

# The Evolution of Banff in the Context of Modernization

Darren Gray  
Statistics Canada

UNECE Workshop on Statistical Data Editing  
Neuchâtel, Switzerland, 18-20 September 2018



100

STATISTICS CANADA  
ONE HUNDRED YEARS AND COUNTING

STATISTIQUE CANADA  
CENT ANS BIEN COMPTÉS



## What is Banff?

- Generalized edit and imputation (data editing) system developed and maintained by Statistics Canada
- Primary tool for business surveys at Statistics Canada
- External users include other Canadian agencies, other National Statistical Institutes (NSIs) and some private organizations

# Modernization

- Statistics Canada is moving beyond a survey-first approach with new methods and integrating data from a variety of existing sources
- Existing methods must be reviewed for suitability, while new methods are tested and developed for use
- How do we maintain the “generalized” aspect of Banff (focusing on methods with broad application) while facilitating the integration of emerging methods in an efficient manner?



## Presentation Outline

- Overview of the Banff research & development plan
- Description of Banff as a process flow tool
- Project details



# Banff Research and Development Plan



Current needs

Future needs

Internal methodology

Discrete /  
categorical variables  
Non-linear  
constraints

Integration of  
external methods  
into Banff  
framework

Process flow using Banff

Simplify / improve  
user control

Integration of  
external methods  
into Banff  
framework

Generalized assessment  
and evaluation tools

Measure impact of  
data editing on  
statistical outputs

Assess and  
compare new  
methods





**Current needs**

**Future needs**

**Internal methodology**

Discrete /  
categorical variables  
Non-linear  
constraints

**Integration of  
external methods  
into Banff  
framework**

**Process flow using Banff**

**Simplify / improve  
user control**

**Generalized assessment  
and evaluation tools**

Measure impact of  
data editing on  
statistical outputs

Assess and  
compare new  
methods



**Process flow using Banff**

Simplify / improve  
user control

Integration of external  
methods into Banff  
framework

1) Standardization of  
Banff procedures

Address data  
management “holes”

