

DATA VETTING REQUESTS

– Standard Operating Procedure –

1. PURPOSE

1.1 To outline the process to be followed, by NB-IRDT Approved Users and Employees, to request vetting of (research) data results. Data result must be vetted prior to disclosing them with anyone not on the Research Team (section 2 of most current & approved DAR). If this process is not followed, no data results can be removed from the secure research environment. Following this process is mandatory.

2. SCOPE

2.1 This procedure applies to all NB-IRDT data released from the secure research environment including data access requests and deliverables for information manager agreements.

3. DEFINITIONS

3.1 Approved Users: Individuals, such as NB-IRDT employees, researchers, students, and government employees, who have been issued an electronic identification card, personal identification number, and project access account following the approval of access according to all relevant NB-IRDT procedures, including a Criminal Record Check (CRC).

3.2 Conflict of Interest: A conflict between an individual's duties and responsibilities as a NB-IRDT employee, or committee member, and that person's professional, business, personal / familial or public interests. Conflicts of interest cause divided loyalties and can lead to biased judgment. Conflicts of interest can be real, potential, or perceived.

- Potential: Arises when a NB-IRDT employee / Committee member becomes aware of an outside interest that may influence the member's ability to act with objectivity, independence, and integrity.
- Perceived: Arises when it appears to others that an outside interest would influence a NB-IRDT employee/Committee member's ability to act with objectivity, independence, and integrity.
- Real: Arises when an outside interest influences a NB-IRDT employee/Committee member's ability to act with objectivity, independence, and integrity.

3.3 Data: Any information used for research, quality assurance or statistical purposes, including personal health information, which is disclosed to under a Data Sharing / Access Agreement.

3.4 Disclosure: making available, sharing, or releasing any data, regardless of format (written, verbal, online, etc.) from the secure custody of NB-IRDT whether authorized under legislation and or agreements or unauthorized such unvetted data results, loss, theft, etc.

3.5 Employees: All full-time and part-time persons currently earning wages or salary from NB-IRDT (including the Director).

3.6 Project Data: Data prepared for a specific project and stored in a project folder, which consists only information approved for a specific research project.

3.7 Vetting: Prior to the disclosure of data from the secure research environment (SRE) a set of vetting rules are applied to the proposed release. The application of these rules protects against the unintentional disclosure of personal or personal health data. These rules do not assess the quality of the results. When output is produced and presented for vetting it is the responsibility of the data user to assess the quality of their results. NB-IRDT screens the output request for data protection and not for publishable quality.

4. PROCEDURES

4.1 Data Vetting Requests

Process Step	Description	Actor
1. Vetting request submitted	<p>When ready to request vetting of results (for disclosure), an Approved User, from the project team, completes a vetting package and stores the vetting package in the 'Disclosure Request' subfolder of the project.</p> <p>The package includes:</p> <ul style="list-style-type: none"> • a Vetting Request Form (to be filled out by the Approved User - see Appendix A), • the data requested for release, and • supporting files, see Appendix B.2 for more details. <p>When ready, the Approved User notifies the Database Administrator (DBA) by email that the completed vetting package is ready for review.</p>	Approved User
2. Vetting package review assigned	The DBA may review the vetting package themselves or assign the review of the vetting package to a Senior Data Analyst (SDA).	
3. Vetting package reviewed	<p>The DBA or SDA reviews the package, applying NB-IRDT vetting rules and ensuring vetting requirements are met.</p> <p>These rules include but are not limited to:</p> <ul style="list-style-type: none"> • verification that aggregated results match the approved project objectives • verification of sample descriptions for aggregated results • verification of variable names and labels • check for risk of residual disclosure 	NB-IRDT Database Administrator or Senior Data Analyst

