National Records of Scotland

COVID-19 Research – Census Cohort Upload

Data Protection Impact Assessment (DPIA)

NRS-DPIA-2020-05

This document is to be completed in accordance with the Cabinet Office Data Ethics Framework and NRS guidance on DPIAs.

Document Control

Title	COVID-19 Research Data Protection Impact Assessment
Prepared by	M Pemble, Head of Security & Privacy, Census 2021
Approved by	*NRS DPO*
Date of approval	11/6/20
Review frequency	*6 Months*
Next review date	11 Dec 2020

Status Control

Version	Date	Status	Prepared by	Reason for Amendment
0.1	18 May 2020	Draft	M Pemble	
0.2	19 May 2020	Draft	L Cavin	Additional Framework Detail
0.3	20 May 2020	Draft	M Pemble	Completed Section 3 – for discussion with ICO
0.4	28 May 2020	Draft	L Cavin	Reviewed draft with ICO and addressed comments
1.0	12 June 2020	Finalised	L Cavin	Finalised following NRS Privacy Group and DPO review
1.1	11 Sept 2020	Update	L Cavin	Updated Section 2 – CHI advisory group approval

1. Project Details

Statistics PBPP reference number	n/a
Project name	COVID-19 Research Hub initial NRS upload
Lead researcher	L Cavin, National Records of Scotland

The Scottish Government COVID Data Taskforce has requested that NRS participate in efforts to assemble a collection of individual, linkable datasets within an extension of the National Safe Haven. This will have defined access routes for researchers and public body analysts completing research to provide evidence for:

- The operational response to COVID-19
- The likely effects of easing lockdown restrictions/the route to recovery
- The possible longer term effects of COVID-19 on Scotland

The request for NRS Census data is:

- A 100% cohort of selected variables from the 2011 Census
- The Scottish Longitudinal Survey (SLS)
- Mother and Father Links (data linkage project in collaboration with Public Health Scotland, currently a work in progress)

Research Data Scotland will form the single point of contact for researchers seeking access to securely held data. The datasets will be held alongside other key health and social datasets, in the National Safe Haven (NSH) developed by the Edinburgh Parallel Computing Centre and administered by eDRIS (Public Health Scotland). This infrastructure will be used to store, process and link data safely. eDRIS will be responsible for the creation of extracts from this data store for individual projects – with projects subject to individual review and approval by the Statistics Public Benefit and Privacy Panels (S-PBPP). NRS has representation on S-PBPP, and retains status as data controller, ensuring that we can review specific projects and prevent inappropriate access to our data. Current arrangements requiring Data Sharing Agreements (establishing joint data controller status between NRS and researchers) and Data Protection Impact Assessments will be maintained, and eDRIS will remain data processors. Outputs from analytical projects using census data would continue to be reviewed before release by NRS statisticians.

The justification for the 100% census cohort is that whole population coverage will enable analysis of the impacts of COVID on small groups within the population, such as geographically small areas or occupational groups. Individual projects will receive cohorts that must be justified at S-PBPP stage, according to the principle of data minimisation. Census variables will cover themes such as demography, economic activity, health and travel to work, which will be vital in assessing the vulnerable population, and wider impacts of the current pandemic.

Supplying custom microdata extracts from the census for research is not new, and the request for data for the COVID-19 Modelling and Analysis Hub largely uses existing processes and data infrastructure. Specifically, the new and revised elements of this proposal are as follows:

- S-PBPP review has been streamlined to a 1 week review process for census data (previously 2 weeks, or 3 when including NRS Privacy Group review)
- SLS available through NSH (previously only available through site visit to Ladywell House)
- Custom extracts of 2011 census restricted to a subset of variables, selected by NRS as relevant to pandemic related research
- 100% cohort of 2011 census remains in NSH secure data management zone (previously in nrscotland, with only custom extracts delivered to NSH)
- Signed Data Sharing Agreements and Data Protection Impact Assessments will still be required to cover projects using the data
- Data linkage based on CHI (encrypted) within NSH
- Mother and Father Links is a new look-up, intended to help in data linkage projects where intergenerational linkages are important.

