separate system that will be developed later and used to disseminate research outputs, some staff information and assessment scores to/from sub-panel members.

2.2. Consultation

Consider how to consult with relevant stakeholders:

Describe when and how you will seek individuals' views – or justify why it's not appropriate to do so. Who else do you need to involve within your organisation? Do you need to ask your processors to assist? Do you plan to consult information security experts, or any other experts?

Extensive stakeholder engagement has already taken place when developing the Guidance on Submissions. This included a full consultation exercise to canvass views from across the sector, for which we received approximately 300 responses.

We expect the submissions system to comply substantially with the approach recommended by the GDS (Government Digital Service) and have held discussions with representatives from GDS to identify how best to integrate the approach into our development and service delivery.

We have circulated the Data Protection Impact Statement to our Data Collection Steering Group (DCSG), and the draft statement was discussed at the March meeting of the DCSG.

2.3. Necessity and proportionality

Describe compliance and proportionality measures, in particular:

Is there another way to achieve the same outcome?

Not without collecting the data set as described in the Guidance on Submissions. Alternatives such as paper return and manual data entry would be more error-prone and less efficient.

In respect of the sharing of HESA data about research doctoral degrees awarded (REF4a) with the HEI(s) where the students were registered, there is not an alternative mechanism that would allow HEIs to verify the data in order to make accurate data returns in their REF submissions.

How will function creep be prevented?

All functionality is determined by the requirements of the Guidance on Submissions. The software is being developed using Agile methodology, and there is representation from the REF Director (the product owner), REF Head of Policy and others on the REF team to ensure that user stories are only scoped to include the functionality required by the Guidance on Submissions.

How will data quality and data minimisation be achieved?

Data quality is subject to audit processes (the full audit guidance will be published in Summer 2019).

Data minimisation is undertaken by ensuring that only the data capture requirements in the Guidance on Submissions are implemented (this is the technical measure that we have taken to ensure data minimisation as required under article 89[1]). Further, the data requires no further processing after December 2021, and will be deleted at that point in time (this is the organisational measure that we have taken to ensure data minimisation as required under article 89[1]).

How will Data Subjects' rights be implemented?

All data in the submission that relates to a data subject can be disclosed under a Subject Access Request. In addition, where audit evidence about the data subject (such as contract or payroll information) is held at the time a Subject Access Request is issued, this will also be disclosed.

A data subject can request rectification, but would need to supply appropriate supporting evidence, which would also need to be verified with the HEI.

A data subject can request deletion, but this could only be a valid request in circumstances where the HEI had not followed their own Code of Practice for identifying eligible staff, or had not followed their own Code of Practice in returning self-disclosed staff circumstances. In either case, this would be dealt with as an audit issue, and the decision whether to delete or not would be determined by the outcome of the audit process.

The lawful bases for processing personal data are as set out in section 1, page 2.

Where data about staff individual circumstances are special category data (as defined in the Data Protection Act 2018 and the GDPR), the specific condition for processing is that 'processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes'.

What measures will be in-place to ensure processors comply?

All requests for access, rectification or deletion will be managed by the Head of Data Verification and Information Management. They will also be copied to UKRI's DPO. There will be a standard process for fulfilling subject access requests.

Requests for rectification or deletion will, by their very nature, initiate an audit query and the compliance will be managed through the audit process.

How do you safeguard any international transfers and personal data sharing?

All data is held in the UK.

There is a small risk of limited disclosure of personal data to CLA if they undertake an audit of our copyright licence agreement. We are including a Non-Disclosure Agreement in the licence agreement which places strict obligations on CLA in terms of any personal data that they may access during the audit process.

If transferring or sharing will be outside of the European Union, please state which countries

None – not applicable.

Any other necessity or proportionality to be noted?

The design of the collection and processing is the minimum necessary to achieve the goals of the REF, and we therefore consider the processing to be minimised in conformance with Article 89 of the GDPR and section 41 of the Data Protection Act 2018.

2.4. Transparency

How will Data Subjects be informed about this processing activity?
This could be via privacy notices, terms & conditions, guidance documents, etc.

HEIs will be issued with a model Data Protection Statement which they can give to data subjects whose information is being returned in the REF. This will explain how their personal information will be used in the REF.

A fair processing notice for staff submitted to the REF has been published. It includes the same information as the model Data Protection Statement, and also the lawful bases and specific processing condition for special category data. It also links to UKRI's privacy notice. It is available from this link: http://www.ref.ac.uk/guidance/data-management-guidance/fair-processing-notice-for-staff-submitted-to-ref-2021/

The privacy notice for users of the submissions system is available from this link: https://www.ref.ac.uk/submission-system/privacy-notice/

The Guidance on Submissions was published on 31st January 2019 on the REF website, and explains how personal data will be used in the REF.

