UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS
Expert Meeting on Statistical Data Editing
3-7 October 2022, (virtual)

Banff's next step: an open-source data editing system for advanced tools and collaboration

Darren Gray, Statistics Canada
darren.gray@statcan.gc.ca

I. Introduction

1. The Banff team is pleased to announce that development has begun on a new version of Banff (Statistics Canada, 2017a), with a planned release in 2024. This will be the most significant change to Banff since its production release in 2002, dramatically increasing user access to the most advanced statistical data editing (SDE) tools available. The development includes three core components:

   (a) Banff will no longer rely on the SAS architecture, shared instead as a free, standalone software package. While the structure and format of this release has not been finalized, the objective is to make Banff as widely accessible as possible. All existing functionality will be included in the conversion.
   (b) Existing Banff procedures that directly perform data editing functions will be updated to a modified modular format. This standardized format will define the input and outputs for all Banff modules, based on the type of statistical data editing functions performed. Banff will provide guidelines and tools allowing users to convert open-source or custom programs to this standard, thereby facilitating the integration of external tools into the Banff framework.
   (c) The Banff Processor will be completely redesigned, with new features improving process flow design, functionality, convenience, efficiency, and diagnostics.

2. The redeveloped Banff system will provide SDE practitioners from all fields a free, generalized platform to design and process large-scale statistical data editing projects. Once released, the Banff team hopes to work with colleagues in the data editing community to build and maintain a diverse catalogue of advanced, expert-vetted data editing tools, open to everyone.

A. Development History

3. The production version of Banff was released in 2002, built on the methodology of the Generalized Edit and Imputation System, or GEIS (Kovar, MacMillan and Whitridge, 1988). A summary of this release by Kozak (2005) highlights many of the same general objectives as the current redevelopment, with respect to flexibility, ease-of-use and modularity: *"For that initial release of Banff, the goal was to reproduce the methodology used in GEIS, but in a more user-friendly and flexible system. . . . The modules in GEIS are linked to one another, while the individual SAS procedures in Banff are independent of one another. . . . Banff is intrinsically much simpler to use and more flexible than GEIS."*

4. Subsequent releases offered new functionality, and introduced the Banff Processor (Statistics Canada, 2017b), a metadata-driven tool designed to automate the edit and imputation process in a sequential data editing process flow. Over the years, Banff's user base grew to include not only Statistics Canada, but other National Statistical Offices (NSOs), other Canadian government agencies, and various private institutions.

5. Plans for future development began in earnest in the mid-2010s and are presented by Thomas (2017), who at the time noted that the scope of development "*could be very small and include minor enhancements to the system or, if warranted, it could be a complete overhaul of the system and its functionality.*" As part of the review process, consultations with users and stakeholders (internal and external to Statistics Canada) were conducted, alongside discussion with other data editing experts. Amongst other objectives, Thomas noted an interest in "*linking Banff to other international modules or systems in order to increase its functionality to process modern data*". Banff's black-box nature was noted as a drawback of the system, in the eyes of external users who were interested in the software, as was its reliance on SAS. Discussing SAS and open-source alternatives, he states: *"Statistics Canada has a strong dependency on SAS and Windows / Unix. Most of the systems that it is using are developed directly using SAS or by developing SAS procedures in C. There is a strong push to consider cheaper, more flexible, options like R for a programming language or Linux as an operating system. Many R 'packages' are readily available and easily shared in an open-source environment. At Statistics Canada, there are several groups evaluating R as an alternative option. There will have to be considerations for options in R, or other statistical languages, that will meet the needs in the development of Banff."*

6. A concrete proposal for improvements to Banff as an automated, modular data editing tool were shared by Gray (2018). This proposal included the following features that are implemented in the current redevelopment:

   (a) Standardized Banff modules: a criterion to define modular, automated data-editing procedures in the Banff framework. This includes definitions of mandatory inputs, outputs, and metadata, based on the type of function performed.
   (b) An improved Banff processor allowing process controls based on status flags.
   (c) The incorporation of open-source software into the Banff framework, as demonstrated using the R package *missForest*[1].

7. Finally, over the past several years, Statistics Canada has undergone an internal review to assess opportunities for analytical diversification; with respect to the agency's suite of generalized software systems, this included a review of current functionality, discussion of corporate business needs, and an assessment of external software alternatives, either proprietary or open-source. The case for Banff, and in particular the advantages of incorporating external SDE software, was made to the Advisory Committee on Statistical Methods (Gray, D. and Matthews, S., 2020). In consideration of open-source packages, the committee gave a series of recommendations when exploring the use of open-source software within the generalized systems, which included the consideration of various use cases, guidelines for the vetting and incorporation of open-source systems, and collaboration with the user and developer community.

B. Release timeline

8. Approval for the Banff redevelopment was finalized in early 2022, with a plan to release the new version in March 2024. This includes new releases of Banff's nine procedures, and a new Banff Processor. For testing and continuity purposes, no new functionality will be introduced to the Banff procedures as they migrate from SAS. However, throughout the development phase, the Banff team will work both on identifying external software packages that might suitably be integrated into the Banff framework, and identifying gaps in the community of SDE tools that may require future in-house development. Following the release in 2024, the Banff team will focus on supporting existing Statistics Canada programs, as well as external users, as they migrate to the new system.

[1] https://cran.r-project.org/web/packages/missForest/index.html

9. The remainder of this paper highlights some of the key changes of the redevelopment. Section II reviews relevant SDE concepts and terminology. Section III discusses the Banff modules, and how standardization will facilitate the integration of external SDE tools into the Banff framework. The new features of Banff's revised Processor are presented in Section IV.