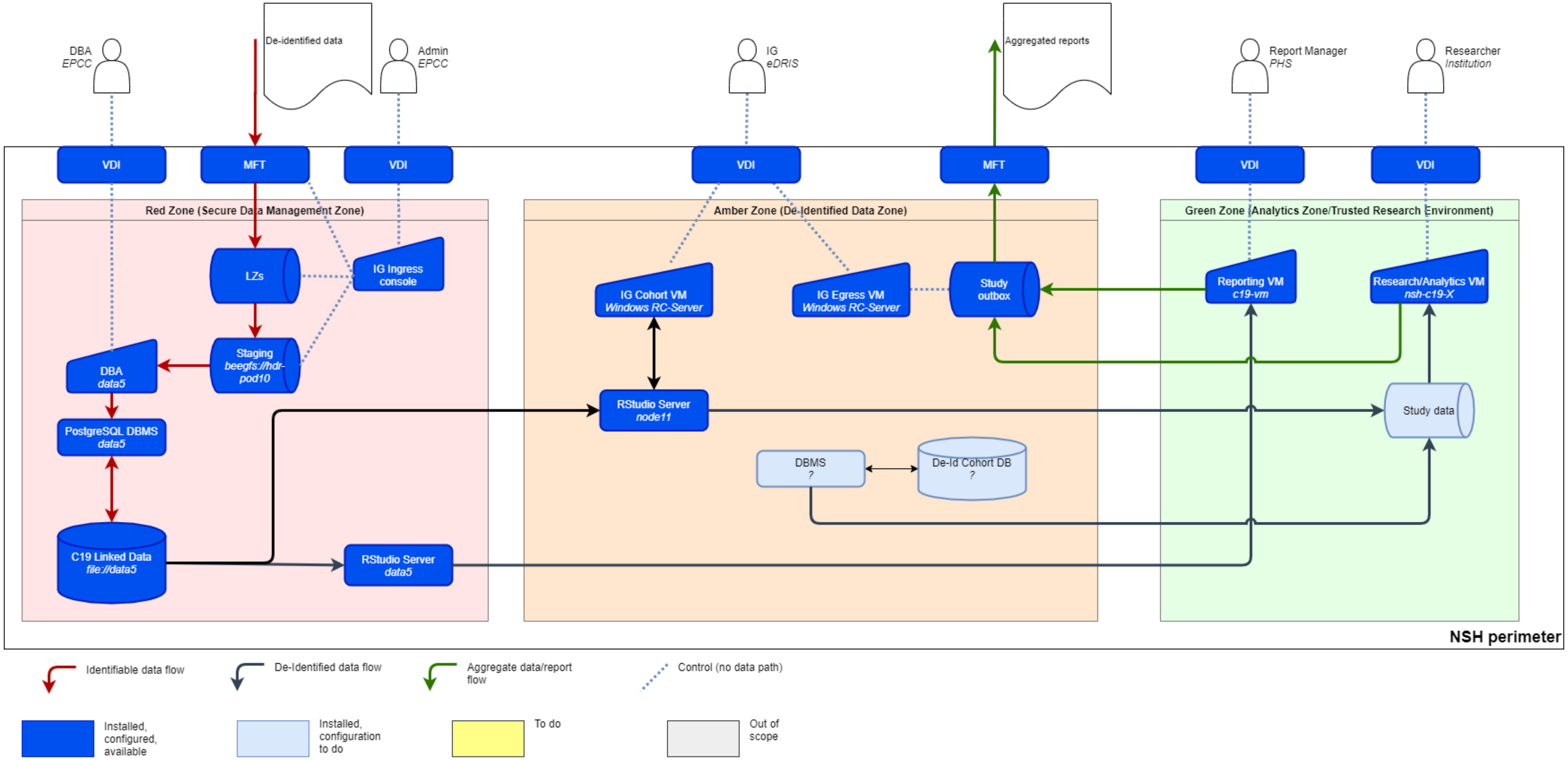
This proposed data sharing is being established in response to exceptional circumstances. NRS are not bound by any precedent.

Access to census data by researchers will be through remote access to the National Safe Haven. Terms for remote access and eDRIS security statement are provided in appendices to this document.

Diagram of proposed infrastructure and data flows within the National Safe Haven:

27/05/2020

C19 Data Workbench / National Safe Haven



2. Data Ethics Framework Assessment

Data ethics is an emerging branch of applied ethics which describes the value judgements and approaches we make when generating, analysing and disseminating data. This includes a sound knowledge of data protection law and other relevant legislation, and the appropriate use of new technologies. It requires a holistic approach incorporating good practice in computing techniques, ethics and information assurance. Provide an assessment of your research proposal against the seven data ethics principles. Guidance on the Data Ethics Framework these principles are drawn from has been published on the UK Government website. Under each principle there are suggested questions to answer. Not all of these will be relevant to every application. You are only expected to address the questions that are pertinent to your research.

