REB review of projects involving secondary use of data

June 2015

BEST PRACTICE DOCUMENT
TIPS AND TRICKS

Produced by the Ethics Working Group of the MSSU Data /Privacy Advisory Committee

MSSU - USSM
MARITIME SPOR SUPPORT UNIT -
UNITE DE SOUTIEN DE LA SRAP MARITIME

Table of contents

INTRODUCTION 3

IDENTIFICATION OF DATA SOURCE(S) 3

ILLUSTRATIVE TOOLS – SPECIFICITY OF DATA ACCESS 4

ACCESS TO PERSONAL IDENTIFIERS 4

DATA TRANSFER 5

RESEARCH DESIGN – DATA LINKAGE 5

JUSTIFICATION FOR CONSENT WAIVER 6

APPENDIX A: DATA FLOW DIAGRAMS 7

APPENDIX B: PAN-CANADIAN MORTALITY STUDY 8

APPENDIX C: NS PET PROJECT 10

INTRODUCTION

This document is aimed at supporting researchers and the REB review process of research projects that may involve:

- the use of data that are not collected primarily for research purposes
- the use of data held by the government or a `data custodian`
- hundreds or thousands of observations in the analysis
- historical data on subjects, who may have since died or moved
- a justification for a consent waiver.

IDENTIFICATION OF DATA SOURCE(S)

If the research involves the use of data held by an external agency or custodian, the formal name of the data holding(s) as well as the name of the agency/custodian should be clearly identified.

If a formal access process is required to obtain the data, a letter from the custodian to the project lead confirming the feasibility of the study should accompany the REB application. Best practice would suggest that the principal investigator should meet with the data custodian in the early stage of research project formulation to ensure that the research purpose is feasible within the context of the available data.

The feasibility of the work is, in part, assessed by way of two aspects of the study design:

1. The specificity of inclusion/exclusion criteria to identify eligible study records from a given data holding
2. The list of variables associated with the eligible subset of records, including a justification for each – broadly identified as variables required for analytical purposes and/or linkage purposes (if applicable).

Please note that this level of detail is usually required by data custodians to give approval, in principle, and conditional on successful ethical review, for any research study.

ILLUSTRATIVE TOOLS – SPECIFICITY OF DATA ACCESS

New Brunswick Institute for Research Data and Training (NBIRDT) Data Access Application

http://www.unb.ca/fredericton/arts/nbirdt/_resources/pdfs/application-data-access-11-2014.pdf

Health Data Nova Scotia Data (HDNS) Request Form

http://medicine.dal.ca/departments/department-sites/community-health/research/hdns/services/Forms.html

Nova Scotia Department of Health and Wellness (NSDHW) Data Access Application

(soon to be released)

ACCESS TO PERSONAL IDENTIFIERS

The data extraction and linkage, if applicable, should be designed in order to minimize the access of both data custodians and study team members. Individuals who will have access to patient-level identifiers at various points in the process should be explicitly identified and noted in a data flow diagram.

In the case of simple data extraction from one database, the custodian will not normally release any personal identifiers unless one of these variables is required for some form of derivation work that the custodian is unable to provide (e.g., postal code mapped to a geographic unit).

In the case of multiple datasets involving linkage, the custodian is routinely the only one to have access to both the identifiers and the analytic data together. If the person carrying out the linkage is a project team member, and not a data custodian, then the person carrying out the linkage (i.e., with access to the identifiers) should not the person who will analyze the final dataset.

In terms of analytical datasets, generally only the designated analyst has access to record-level data - the other team members would see only aggregate results.

Identification of the 'touch points' for data access can be clearly illustrated by way of a data flow diagram, which is recommended when the project involves extractions, and possibly linkage, from more than one source dataset.

Illustrative tool: Data flow diagrams (see appendices for examples).

DATA TRANSFER

If the research involves the transfer of data between custodians, and possibly among members of the research team, it should be clear how the data transfer will take place. Best practice would suggest this be done by way of a secure file transfer method (e.g. secure email transfer, encrypted USB).

RESEARCH DESIGN – DATA LINKAGE

The linkage of two or more datasets should be planned in keeping with the following principles:

1. Analytic data from one custodian should not be shared with another custodian
2. Identifiers (i.e., variables that identify an individual or institution) should not `travel` with analytic data

A data flow diagram is an easy way to help design the process of the linkage and then to clearly explain this flow, including the `touch points` in the process, specifying who will carry out the extraction, who will arrange for data transfer, who will receive the data, and who will analyze the dataset. It should also be clear at

which point along this path where a study identification number is used as well as the existence of identifiers (i.e., variables that identify an individual or institution).