AS per article 14(5)(b) of the GDPR, we have undertaken a balancing test and concluded that it would be disproportionate to provide the information required in article 14(1) and 14(2) directly to the data subjects, where they are staff submitted to the REF. This is because there is no necessity within the scope of the exercise to collect contact (address / telephone / email) information from the data subjects, and we have therefore not included collection of this information within our published data specification. Further, we are unsure whether we could collect this information through the REF submissions system without compromising our obligations under articles 5(1)(c) and 89(1). We have made our best endeavours to make this information available to staff submitted to the REF, via HEIs through the aforementioned model Data Protection Statement and via the fair processing notice published on the REF website.

2.5. Risks and mitigations

Intro text – who approves the measures?

	Assess the risk				**Manage the risk**			
Risk id	Source of risk on individuals	Likelihood of harm	Severity of harm	Risk	Measures to reduce or eliminate risk	Effect on risk	Residual risk	Measure approved
1	Data breach (disclosure) from direct hacking of the submissions system	Remote	Serious	Low	Procure penetration testing of the submissions system, and rectify where any vulnerabilities are found. Comply substantially with the approach recommended by the GDS (Government Digital Service) for service delivery.	Reduced	Low	Yes
2	Data breach (unauthorised amendment/deletion of data) from direct hacking of the submissions system	Remote	Serious	Low	Procure penetration testing of the submissions system, and rectify where any vulnerabilities are found. Comply substantially with the approach recommended by the GDS (Government Digital Service) for service delivery.	Reduced	Low	Yes
3	Data breach (disclosure) from inadequate processes, such as access to unauthorised areas of the system by authorised persons (e.g. unauthorised access by an HEI to another HEI's data, or unauthorised access by CLA during audit)	Remote	Serious	Low	Test access controls thoroughly to ensure authorisation is limited to specific and defined access controls.	Reduced	Low	Yes

		Likelihood of harm		
		Remote	Reasonable possibility	More likely than not
Severity of Impact	Serious harm	**Low risk**	**High risk**	**High risk**
	Some impact	**Low risk**	**Medium risk**	**High risk**
	Minimal impact	**Low risk**	**Low risk**	**Low risk**

2.6. Sign-off and record outcomes

Item	Name / date	Notes
Measures approved by:	Catriona Firth – 01/04/2019	Integrate actions back into project plan, with date and responsibility for completion
Residual risks approved by:	Catriona Firth – 01/04/2019	If accepting any residual high risk, consult the ICO before going ahead
DPO advice provided:	Moyo Taiwo – 18/02/2019	DPO should advise on compliance, step 6 measures and whether processing can proceed
Summary of DPO advice: I have had a thorough look at the DPIA and in my opinion, it looks absolutely fine. It covers all the key elements of a DPIA and also very detailed and specific on why we need a DPIA and also the completion of the DPIA. As it looks fine, I will not be updating anything from this end.		
DPO advice accepted or overruled by:	Catriona Firth – 01/04/2019	
If overruled, you must explain your reasons: n/a		
Consultation responses reviewed by:	Catriona Firth – 01/04/2019	
If your decision departs from individuals' views, you must explain your reasons: n/a		
This DPIA will be kept under review by:	Andy Hepburn	The DPO should also review ongoing compliance with DPIA

Section 3. Guidance notes

3.1. Why is this required?

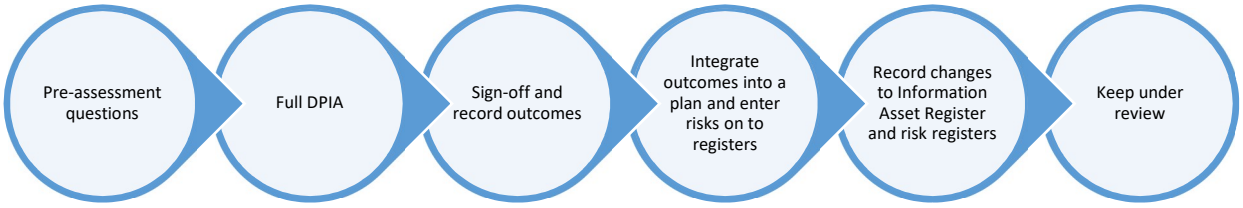
A data protection impact assessment is a process to help identify and minimise data protection risks of a project or whilst reviewing processing activities or systems, by:

- Helping to identify the data protection risks and their mitigations
- Creating a register of DPIAs to help evidence compliance
- Demonstrates personal data risks are managed in a compliant way.

3.2. When to use

- If, and only if, personal data is being processed
- At the start of a project
- If there's a significant change to an existing process or IT system.

3.3. The process



The pre-assessment questions will help to determine if a full DPIA is required by understanding the context within which the personal data is being processed. Your council Data Protection Lead will need to be involved and sign-off the pre-assessment.