	If the request for release does not meet the vetting requirements, the DBA/SDA will suggest changes and communicate them to the approved user in writing.	
3. Vetting packaged edited as necessary	If edits are needed, the Approved User revises the vetting request form, output files, and/or supporting files. The Approved User may also choose to withdraw the vetting request. When the revisions are completed, the Approved User re-submits and re-notifies the DBA by email.	Approved User
4. Vetting package completion	Steps 2 and 3 may be repeated until all the vetting requirements are met.	Approved User NB-IRDT Database Administrator or Senior Data Analyst
5. Results release prepared	The DBA/SDA adds the files to be released into the NB-IRDT Released Data Archive. These files are indexed by project number, the Approved User's username, and date. Using the secured desktop ¹ in the server room, the DBA/SDA copies the files for release to a dedicated USB stick. The DBA/SDA logs on the NB-IRDT desktop ² with Internet connection in the server room and sends the files for release to the Approved User by email. ¹ There is only one USB enabled computer in the server room within the SRE. Only the DBA has access to this computer. ² This is the only computer in the SRE has internet connection.	NB-IRDT Database Administrator Senior Data Analyst

4.2 Data Vetting Requests – In the absence of NB-IRDT Database Administrator or in event the NB-IRDT *Conflict of Interest Policy* applies

Process Step	Description	Actor
1. Designated Data Analyst (DDA) is assigned.	In preparation of the absence, the NB-IRDT Database Administrator will assign a Senior Data Analyst to the role of vetting data disclosure requests.	NB-IRDT Database Administrator

	<p>The DBA will send an email notification to all Approved Users stating who will be their replacement during their absence.</p> <p>If the NB-IRDT Conflict of Interest Policy applies, a Senior Data Analyst will be selected in consultation with the NB-IRDT Director.</p>	
2. Assignment credentials	The NB-IRDT Director informs the NB-IRDT Systems Administrator to create an account for the SDA with credentials needed to access and vet data disclosure requests.	NB-IRDT Systems Administrator
3. Data disclosure requests steps 1-5 apply	Approved Users follow all steps in 4.1, above, with the Senior Data Analyst who was assigned by the DBA.	Approved User Senior Data analyst

Note: If disagreement occurs during the Data Vetting Request process, there is a mechanism in place for resolution (Data Vetting Committee).

5. DOCUMENT VERSION, REVIEW, AND APPROVAL HISTORY

Version	Author	Nature of Change	Date
1.0	Afton Lavigne /Donna Curtis Maillet	Document development	April 7 2022
			Review Date April 2023

Version	Author	Nature of Change	Date
1.1			

Version	Author	Nature of Change	Date
1.2			

APPENDIX A – VETTING REQUEST FORM

NB-IRDT Confidentiality Vetting Request Form

Name:	Email:	
User Name:	Project #:	Date: (YYYY-MM-DD)
Project Title:		
Folder name:		

	Yes / No
1. Variables are labeled, especially recoded variables (provide a list of recoded variables in supporting docs folder)	
2. Sub-samples are clearly described in Output Request form	
3. For simple descriptive statistics, each statistic (e.g., mean, frequency) meets the minimum cell size requirement. <ul style="list-style-type: none"> <i>Outputs are labeled.</i> <i>Categories have been collapsed or table dimensions reduced, if necessary.</i> 	
4. Do you have graphs or model outputs which are equivalent to simple descriptive statistics? (e.g., histograms, a model with one categorical independent variable or a model with all possible interactions – consult your Analyst if you are not sure) <ul style="list-style-type: none"> <i>If Yes, please provide the corresponding frequency table provided in supporting documentation</i> 	
5. Is there a correlation or covariance matrix with categorical variables in your output? <ul style="list-style-type: none"> <i>If Yes, please provide the corresponding cross-tabulation table provided in supporting documentation</i> 	
6. All applicable rules have been applied. For example, rounding, weighting, and any additional database specific rules when applicable.	