2.1 Framework Scoring

Question Set	0	1	2	3	4	5
1. Start with clear user need and public benefit	User need is not well defined					User need is clearly defined
						*
2. Be aware of relevant legislation and codes of practice	Needs clarification or expert input					Relevant laws are well understood
						*
3. Use data that is proportionate to the user need	Reuse not proportionate					Reuse of data is clearly proportionate to achieve user need
						*
4. Understand the limitations of the data	Unreliable, unsuitable data					Data is representative and accurate
					*	
5. User robust practices and work within your skillset	Needs further expert input					Methodologies clearly designed and understood.
						*
6. Make your work transparent and be accountable	No scrutiny or peer review available					Oversight built in through life cycle of project

						*
7. Embed data use responsibly	No ongoing plan determined					Evaluation plan developed and resource in place to deliver it
						*

2.2 Detailed Answers to Framework Principles

Principle 1: Start with a clear user need and public benefit

Using data in more innovative ways has the potential to transform how public services are delivered. We must always be clear about what we are trying to achieve.

Does everyone in the team understand the user need?

Yes

How does this benefit the public?

This proposed data sharing will enable collaborative research and analysis that provides evidence for important COVID-19 related decisions. Crucially, the increase in timely delivery of data will improve the evidence base for pandemic related decision making.

What would be the harm in not proceeding - what needs might not be met?

Sharing of census data already exists - this is a process change. Census microdata can be accessed for analytical projects under existing arrangements. However, we advise applicants that 6 months or more is a realistic timescale for delivery of data through that process. This is insufficient to meet the needs analysis related to the current public health crisis.

Do you have supporting evidence for the approach being likely to meet a user need or provide public benefit?

NRS and eDRIS have a proven track record of meeting user needs by delivering census data for use in research projects. eDRIS currently have expressions of interest in access to NRS data through the COVID-19 Data Holding.

Principle 2: Be aware of relevant legislation and codes of practice

You must have an understanding of the relevant laws and codes of practice that relate to the use of data. When in doubt, you must consult relevant experts.

List the legislation, codes of practice, and guidance that apply to your project.

GDPR & DPA-18
Census Act 1920
Code of Practice for Official Statistics
Approval for each project will be required from the Statistics Public Benefit and Privacy Panel
Approval for use of CHI as an (encrypted) linking variable has been formally approved by the CHI advisory group

Do all team members understand how relevant laws apply to the project?

There is a potential wide nature of projects, and the S-PBPP will ensure suitable review and assessment of laws relevant to each. Mandatory training information governance and security training is required for all staff engaged in processing or analysing the data. Census data is pseudonymised not anonymised and must be considered personal data. General Data Protection Regulation (GDPR) Recital 26 should be read and understood by researchers.

Identify your legal gateway and lawful basis for processing personal data under the GDPR.

GDPR Article 6(1)(e) “public task”, based on Census Act 1920 s4(2)

Processing special category data is exempted by GDPR Article 9 conditions (2)(g), (i) and (j), which in UK law are DPA Schedule 1(6), Schedule 1(3), and Schedule 1(4) respectively. Article 89(1) and DPA18 Schedule 2(27) exemptions apply as all outputs will be disclosure controlled.

Have you spoken to your information assurance team? – Yes

If using personal data, do you understand obligations under data protection legislation? – Yes. This has also been discussed with ICO Regions.

Principle 3: Use data that is proportionate to the user need

The use of data must be proportionate to the need. You must use the minimum data necessary to achieve the desired outcome.

If using personal data, have you answered the questions for determining proportionality? You must include evidence to support any decision.

> The user need for transfer of a 100% census cohort to the NSH is that due to its whole population coverage, census data will enable analysis of the impacts of COVID on small groups within the population, such as geographically small areas or occupational groups. Variables covering themes such as demography, economic activity, health and travel to work will be vital in assessing the vulnerable population, and economic impact of the current pandemic. This subset of census variables is not greater than we would normally consider for an individual research project.
>
> Whilst preparation of extracts is not normally the most time consuming part of the process of data delivery, we still expect the proposal to provide time savings in both extract creation, and linkage. Furthermore, an increased volume of requests for COVID related uses of census data could be beyond the capacity of the NRS Data Access team’s capacity to process. The processing of census extracts by eDRIS is therefore justified.
>
> Individual research projects will be provided with a bespoke cut that meets (but does not exceed) their requirements as approved by S-PBPP. The cohort provided will be subject to data minimisation on the following levels: cohort size & variables included.