Illustrative tool: Data flow diagrams (see appendices for examples).

JUSTIFICATION FOR CONSENT WAIVER

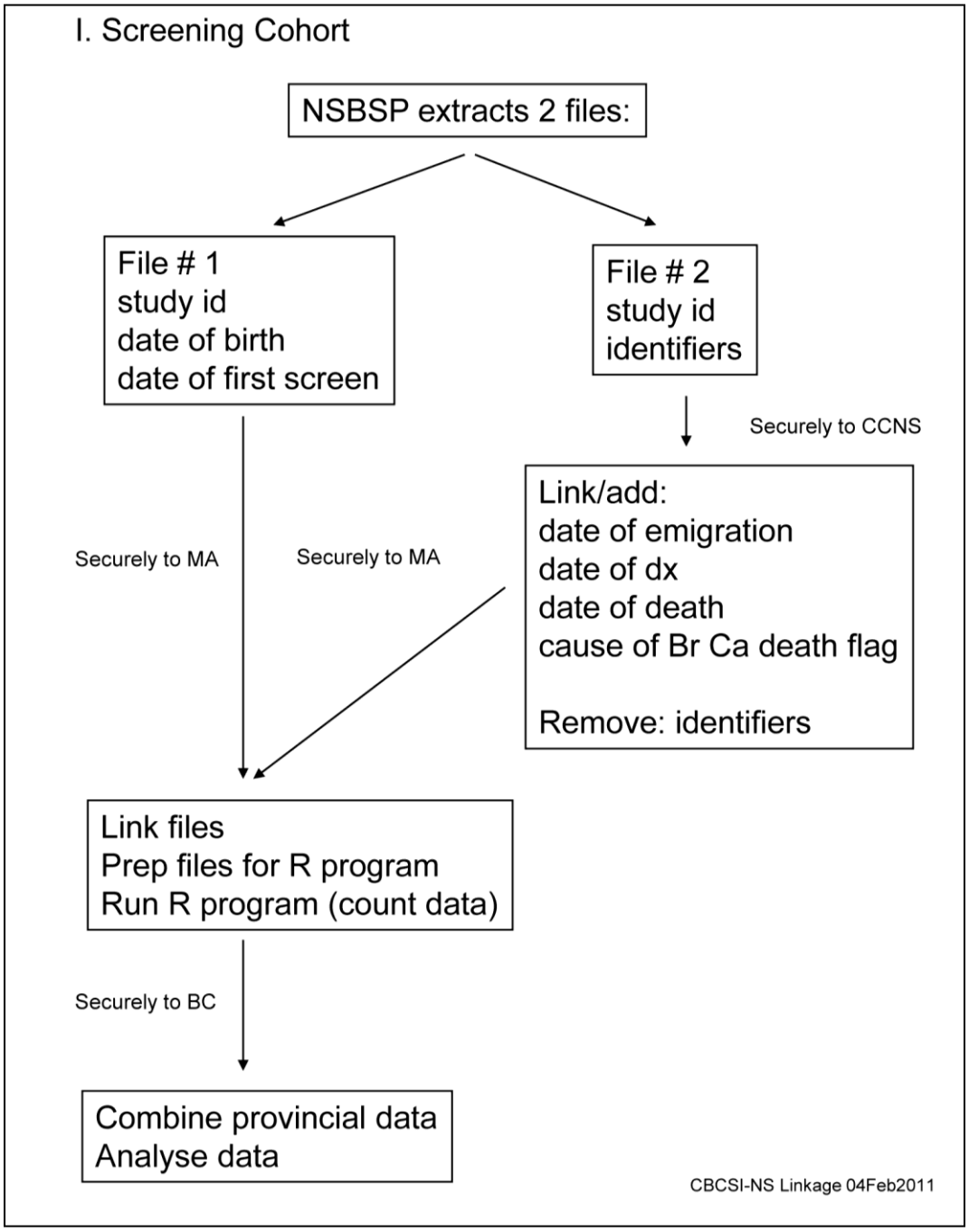
REB Application forms will generally have space where a justification can be made for a consent waiver, if applicable. There are several grounds that can be used for such a waiver.