Residual Disclosure Risk: Comparison with previous vetting requests	
1. Has a version of this output, in part or in whole, been previously released?	
2. Have any variables been recoded or modified even slightly?	
3. Has the sub-sample changed?	
4. Have individual cases or outliers been dropped?	
If the answer is YES to any of these, discuss with the DBA	

Total number of files for release¹:

File Name	Database name and years	Specify method used or insert number from list below	Subsample Description (e.g., employed females, aged 21-45, in New Brunswick)
1			
2			
4			
5			
6			
7			
8			

Types of Methods

1. Descriptive methods (e.g., Frequencies, Cross-tabular analysis, means and distributions)
2. Scaling methods (e.g., Factor Analysis)
3. Graphs (e.g., histograms) – please remember to include supporting tabulations
4. Regression methods (e.g., OLS, ANOVA, Logistic, Probit, Tobit)
5. Complex methods of modeling (e.g., Structural equation modeling, Hierarchical Linear Modeling, Growth analysis, survival analysis, Event History Analysis, Simultaneous-equations Models, Fixed Effects Models, Random Effects Models)

¹ Please use a different row for each requested file and add more rows as required.

These files are to support the vetting request and **will not be released**.
Total number of supporting files:

Place these files in your Supporting Documents folder

1. Supporting frequencies

Reminder: Correlations and linear or logistic regressions with a single independent variable give results equivalent to descriptive statistics. Therefore, for categorical dependent variables provide the tabular frequencies and for continuous dependent variables provide the sample sizes.

2. Other files as required by the database-specific Guidelines when applicable

3. Syntax files (if requested by the analyst)

File name	Notes

Additional comments which may be helpful to the analyst:

APPENDIX B.1 - VETTING CHECKLIST

Task	Consider	Task Complete
1. Have you checked the latest set of vetting requirements for your dataset(s)?	<ul style="list-style-type: none"> • Vetting rules can change over time • The minimum cell count rules are strictly applied. • Expect stricter rules to be applied for sensitive variables and narrowly defined populations 	Yes <input type="checkbox"/> No <input type="checkbox"/>
2. Have you left yourself enough time?	<ul style="list-style-type: none"> • Be sure to consider the processing timelines associated with preparing and reviewing a vetting package • Allow yourself more time if (a) output has been previously released, (b) if the vetting request is large, and/or (c) if vetting during a 'peak' period. 	Yes <input type="checkbox"/> No <input type="checkbox"/>
3. Are the results in this vetting request within the scope of:	<ul style="list-style-type: none"> • The accepted proposal? • A typical NB-IRDT project? 	Yes <input type="checkbox"/> No <input type="checkbox"/>
4. Are you sure your outputs are correct?		Yes <input type="checkbox"/> No <input type="checkbox"/>
5. Has your supervisor/PI/CI's/other stakeholders reviewed the outputs and approved the vetting request? Are they familiar with the vetting rules for your dataset(s)?		Yes <input type="checkbox"/> No <input type="checkbox"/>
6. Do you/your team really need to have access to the requested output outside of the NB-IRDT?		Yes <input type="checkbox"/> No <input type="checkbox"/>
7. Is this your final output for the project?	<ul style="list-style-type: none"> • If no, will the release of the output prevent the release of future requests? • Do you really need to have access to the requested output outside of NB-IRDT lab at this time? • Have you considered or taken steps to try to lessen the risk of potential issues with the release of future requests? 	Yes <input type="checkbox"/> No <input type="checkbox"/>
8. Submitted a vetting package that satisfies the confidentiality vetting requirements?	<ul style="list-style-type: none"> • Fill out the vetting request form- must be complete and accurate. Vetting requests will not be considered until this has happened. • Provide files that contain outputs that can be released. • Provide all of the required supporting files in the correct formats (e.g., excel, word, etc)? • Flag any potential issues or concerns with your vetting request? 	Yes <input type="checkbox"/> No <input type="checkbox"/>
9. Made sure the DBA or DDA is:	<ul style="list-style-type: none"> • Aware that you have submitted a vetting request. 	Yes <input type="checkbox"/> No <input type="checkbox"/>