If using personal data, what measures are in place to control access? How widely are you searching personal data?

The COVID-19 Data Infrastructure sits within the National Safe Haven alongside the existing infrastructure and its architecture has been developed on purpose to use the same design patterns. This means that the main data holdings would sit within a secure data management zone, with extract creation taking place there. Access by approved researchers to those extracts would be within a separate zone, and limited to those involved in their specific project.

How can you meet the project aim using the minimum personal data possible?

Each project will require S-PBPP review and approval, which will include data minimisation from the requested datasets.

Is there a way to achieve the same aim with less identifiable data?

All data will be pseudonymised before upload to NSH, who will further pseudonymise the widely distributed CHI field using encryption. Any projects seeking access to census data will have data minimisation applied at the S-PBPP project review stage.

Can you use synthetic data?

No

Has the data being used been provided for your analysis?

N/A

By using data that the public have freely volunteered, would your project jeopardise people providing this again in the future?

Under current rules, data would be provided for S-PBPP approved projects. The nature of the global pandemic and the requirement for high quality research to support public policy decision making provides a rational and legitimate justification for the change in the proposed process.

Could you clearly explain why you need to use that data to members of the public?

Yes, and this will be communicated through the Research Data Scotland website, and blog post by the Chief Statistician.

Is there a fair balance between the rights of individuals and the interests of the community?

This will be assessed on a per-project basis by S-PBPP.

Principle 4: Understand the limitations of the data

Data used in research must be well understood. It is essential to consider the limitations of data when assessing if it is appropriate to use it for a user need.

Describe the data sources being used.

Census 2011 (subset of variables covering demography, economic activity, health and travel to work), Scottish Longitudinal Survey and, when available, the Mother and Father Links.

Identify the potential limitations of the data sources and how they are being mitigated.

Census data is the gold standard population level data source. It is limited in that the population may have changed in distribution from 2011, and the questions in the data collection were not specifically designed to enable pandemic research. The supplementation with the Scottish Longitudinal Survey will assist in updating some aspects of distribution change, however this is a population sample which does not provide the same statistical power for the analysis of small population groups.

When complete, the Mother and Father links will enable inter-generational analysis of data, particularly for families not living in the same household. This is a tool that can link records from other datasets together rather than a dataset with research data attached. It will use encrypted CHI numbers as the basis for linkage and any additional risks associated with this data being included in the COVID-19 data holding will be assessed by the data controller, NRS Vital Events, before a final decision on its inclusion.

What processes do you have in place to ensure and maintain data integrity?

The 2011 census dataset, and SLS are the result of rigorous, well established QA procedures

Is there a plan in place to identify errors and biases?

The proposed methodology of any projects will be assessed at S-PBPP stage, giving relevant data experts the opportunity to comment on any likely biases. NRS will not be engaged in QA of project outputs, but will be reviewing outputs to ensure no release of personal information and would comment on any obvious errors at that stage.

What are the caveats?

The caveats around the use of census data are that 9 years have elapsed since collection of this dataset, and thus does not reflect changes in Scotland's population since then.

Principle 5: Use robust practices and work within your skillset

Insights from analysis are only as good as the data and practices used to create them. You must work within your skillset recognising where you do not have the skills or experience to use a particular approach or tool to a high standard.

Explain the relevant expertise and approaches that are being employed to maximise the efficacy of the project.

The COVID-19 Data Holding makes use of existing approaches and expertise from NRS on extracting and transferring census data, linkage, and reviewing proposed projects. Additionally, expertise from eDRIS and the Edinburgh Parallel Computing Centre on running data established infrastructure with a track record of safe storage and safe usage and delivery of sensitive data.

Describe the disciplines involved and why.

For the upload project only, NRS statistical, linkage and disclosure experts, along with advisors from our Census Security & Privacy and corporate Information Governance teams have been involved.

Is there expertise that the project requires that you don’t currently have?

Review of projects at S-PBPP stage requires a range of information governance and statistics experts from NRS and SG, plus independent members.