Illustrative tool:

IWK Health Centre Consent Waiver Form (under 'Privacy')

http://www.iwk.nshealth.ca/research/application-materials-forms

APPENDIX A: DATA FLOW DIAGRAMS

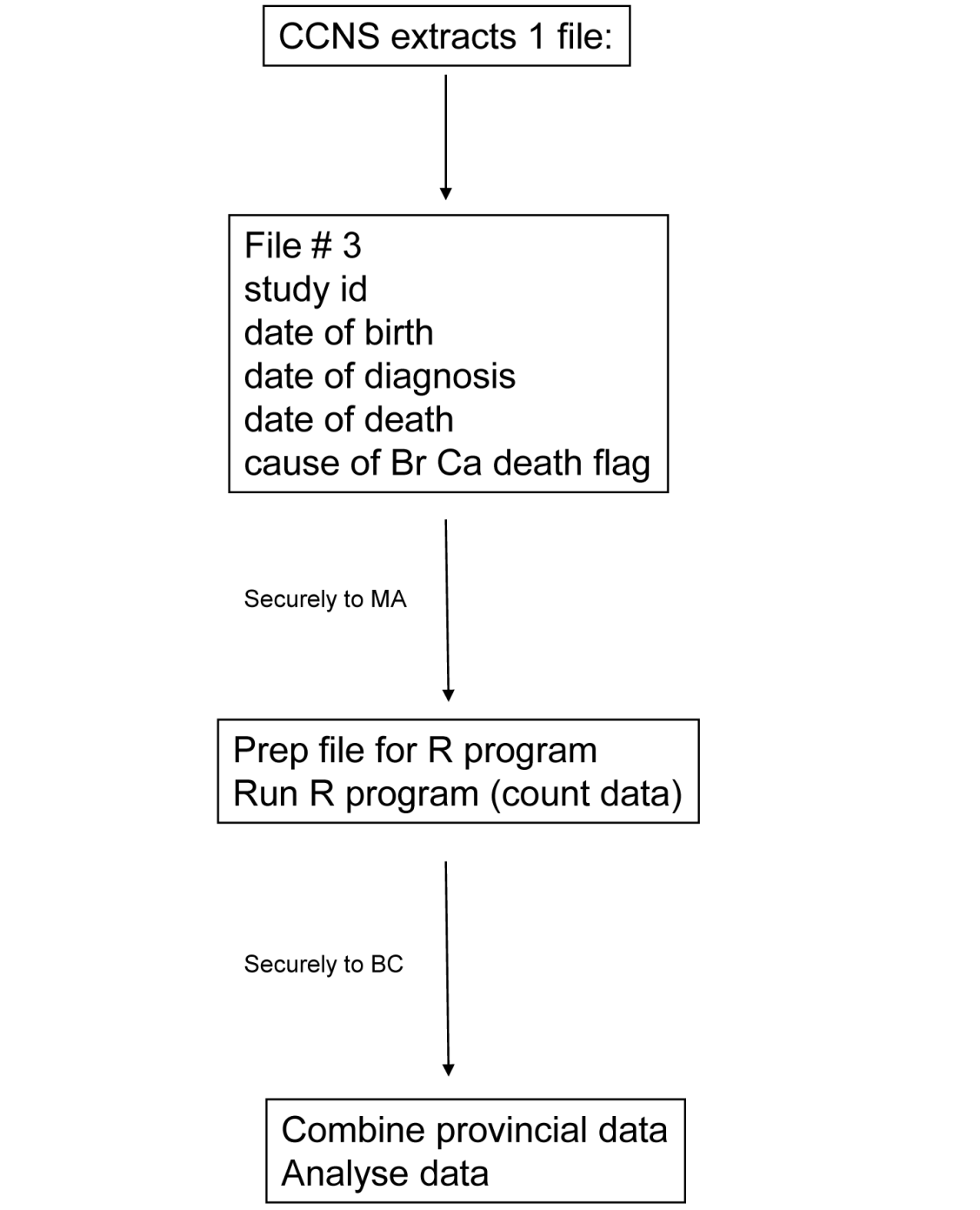


APPENDIX B: PAN-CANADIAN MORTALITY STUDY

NS Team: Jennifer Payne, Mohamed Abdolell, Judy Caines, Theresa Foley

This diagram describes the data flow for the NS arm of a multi-provincial breast cancer mortality study. Each province followed a common protocol and executed a common analytical plan (R programs) to produce aggregate data which then flowed to a statistician in British Columbia (BC) for inter-provincial comparisons and combined analysis. This diagram was used to describe how data would flow between the Nova Scotia Breast Screening Program (NSBSP), Cancer Care Nova Scotia's Cancer Registry (CCNS) and the research team (biostatistician identified by his initials). The diagram describes the data flow to produce an analytic file for two different cohorts (screening cohort & cancer cohort) and specifies the actual variables that flow, highlighting the use of personal identifiers and a study identification number. The data flow was designed to minimize sharing of data between the two custodians (NSBSP and CCNS) and to ensure that the biostatistician did not receive any identifiers (note that date of birth was required to be maintained for analytical purposes – calculation of person-time).