When the pre-assessment determines the processing 'is likely to result in a high risk to the rights and freedoms of natural persons', a full DPIA will need to be completed to determine the risks and mitigations.

In order to agree the mitigations, various stakeholders identified in the consultation section will need to be involved as wider parameters (such as longevity of the system and budget) will need to be considered. You should involve your local Data Protection Lead with this too.

Once the mitigations are agreed, the Data Protection Officer needs to sign off on the plan before being taken forward, and risks need to be added to council risk registers.

Having implemented the modifications to the processing, Information Asset Registers and risk registers need to be updated.

DPIAs need to be reviewed periodically, at least every 3 years, requiring a number of stakeholders.

3.4. Types of DPIAs and how to group DPIAs

DPIAs can be used to assess:

- A complete processing activity which might include one or more IT systems e.g.
- Just an IT system where the processing activity isn't assessed but the controls for dealing with data processing are e.g. Office 365

- A group of similar IT systems providing base functionality e.g.

3.5. Document information

Name of the process

DPIA reference

Owner of the process or system
The outcomes from a DPIA necessitate changes to a process or IT system, and these changes can have ramifications to a processing activity, or further down in other processing activities, which is why it is important to identify a group or person with knowledge across the broader spectrum and involve them in the consultation.

3.6. DPIA pre-assessment

Definitions of the criteria

Qu	How to interpret the criteria
1	Will the processing create an evaluation or score? This includes profiling and predicting especially from aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements (recitals 71 and 91). Broadly speaking, profiling means gathering information about an individual (or group of individuals) and analysing their characteristics or behaviour patterns in order to place them into a certain category or group, and/or to make predictions or assessments about, for example, their: • ability to perform a task • interests • likely behaviour UKRI examples:
2	Will the processing result in an automated decision having a legal or similar significant effect? Automated decision-making has a different scope and may partially overlap with profiling. Solely automated decision-making is the ability to make decisions by technological means without human involvement. Automated decisions can be based on any type of data, for example: • data provided directly by the individuals concerned (such as responses to a questionnaire) • data observed about the individuals (such as location data collected via an application) • derived or inferred data such as a profile of the individual that has already been created (e.g. a credit score) UKRI examples:
3	Does the processing involve systematic monitoring? Processing used to observe, monitor or control data subjects, including data collected through a systematic monitoring of a publicly accessible area. Additionally, it may be impossible for individuals to avoid being subject to such processing, i.e. CCTV. UKRI examples: Office outside security CCTV cameras, web traffic monitoring systems.

4	Does it involve special category data, or highly personal data? This includes special categories of data, and data that may be considered as increasing the possible risk to the rights and freedoms of individuals, i.e. ethnicity, political opinions and personal data relating to criminal convictions and offences. Special category data is defined as: Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership; data concerning health or sexual orientation; genetic data or biometric data UKRI examples:
5	Does it involve large scale processing? when determining whether the processing is carried out on a large scale, take account of: • the number of data subjects concerned, either as a specific number or as a proportion of the relevant population • the volume of data and/or the range of different data items being processed • the duration, or permanence, of the data processing activity • the geographical extent of the processing activity UKRI examples:
6	Have datasets been matched or combined? Datasets originating from two or more data processing operations performed for different purposes and/or by different data controllers in a way that would exceed the reasonable expectations of the data subject. UKRI examples:
7	Are vulnerable data subjects involved? Vulnerable subjects occur where there is an increased power imbalance between the data subjects and the data controller, meaning the individuals may be unable to easily consent to, or oppose, the processing of their data, or exercise their rights. This may include children (they can be considered as not able to knowingly and thoughtfully oppose or consent to the processing of their data), **employees**, more vulnerable segments of the population requiring special protection (mentally ill persons, asylum seekers, or the elderly, patients, etc.). UKRI examples: employees, children.
8	Is Innovative technology or current technology being used in an innovative way? This could include the use of innovative technology, or use of current technology in an innovative way. UKRI examples:
9	Does the processing prevent data subjects from exercising a right, using a service or a contract? (Article 22 and recital 91) processing performed in a public area that people passing by cannot avoid, or processing that aims at allowing, modifying or reusing data subjects' access to a service or entry into a contract. UKRI examples: photographing events

3.7. The processing

3.8. Consultation

3.9. Necessity and proportionality

3.10. Transparency

3.11. Assessing risks and mitigation

3.12. Sign-off and recording outcomes

Section 4. Related documents

Advice on DPIAs on the ICO website:
https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/

This template is a modified version of the ICO DPIA template:
https://ico.org.uk/media/about-the-ico/consultations/2258461/dpia-template-v04-post-comms-review-20180308.pdf