Have you designed the approach with a policy team or subject matter expert(s), taking into account subject matter context?

The proposed approach has been developed by discussion with security, statistics and information governance colleagues within NRS. Additionally, as a collaborative project across the public sector we have designed this approach alongside partners from SG, eDRIS, EPCC, HPS, ADR-S and HDR-UK.

How has reproducibility been ensured? Could another analyst repeat your procedure based on your documentation?

This will be assessed on a per-project basis by S-PBPP.

If using data about people, is it possible that your methodology uses proxies for protected variables which could lead to a discriminatory policy decision?

This will be assessed on a per-project basis by S-PBPP.

Principle 6: Make your work transparent and be accountable

You should be transparent about the tools, data and algorithms you used to conduct your work, working in the open where possible. This allows other researchers to scrutinise your findings and citizens to understand the new types of work we are doing.

Describe how you have considered making your research transparent and accountable.

The creation of a COVID-19 Data Holding will be communicated through explanatory pages of the RDS website, and through a blog post by the Chief Statistician. The privacy notice for the census will be updated to include details of this work.

How will you tell individuals about the use of their personal data?

The record level data used in research projects will be pseudonymised, and therefore direct contact with individuals included in research extracts is not possible. The use of census information in research in the public interest is explained when the data is collected, and on the Scotland’s Census website.

Do you need to amend your privacy notices?

The Scotland’s Census website privacy notice contains details on the use of census data in research projects and could be amended to include details of the COVID-19 Data Holding.

Have you considered how both internal and external engagement could benefit your project?

This will be assessed on a per-project basis by S-PBPP.

How interpretable are the outputs of your work?

This will be assessed on a per-project basis by S-PBPP.

How are you explaining how approaches were designed in plain English to other practitioners, policy makers and if appropriate, the public?

S-PBPP review would expect researchers to justify how they had communicated with and included input from the public/laypeople in their proposal.

Principle 7: Embed data use responsibly

It is essential that there is a plan to make sure insights from data are used responsibly. This means that both development and implementation teams understand how findings and data models should be used and monitored with a robust evaluation plan.

Describe the steps taken to ensure that outputs from your work are managed responsibly

Any proposed outputs using census data would be reviewed by NRS statisticians. Using established and robust protocols for statistical disclosure control, we would ensure that no personal information was contained in any outputs released, or could likely be deduced by comparison with other published sources of information.

How many people will be affected by any new model, insight or service arising from your work?

The COVID-19 Data Holding will be used as part of the evidence base in shaping the public sector response to the pandemic, therefore has potential to affect a large number of people.

Who are the users of your outputs?

The immediate users of our data will be researchers and public body analysts, with secondary users of their outputs being policy makers and service delivery.

Do users have the appropriate support and training to understand and maintain those outputs?

Yes. Immediate users will be assessed at S-PBPP stage to confirm their technical and topic matter expertise. Users will be expected to provide evidence of training and/or policies from their institution on security, including secure remote working. Users will be required to sign the Census Confidentiality Undertaking and by doing so demonstrate that they are aware of their obligations and of the consequences of unlawful disclosure of census data.

Have future events been planned for?

No.

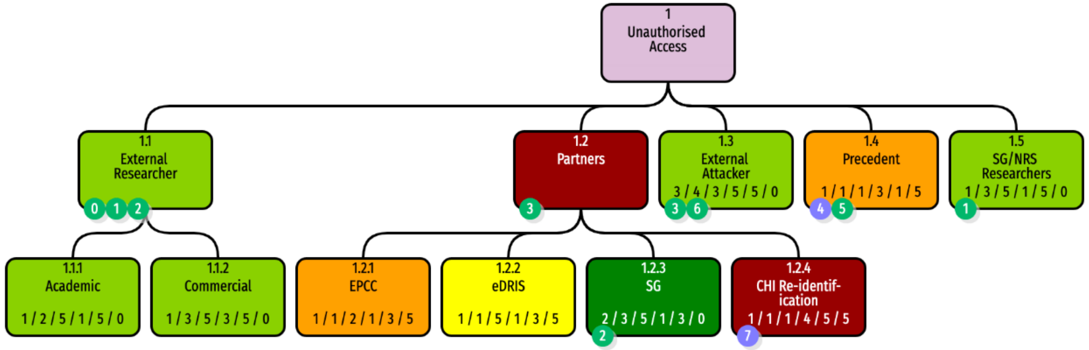
Is your implementation plan correlated with the impact of any particular model?