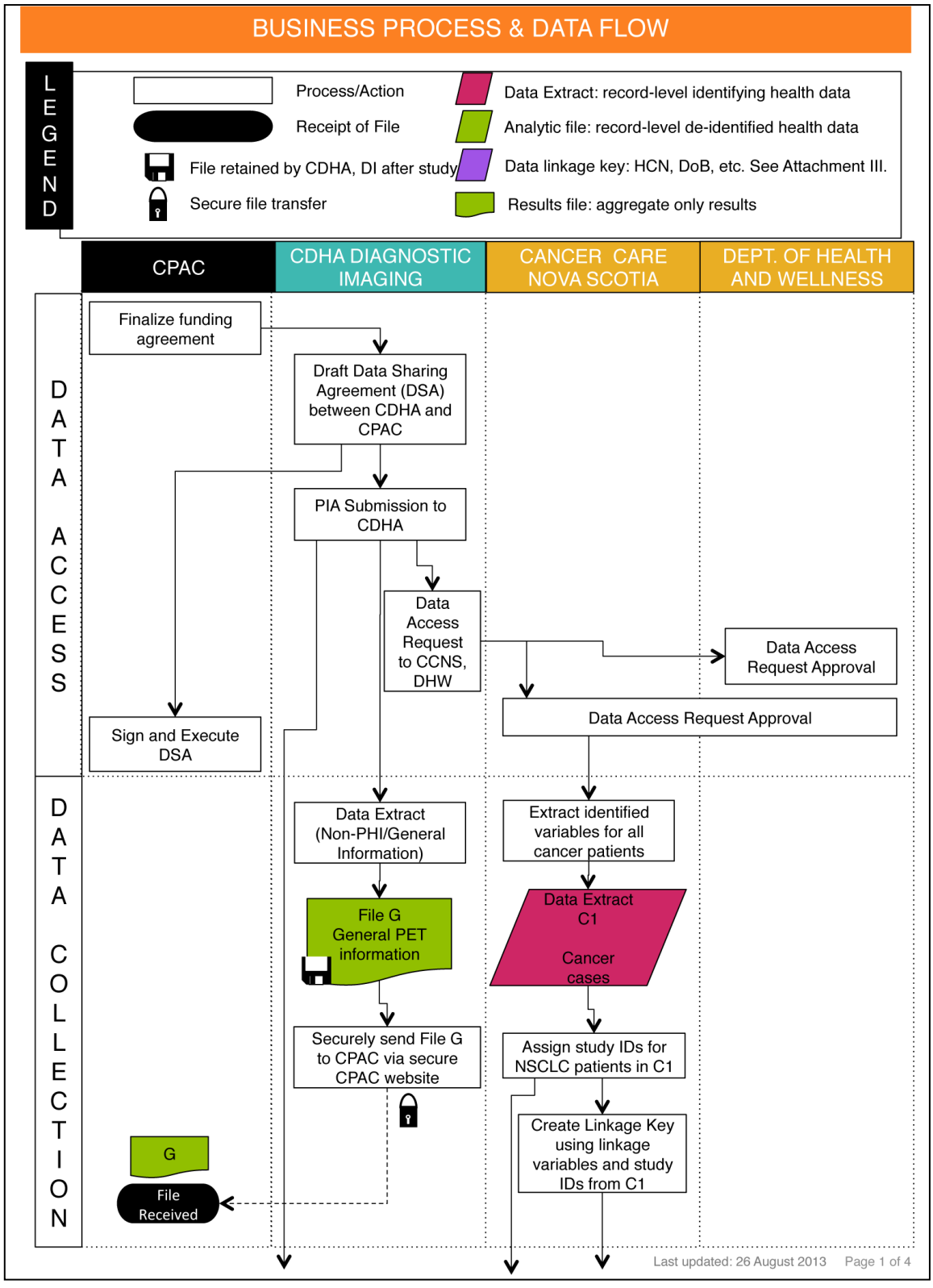
II. Breast Cancer Cohort

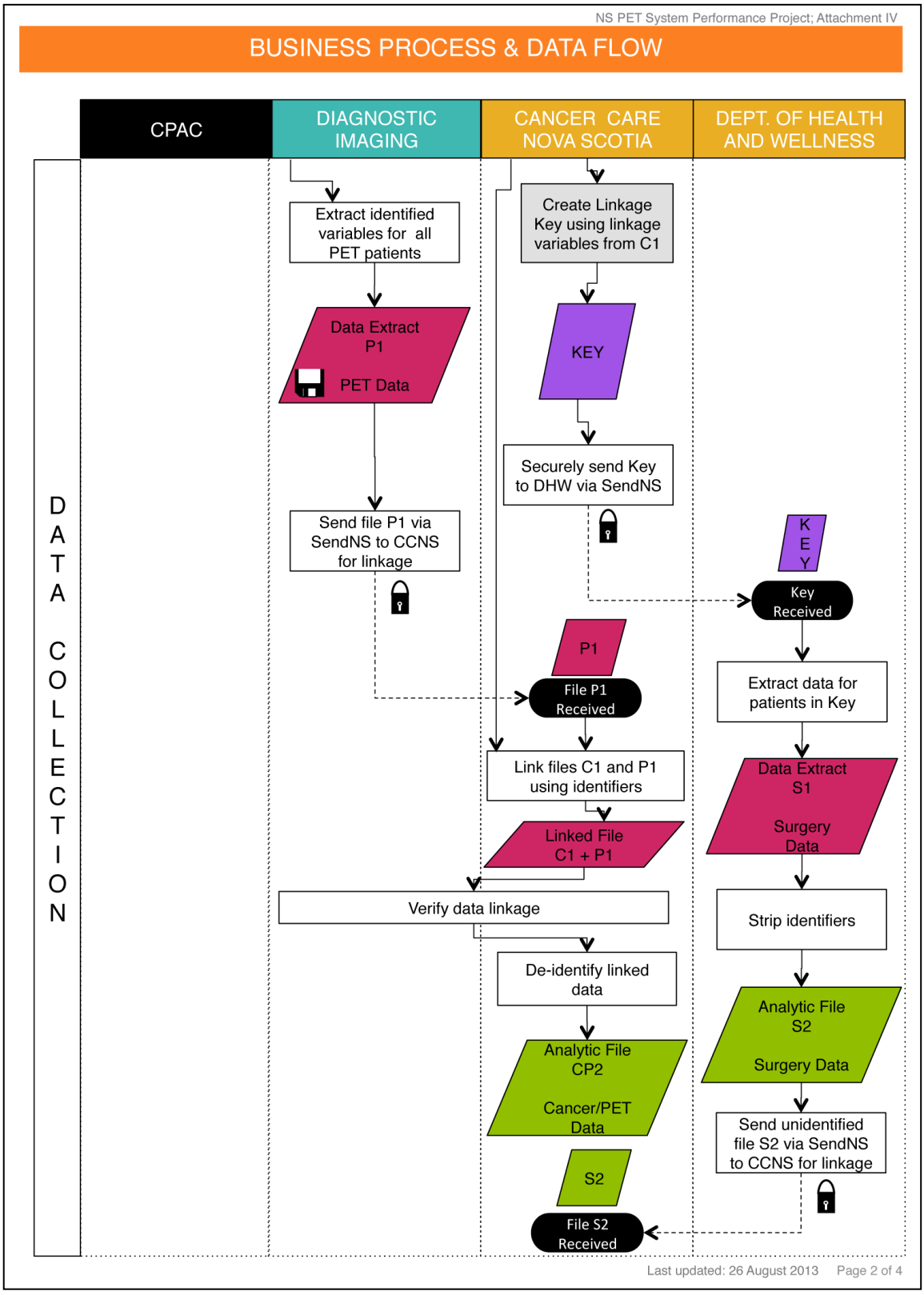


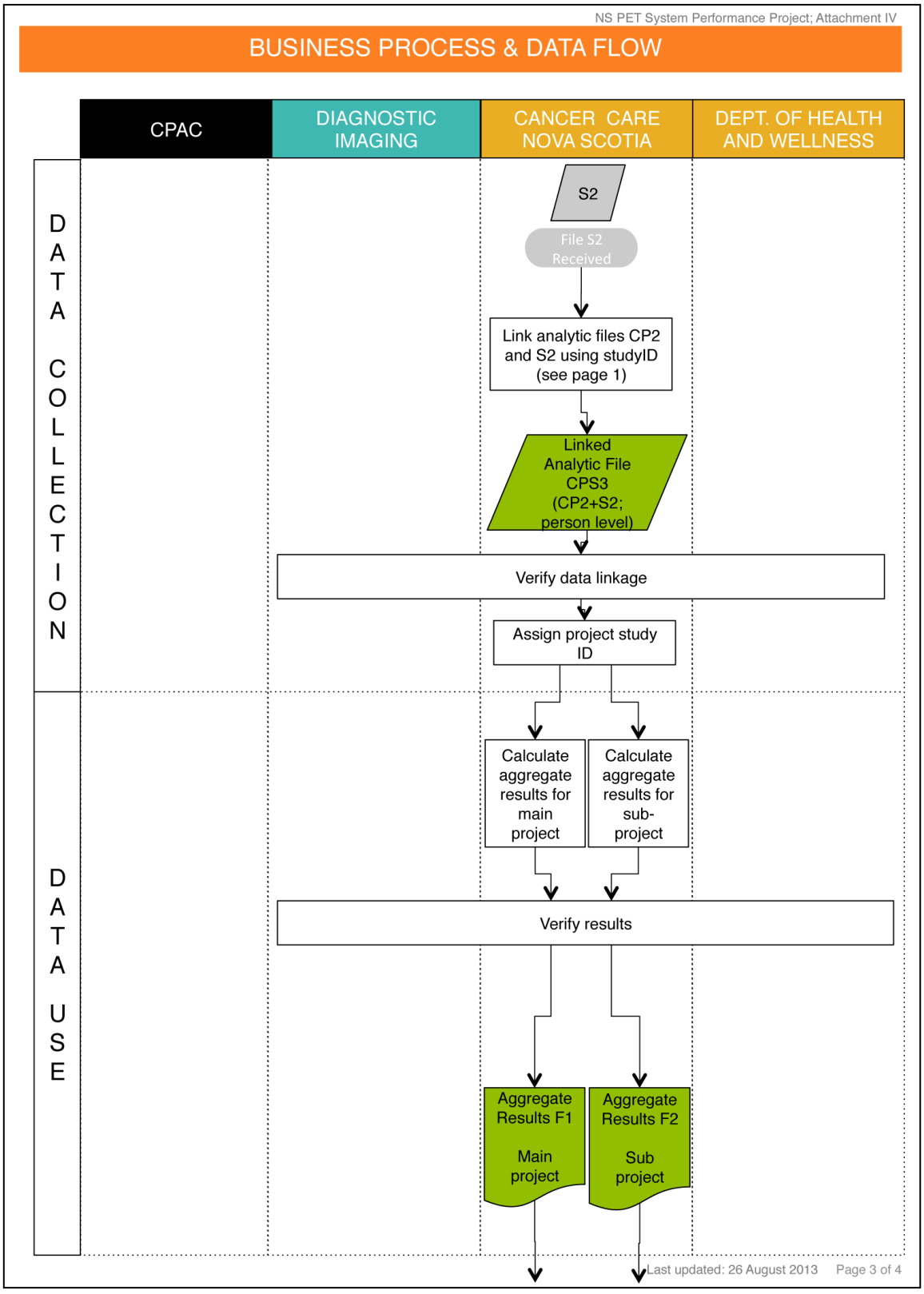
APPENDIX C: NS PET PROJECT

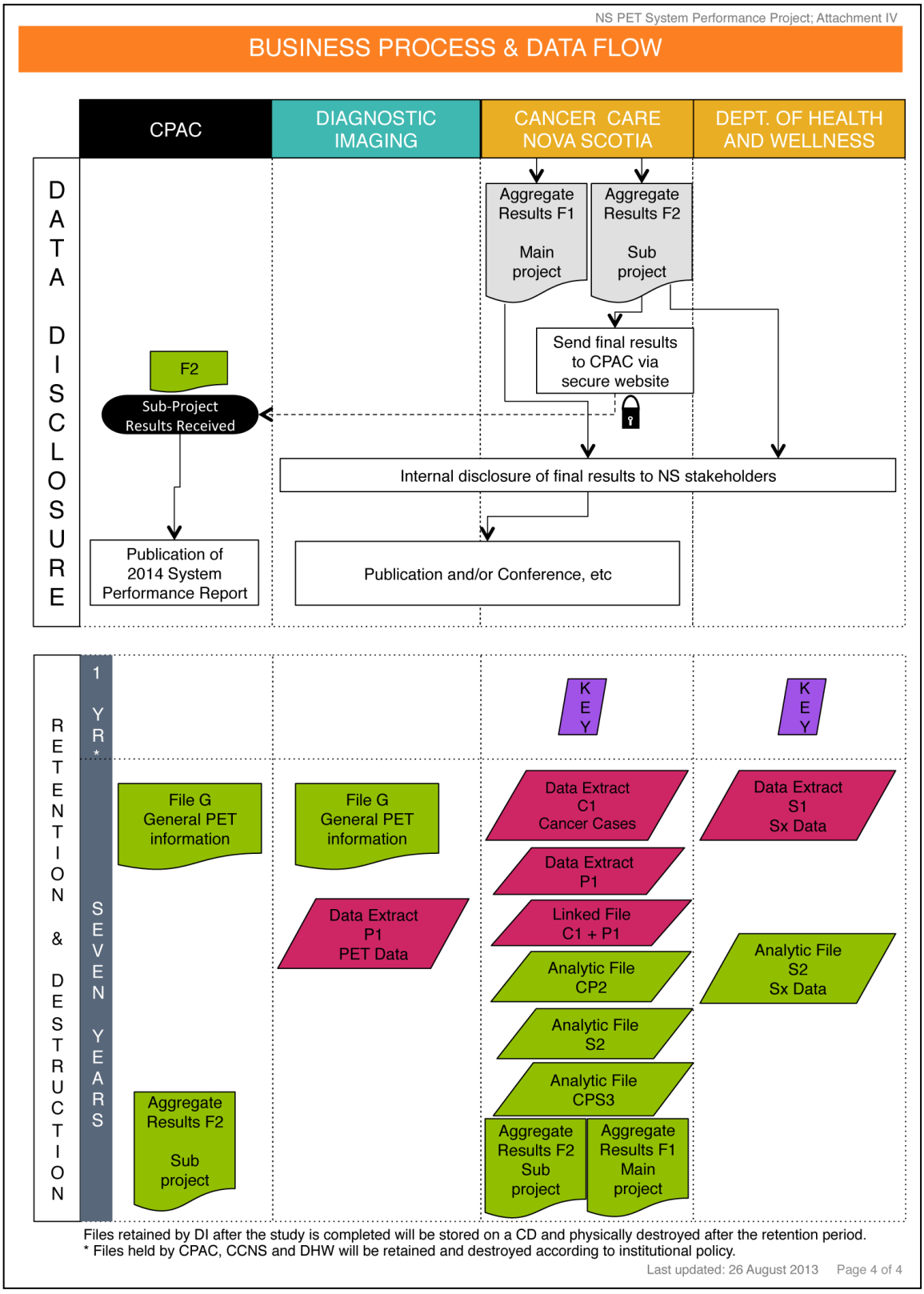
NS Team: Andrew Ross, Jennifer Payne, Ron Dewar, Devbani Raha, Kaitlyn Tsuruda

This diagram describes the data flow for the NS arm of a multi-provincial study of the use of PET diagnostic imaging in the diagnosis and management of non-small cell lung cancer, which was then expanded in NS to allow an assessment of the use of PET imaging in the management of cancer as a whole. This diagram was used to support a data access application with the end result being a Privacy Impact Assessment (PIA), as this was not a research project. The diagram describes the data access and data flow processes through to retention, using the language of the NS Personal Health Information Act (i.e., collection, use, disclosure, retention, destruction). The project was funded by the Canadian Partnership Against Cancer (CPAC) and was hosted by Capital District Health Authority's (CDHA) Dept of Diagnostic Imaging and involved Cancer Care Nova Scotia's (CCNS) Cancer Registry. The detail provided was meant to outline how the data flowed, including identifiers and analytic data, how the data would be transferred, and who would retain the data.