Framework for assessing  
and “wrapping” external  
methods

2) Improvements to  
Banff processor

User control and  
flexibility

Users control integration  
of external method into a  
process flow

3) Introduction of  
record-level statuses

Simplifies certain  
procedures

Allows for a wider variety  
of external methods





# Process flow using Banff

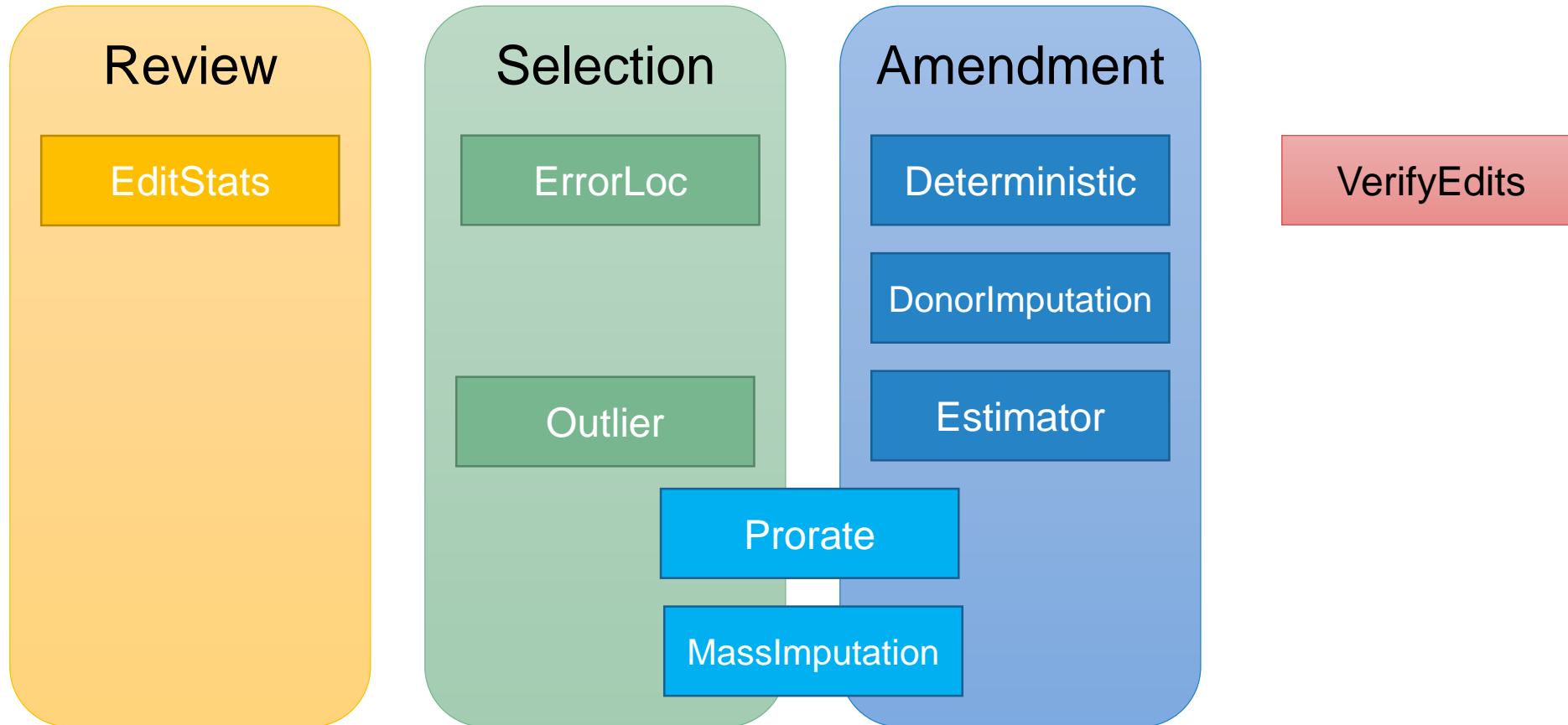


## Description of Banff as a modular, automated data editing tool

- Banff consists of nine SAS procedures, performing various edit and imputation related functions
- With one exception, these procedures fit naturally within the Generic Statistical Data Editing Model (GSDEM) developed by the UNECE



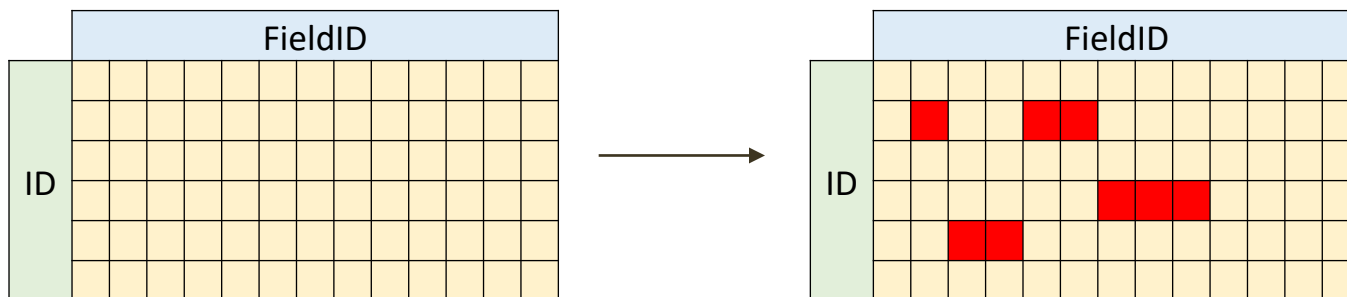
# GSDEM Data Editing Function Classifications





## Module interaction

- Banff passes information from one procedure to another using *status flags*
- Selection procedures identify values requiring further treatment

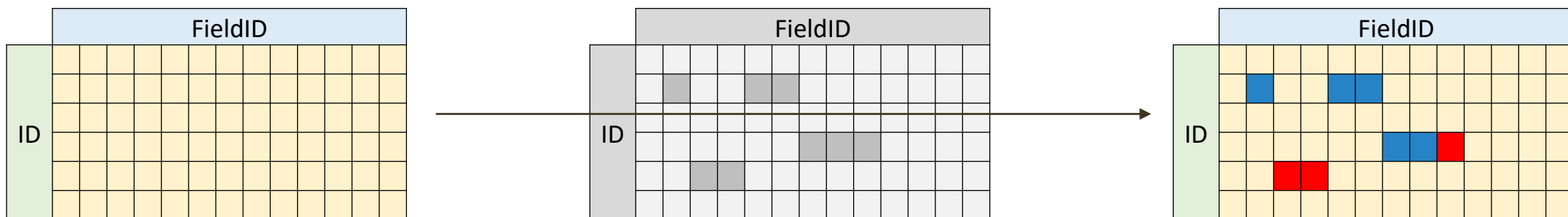






## Module interaction

- Banff passes information from one procedure to another using *status flags*
- Selection procedures identify values requiring further treatment
- Amendment procedures make changes to the data
- All-in-one procedures do both





# Project details



100



# 1) Standardizing the Banff procedures

- Objective: develop formal criteria defining a modular, automated function within the Banff data editing framework
- Elements:
  - Function classification
  - Standardized inputs and outputs
  - Reserved status flags



- Most procedures already fit this criteria, but we have identified some gaps
- Key change: requiring that procedures output updated statistical data and status files