	<ul style="list-style-type: none"> • Able to contact you if there is a question? • Aware of your deadline(s). Note that while the DBA/DDA will do their best to meet your deadline, this cannot be guaranteed. 	
--	--	--

APPENDIX B.2 - Supporting Material Checklist

Material	Confirmed
1. All of your syntax (e.g., preparing the analytical sample(s)/sub-setting, variable creation/modification, and analysis)?	Yes <input type="checkbox"/> No <input type="checkbox"/>
2. Have you linked (merged data from different databases, e.g. Citizen Database and NB Physician Billing data) or pooled data (put together multiple datasets of same layout, e.g. multiple years of DAD datasets)? If yes, be sure to describe your approach.	Yes <input type="checkbox"/> No <input type="checkbox"/>
3. Provide a description of the geographic region(s) you have used.	Yes <input type="checkbox"/> No <input type="checkbox"/>
4. Have you subsetted or selected only a certain set of respondents from the dataset(s) for all or part of the analysis? (e.g., the analysis is only looking at males over the age of 50)? If 'yes,' then: <ul style="list-style-type: none"> • Provide a description of your approach to sub-setting • Consult with the DBA. You may need to provide the counts for the excluded cases. 	Yes <input type="checkbox"/> No <input type="checkbox"/>
5. Have you included or excluded cases that are 'not applicable' or 'valid skip'? Consult with the DBA. You may need to provide the supporting counts for these cases.	Yes <input type="checkbox"/> No <input type="checkbox"/>
6. Have you created, derived, or transformed any variables? This includes collapsing or recoding. If 'yes' then: <ul style="list-style-type: none"> • Ensure that your variables have been labelled; this includes the levels of variables (categories). • If you have previously released output and have recoded variables, then provide a description of the change(s) and consult with the DBA. 	Yes <input type="checkbox"/> No <input type="checkbox"/>
7. Descriptive statistics and equivalent <ul style="list-style-type: none"> • Does every count / descriptive statistic being requested for vetting meet the minimum cell size requirement? If not, consult with the DBA. • Provide the number of individuals used to produce each output. 	Yes <input type="checkbox"/> No <input type="checkbox"/>
8. Tables <ul style="list-style-type: none"> • Have they been titled and labelled? • Consult with the DBA about the format – you may need to use excel or word. • Include row and column totals for supporting counts. • For large vetting requests, provide an appendix with a list of the tables. 	Yes <input type="checkbox"/> No <input type="checkbox"/>

<p>10. Rounding</p> <ul style="list-style-type: none"> • If there was rounding, note how was the rounding conducted. • Note if there was any forced rounding to zero? 	<p>Yes <input type="checkbox"/> No <input type="checkbox"/></p>
<p>11. Sensitive variables (e.g., lower levels of geography, narrowly defined populations, dollar-value variables, and other sensitive information). Consult with the DBA – you may need to provide additional supporting information.</p>	<p>Yes <input type="checkbox"/> No <input type="checkbox"/></p>
<p>12. Residual disclosure risk. Do you have:</p> <ul style="list-style-type: none"> • Time-Based Variables (e.g., has a depression diagnosis in years 1 and 2 (yes/no) • Non-mutually exclusive variables/the potential for overlap with aggregated variables (e.g., has COPD and had a respiratory condition) • Has a version of this output, in part or in whole, been previously released? <p>If 'yes' to any of the above, then consult with the DBA – you may need to provide additional supporting information (residual checks- which should be in excel format).</p>	<p>Yes <input type="checkbox"/> No <input type="checkbox"/></p>
<p>13. Do you have graphs, maps, or correlations?</p> <ul style="list-style-type: none"> • If you can make your graph or map outside the NB-IRDT using vetted data, please do so. • Consult with the DBA – you may need to provide additional supporting information. 	<p>Yes <input type="checkbox"/> No <input type="checkbox"/></p>