No, potential impacts of any particular research project will be assessed by S-PBPP.

How often will you report on these plans to senior reporting officer?

The proposed data sharing will require review and approval at director level in NRS before proceeding.

3. Risk Identification and Mitigation


1 Unauthorised Access
1.1 External Researcher
0 1 2
1.1.1 Academic 1 / 2 / 5 / 1 / 5 / 0
1.1.2 Commercial 1 / 3 / 5 / 3 / 5 / 0
1.2 Partners
3
1.2.1 EPCC 1 / 1 / 2 / 1 / 3 / 5
1.2.2 eDRIS 1 / 1 / 5 / 1 / 3 / 5
1.2.3 SG 2 / 3 / 5 / 1 / 3 / 0
2
1.2.4 CHI Re-identif-ication 1 / 1 / 1 / 4 / 5 / 5
7
1.3 External Attacker 3 / 4 / 3 / 5 / 5 / 0
3 6
1.4 Precedent 1 / 1 / 1 / 3 / 1 / 5
4 5
1.5 SG/NRS Researchers 1 / 3 / 5 / 1 / 5 / 0
1

Step 1: Identify and assess risks				
Describe source of risk and nature of potential impact on individuals.				
No.	**Risk and potential impact**	**Likelihood of harm** (Remote, possible or probable)	**Severity of harm** (minimal, significant or severe)	**Overall risk** (low, medium or high)
01	**There is a risk that:** Partner organisations, particularly EPCC and eDRIS will abuse their access to the uploaded Census data **Resulting in:** Possible re-identification of individuals **Due to:**	**Remote**	**Severe**	**Medium**

	The upload of a 100% cohort for a subset of census variables. Note: CHI has been an approved variable in previous uploads.			
02	**There is a risk that:** External attackers will abuse the internet access route in to the NSH copy to access Census data **Resulting in:** Disclosure of pseudonymised Census data and possible re-identification of individuals **Due to:** The necessary establishment of remote access to the NSH, in order to maintain appropriate social distancing creates the possibility of internet access to both full cohort and specific project data areas.	**Possible**	**Severe**	**High**
03	**There is a risk that:** Researchers apply for a 100% Census 2011 cohort for specific research projects **Resulting in:** Project proposals that are unjustifiable under the Data Minimisation principle **Due to:**	**Probable**	**Minimal**	**Medium**

	The knowledge of the presence of the full cohort within the Research Hub			
04	**There is a risk that:** Census data cohorts could be trivially re-identifiable using the CHI number variable **Resulting in:** Disclosure of individually identifiable personal data. **Due to:** CHI being a variable contained in numerous datasets available to the research and medical communities.	**Possible**	**Severe**	**High**

Step 2: Identify measures to reduce risk				
Identify additional measures you could take to reduce or eliminate risks identified above as medium or high risk.				
No.	**Options to reduce or eliminate risk**	**Effect of risk** **(**eliminated, reduced or accepted)	**Residual risk** (low, medium or high)	**Measure approved** (yes, no)
01	Both eDRIS and EPCC are trusted partners in the use of Census data for research proposals and deal with data of significantly more sensitivity from medical and other cohorts. All staff are appropriately trained and vetted and there is no history of abuse.	**Accepted**	**Medium**	***For NRS Privacy Group***

	The full COVID-19 cohort will be held in a separate area from the NSH, with specific access controls applied, restricting access to the 100% data set to the minimum necessary personnel.			
02	This access is permitted via remote desktop (VDI) installations within the NSH environment that have specifically limited permissions and use strong authentication. This environment is being specifically pen-tested by eDRIS as part of their control measures over the Research Hub implementation.	**Reduced**	**Medium**	***For NRS Privacy Group***
03	Data Minimisation will be explicitly managed by the S-PBPP and the necessity of adhering to data minimization is made clear to researchers in accompanying documentation.	**Accepted**	**Medium**	***For NRS Privacy Group***
04	The CHI variable will be encrypted. On receipt of the Census cohort, eDRIS will encrypt the CHI variable before any release to projects is undertaken. CHI has been included, in plaintext, in previous uploads where justified by the project requirements and suitably approved. eDRIS are a trusted partner for NRS and, as part of NHS shared services, are used to and have suitable perimeter and access controls for dealing with sensitive health data and the temporary, unencrypted CHI variable.	**Reduced**	**Low**	***For NRS Privacy Group***

	It is essential that eDRIS have access to the encryption keys, as they require to encrypt the corresponding variable in medical data sets to which NRS will not have access.			