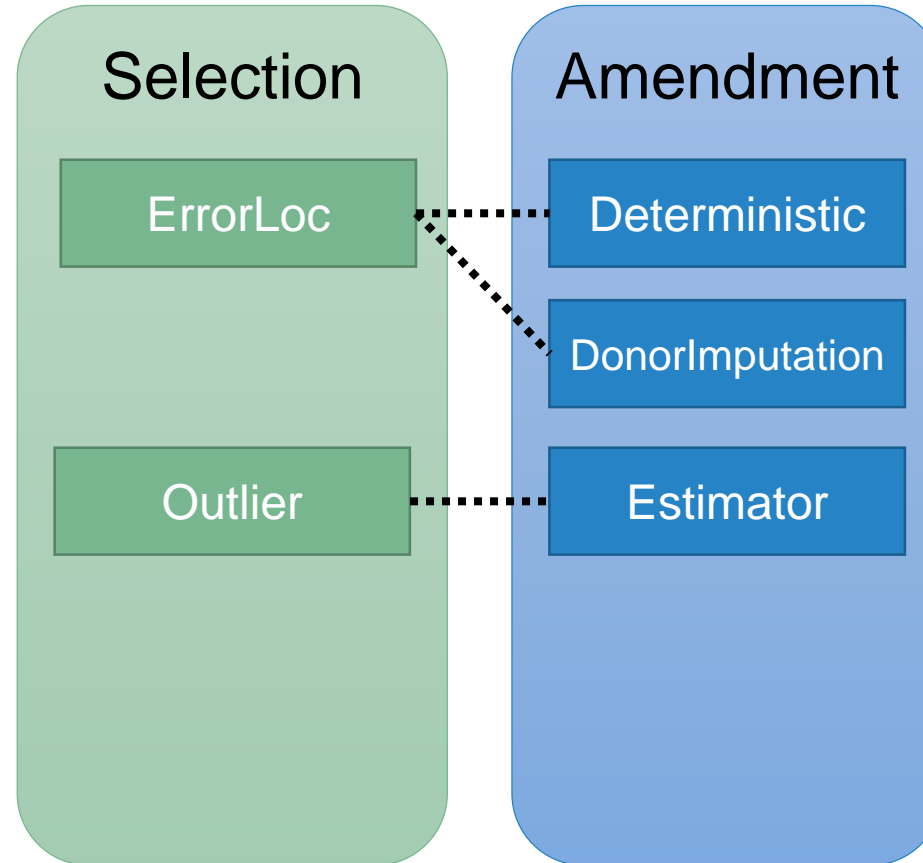


## 2) Improving the Banff processor

- Banff processor: Metadata-driven tool designed to facilitate the use of Banff procedures in a sequential data editing process flow
- Advantages:
  - Data management between steps is handled behind the scenes
  - Overall process flow and procedure parameters are all stored in a single, editable table
- Drawback: limited control of process flow details

## Linking procedures together

- Process flows often need to restrict how procedures interact
- Example: some amendment procedures are designed to fix specific errors





## Linking procedures together

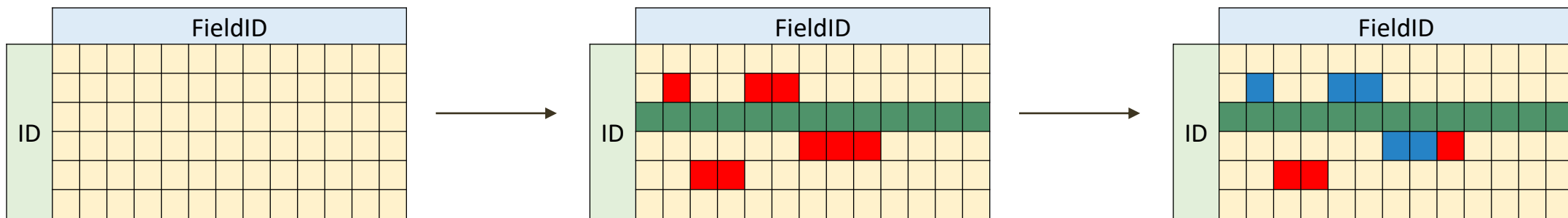
- Some restrictions are built into the Banff processor
  - e.g., ErrorLoc and DonorImputation linked by “edit groups”
- Proposal: give control to users by implementing *status filters*





### 3) Record-level statuses

- Status flags are currently associated with specific values, identified by <ID, FieldID>
- Certain information can only be captured at the record level, but is required for subsequent procedures
- Propose introducing record-level status fields





## Remarks

- Standardizing the Banff modules is a key component of our plan to integrate external methods within the Banff framework
- Presents opportunity for open-source collaboration

# THANK YOU! MERCI!

For more information  
please contact,  
[Darren.Gray@canada.ca](mailto:Darren.Gray@canada.ca)



#StatCan100



Statistics  
Canada Statistique  
Canada

[www.statcan.gc.ca](http://www.statcan.gc.ca)

Canada