BUSINESS PROCESS & DATA FLOW
LEGEND
Process/Action
Receipt of File
File retained by CDHA, DI after study
Secure file transfer
Data Extract: record-level identifying health data
Analytic file: record-level de-identified health data
Data linkage key: HCN, DoB, etc. See Attachment III.
Results file: aggregate only results
CPAC
CDHA DIAGNOSTIC IMAGING
CANCER CARE NOVA SCOTIA
DEPT. OF HEALTH AND WELLNESS
DATA ACCESS
DATA COLLECTION
Finalize funding agreement
Draft Data Sharing Agreement (DSA) between CDHA and CPAC
PIA Submission to CDHA
Data Access Request to CCNS, DHW
Data Access Request Approval
Sign and Execute DSA
Data Access Request Approval
Data Extract (Non-PHI/General Information)
Extract identified variables for all cancer patients
File G General PET information
Data Extract C1 Cancer cases
Securely send File G to CPAC via secure CPAC website
Assign study IDs for NSCLC patients in C1
Create Linkage Key using linkage variables and study IDs from C1
G
File Received
Last updated: 26 August 2013
Page 1 of 4


NS PET System Performance Project; Attachment IV
BUSINESS PROCESS & DATA FLOW
CPAC
DIAGNOSTIC IMAGING
CANCER CARE NOVA SCOTIA
DEPT. OF HEALTH AND WELLNESS
DATA COLLECTION
Extract identified variables for all PET patients
Data Extract P1
PET Data
Send file P1 via SendNS to CCNS for linkage
Create Linkage Key using linkage variables from C1
KEY
Securely send Key to DHW via SendNS
K E Y
Key Received
P1
File P1 Received
Link files C1 and P1 using identifiers
Linked File C1 + P1
Verify data linkage
De-identify linked data
Analytic File CP2
Cancer/PET Data
S2
File S2 Received
Extract data for patients in Key
Data Extract S1
Surgery Data
Strip identifiers
Analytic File S2
Surgery Data
Send unidentified file S2 via SendNS to CCNS for linkage
Last updated: 26 August 2013
Page 2 of 4


NS PET System Performance Project; Attachment IV
BUSINESS PROCESS & DATA FLOW
CPAC
DIAGNOSTIC IMAGING
CANCER CARE NOVA SCOTIA
DEPT. OF HEALTH AND WELLNESS
DATA COLLECTION
DATA USE
S2
File S2 Received
Link analytic files CP2 and S2 using studyID (see page 1)
Linked Analytic File CPS3 (CP2+S2; person level)
Verify data linkage
Assign project study ID
Calculate aggregate results for main project
Calculate aggregate results for sub-project
Verify results
Aggregate Results F1 Main project
Aggregate Results F2 Sub project
Last updated: 26 August 2013
Page 3 of 4


NS PET System Performance Project; Attachment IV
BUSINESS PROCESS & DATA FLOW
CPAC
DIAGNOSTIC IMAGING
CANCER CARE NOVA SCOTIA
DEPT. OF HEALTH AND WELLNESS
DATA DISCLOSURE
Aggregate Results F1 Main project
Aggregate Results F2 Sub project
Send final results to CPAC via secure website
F2
Sub-Project Results Received
Internal disclosure of final results to NS stakeholders
Publication of 2014 System Performance Report
Publication and/or Conference, etc
RETENTION & DESTRUCTION
1 YR *
SEVEN YEARS
KEY
KEY
File G General PET information
File G General PET information
Data Extract C1 Cancer Cases
Data Extract S1 Sx Data
Data Extract P1
Data Extract P1 PET Data
Linked File C1 + P1
Analytic File CP2
Analytic File S2 Sx Data
Analytic File S2
Analytic File CPS3
Aggregate Results F2 Sub project
Aggregate Results F1 Main project
Aggregate Results F2 Sub project
Files retained by DI after the study is completed will be stored on a CD and physically destroyed after the retention period.
* Files held by CPAC, CCNS and DHW will be retained and destroyed according to institutional policy.
Last updated: 26 August 2013
Page 4 of 4