4. Sign off and record outcomes

Item	Name/date	Notes
Measures approved by:		Integrate actions back into project plan, with date and responsibility for completion
Residual risks approved by:		If accepting any residual high risk, consult the ICO before going ahead
Data protection advice provided by:	Laura Mitchell, DPO	Your Data Protection Officer (DPO) should advise on compliance and whether processing can proceed
Summary of data protection advice: The DPIA demonstrated that proportionate controls will be applied to safeguard privacy. In particular, I am reassured to see that the ICO has been consulted; the data will all be pseudonymised; users of the data will be required to sign various undertakings to respect the confidentiality of the data; applications for the use of the data will have to be approved by the NRS Privacy Group and S-PBPP; appropriate information governance documentation will be in place. This is an exceptional arrangement to deal with exceptional circumstances and should not be viewed as setting any precedent. Any further proposal to repurpose elements of this data holding would need to be treated as a new proposal, evaluated on its own merits, and subject to all necessary governance and review.		

DPO advice accepted or overruled by:	L Cavin	If overruled, you must explain your reasons
Comments:		
This DPIA will kept under review by:	L Cavin	Your DPO should also review ongoing compliance with DPIA

APPENDIX A: TERMS FOR REMOTE ACCESS TO CENSUS DATA IN NATIONAL SAFE HAVEN

Due to the COVID-19 pandemic, Scottish Government guidance is that all those who are able to should work from home. Whilst this guidance is in effect, access for researchers to the physical Safe Haven room at the Bioquarter and other Safe Settings has ceased. As a temporary measure to enable the Recipient to continue research activity, remote access to the National Safe Haven is granted under the following terms:

1. Remote access will be for a period of 6 months from the signing of this document or until physical access to Safe Settings is reinstated, whichever is sooner.

2. Remote access will be via your institution’s virtual private network (VPN) only. You will not be able to directly access the National Safe Haven from your home IP address.

3. There will be no attempt to access the National Safe Haven from a location outside the EEA, or a country deemed by the Information Commissioner’s Office to have inadequate levels of protection even when connecting through your institution’s VPN.

4. Under no circumstances should all or any of the data be copied by hand from the workstation screen or attempts made to save screen shots or photograph the screen.

5. Under no circumstances should attempts be made to store information from the National Safe Haven on your computer, or on external storage devices (e.g. USB storage devices, memory pens/sticks, personal digital assistants (PDAs), etc).

6. You must not access the Safe Haven from any public place.

7. You must lock your computer at all times when not in use or you are away from your desk.

8. You must not permit friends or family to access your works computer and must lock your computer when they are in close proximity.

9. Any data breaches or hardware losses/theft should be reported to NRS and eDRIS, as well as your employer.

10. Any deviation from these terms could result in loss of your access to the Safe Haven, and for your organisation.

The Provider reserves the right to withdraw remote access to the data.

APPENDIX B: E-DRIS SECURITY STATEMENT

Version 3.2
22 May 2019

1. Introduction

1.1 This document provides a short security statement about the National Safe Haven (NSH).

1.2 The following key aspects of the environment are covered.

- Secure File Exchange
- Physical hosting and storage
- Backup and Business Continuity
- Access controls

1.3 National Services Scotland (NSS) Information Security has assessed and approved the overall system security of the National Safe Haven. In addition the information security departments of the Scottish Government (SG) and National Records of Scotland (NRS) have reviewed and approved the environment.

1.4 The Scottish Government Health Performance and Delivery Directorate has formally approved the National Safe Haven as an accredited safe haven under the terms of the Charter for Safe Havens (2015).

1.5 An independent Certified Cyber Security Consultancy (CCSC) assessment confirmed "that the National Safe Haven environment operates with a security profile consistent with what would be expected of a system processing 'OFFICIAL-SENSITIVE' data".

2. Secure File Exchange (SFE)

2.1 Transfer of demographic/indexed data from/to data providers uses Serv-U. This is certified as part of the National Safe Haven solution. NHSmail is used if both parties have NHSmail accounts.

2.2 Where indexing is carried out by the NSS Indexing Team, the transfer of demographic data/index numbers to/from NSS uses Globalscape. NHSmail is used if both parties have NHSmail accounts.

2.3 Where NRS Indexing is doing the indexing, the transfer of demographic data/index numbers to/from NRS uses xxxx.

2.4 Transfer of disclosure controlled results for a study from the Research Coordinator to the Researcher uses Serv-U or is done via email.

2.5 A request specific, password protected, time limited URL is generated by the SFE platform for the upload/download of data.

2.6 SFE is implemented as a single instance within the National Safe Haven.

2.7 Access to SFE to generate requests is only available from within the National Safe Haven.

- Web based access to the application is encrypted using Secure Socket Layer (SSL)

2.8 Only Research Coordinators and EPCC support staff have access to generate requests i.e. the request has to be initiated by a member of one of these teams.

2.9 The SFE platform is subjected to external penetration testing and server build review (last completed May 2019). Findings from the test are being addressed.

3. Physical hosting and storage

3.1 EPCC, part of the University of Edinburgh, operates the secure environment for the National Safe Haven.

- The infrastructure is located at the University of Edinburgh’s Advanced Computing Facility (ACF).

- The NSH infrastructure at ACF is isolated from the rest of the ACF infrastructure by the McAfee Firewalls.
- Access to the ACF is limited to the University staff who operate the facility and contractors (e.g. those employed by equipment vendors) but only in the presence of University staff.
- The NSH infrastructure is subjected to external penetration testing and server build review (last completed May 2019), and findings from the test are being addressed.

3.2 NSS and NRS office locations, where the indexing will take place, employ good practice physical controls including a clear desk policy and all employees completing mandatory security awareness and training.

3.3 The Indexing team store the demographic data on a secure, access controlled file server in their organisation in order to perform the indexing. Only the Indexing team have access to this area.

3.4 Linked payload data for a study is stored in its own secure study area within the National Safe Haven. Researchers and Research Coordinators access this area using their unique user account details to perform their analyses. Access to the secure study area is enabled through two-factor authentication.

3.5 Research Coordinators have a Citrix VDI Management Node virtual desktop and Researchers have a Citrix VDI Analysis Node virtual desktop.

- The Citrix remote access platform has been configured to remove any functions not needed by end users e.g. internet access, copying and pasting into or out of the National Safe Haven, printing.
- The Citrix user environment and server builds are, like the SFE service, subject to penetration testing and findings are being addressed.

3.6 On completion of the study the linked payload data is archived from the active system and all permissions are revoked from user accounts. PROTECT
eDRIS Security Statement Page 5 of 5

4. Backup and Business Continuity

4.1 Data security and back up is provided through the use of two separate data centres operated by EPCC/UoE, thus protecting Researchers from the risk of data loss. Both data centres are operated to the same standard (refer to 3.1).

4.2 While a study is active, nightly backups to tape are taken of the linked payload data and Researchers analysis and output files.

4.3 Completed study data is moved to a secure long term storage area held within the EPCC/UoE data centres. This area is not accessible to Researchers.

4.4 No backups are taken of data on the SFE server.

5. Access controls

5.1 The Research Coordinator is responsible for verifying the identity of the researcher.

5.2 The Research Coordinator is responsible for checking the approval status of the Researcher. Only Approved Researchers are able to access the service. Researchers must undertake approved information governance training and sign the eDRIS User Agreement.

5.3 The Researcher will be expected to provide evidence to the Public Benefit and Privacy Panel for Health and Social Care (PBPP) (or to the Research Coordinator if PBPP is not required) that that they have received permission from the data providers for NSS to receive the data for linkage.

5.4 Once the pre-requisites (5.1 to 5.3) have been met the Research Coordinator initiates the creation of user accounts with EPCC.

5.5 User accounts for Citrix are controlled by EPCC. Note that EPCC will not verify the Researcher’s details. The Research Coordinator is responsible for this step (5.1 to 5.3).

5.6 Permanent SFE user accounts are set up for Research Coordinators. This enables Research Coordinators to request upload/download of files by remote file exchange partners using the request specific URLs generated (2.5).

5.7 Disclosure controls are applied to the final outputs requested by the Researcher before the results are released.

- The Research Coordinator will disclosure control the outputs.
- Their assessment will be checked and authorised for release by an Information Consultant or Head of Service.
- The Research Coordinator will use the SFE or email to send the outputs to the Researcher.
- The Researcher cannot export the results themselves.