II. Generic Statistical Data Editing Model: Concepts and Terminology

10. The Generic Statistical Data Editing Model (GSDEM) (UNECE, 2019) was developed under the HLG-MOS "ModernStats" initiative, and is a useful reference for statistical data editing (SDE) terminology and models. Wherever possible, efforts have been made to align Banff's redevelopment proposal with the GSDEM, and in this section, we review some of the key concepts and terminology.

11. The SDE process, also commonly referred to as editing and imputation (or E&I), typically consists of one or more steps whose purpose is to review and treat data to achieve fitness for use. In the presence of non-response (at both the unit and item level) and measurement error, it is generally considered an essential part of improving data quality.

12. The SDE process is often broken into many smaller **process steps**, each of which performs one or more key functions. The GSDEM categorizes statistical **data editing functions** into the following types:

(a) **Review**. Functions that examine the data to identify potential problems.
(b) **Selection**. Functions that select units or fields within units for specified further treatment.
(c) **Treatment**. Functions that change the data in a way that is considered appropriate to improve the data quality. The modification of specific fields within a unit (i.e. filling in missing values or changing erroneous ones) is referred to as imputation.

13. There is a natural link between the three function types: "*review functions lead to quality indicators or measures that can point out specific problems in the data; selection functions take quality indicators and/or selection criteria (thresholds) and data as input and produce indicators identifying records or fields within records for further treatment; finally, treatment functions change or impute the selected data values in order to resolve the problems detected earlier.*" (UNECE)

14. These process steps typically need as inputs not just the statistical data that is the subject of the SDE process, but also input metadata; descriptions of the statistical data, user-specified parameters, auxiliary data, and so on. As output, the process step may produce transformed statistical data (if a treatment function is included), but output metadata as well; information about the process.

15. To link the various process steps together, users must also define a **process flow**, describing the process steps and the sequence in which they are performed. A process flow can be trivial (i.e., steps performed in sequence without any intermediate intervention) or complex, with multiple branches or alternative steps. The conditional logic that controls these decisions is referred to as a **process control**.

III. Banff Modules

16. The current set of nine Banff procedures perform various tasks related to statistical data editing, including procedures performing the review, selection, and treatment functions described in the previous section. This collection of procedures, initially developed within GEIS, was envisioned as a data editing "one-stop shop" for continuous numerical data constrained by linear edits, typical for business surveys.

17. Nonetheless, certain SDE functionality is not covered by the Banff procedures, such as the treatment of categorical variables, simple conditional editing based on IF THEN rules, and multivariate outlier detection. Other software packages offer similar tools, such as error localization and donor imputation,

using alternative approaches that may appeal to Banff users. Another Statistics Canada product, CANCEIS (the Canadian Census Edit and Imputation System), offers alternative approaches to error localization and donor imputation, amongst other features; in particular, it is designed to treat categorical variables, a repeatedly cited need for Banff users. A number of free, open-source data editing tools used in the production of official statistics are listed here: Awesome official statistics software[2]; they include software packages developed and maintained by other NSOs. Finally, the proliferation of new and advanced imputation tools based on machine learning (ML) and/or artificial intelligence (AI) has generated immense interest in the SDE community; an in-depth discussion and comparison of some such methods, demonstrating their potential in relation to existing methods, is given by Dagdoud, Goga, and Haziza (2020).

18. An essential element of Banff's redevelopment is to take advantage of these software alternatives, working alongside the existing Banff procedures. To do this, the Banff processor will allow users to call a combination of Banff procedures and external tools in sequence. At a minimum, this will likely require some wrapping or modification of the external tools to accommodate the appropriate file formats and structure, for which the Banff team will provide support and tools wherever possible.

19. However, the Banff team envisions not just a framework where tools from different developers can run in sequence, but where they effectively "speak" to one another. To do this, the Banff team has developed a criterion to standardize the Banff modules, with two purposes in mind:

(a) Modules run in sequence should not require any intermediate data management on the user's part.
(b) Modules performing similar functions should be interchangeable, allowing users to swap methods within a process flow without needing to modify other steps.

20. As an example, consider a process flow originally designed with two process steps: error localization and imputation, respectively using Banff's error localization and donor imputation procedures. Suppose a user wishes to test an alternative error localization package, developed externally. Suitably modified to meet the Banff module standard, this user would simply swap out the old error localization package for the new one in the Banff processor, without needing to touch any inputs or outputs. To do this, the selection flags generated by both error localization procedures need to be in the same format. Similarly, swapping out the donor imputation procedure for an alterative imputation package requires that the alternative appropriately reads in the selection flags.

21. This modular approach relies on two datasets that act as input and outputs for each module: the statistical data that is the target of the overall SDE process, and the Banff status file containing relevant metadata, used both for subsequent process steps and as an audit trail. One of the redevelopment tasks will be to determine the data format and structure of these files; while the data format has not been finalized (the Banff team will consult with clients before making any final decisions), the structure of each will resemble the current Banff structure in most respects.

A. Statistical Data

22. We begin by making some assumptions about the statistical data: that it consists of records, each with a unique record identifier, and that the elements of a record are associated with a specified field (or variable). Each field must also have a unique label. Then every value in the statistical dataset can be identified by both the record and field identifiers, which we label Record ID and Field ID in the Banff system.

23. Any Banff module performing treatment functions must include the option to produce an updated version of the input statistical data to be used in subsequent process steps. In most cases, this should consist of a copy of the input data with resulting modifications applied. Because it is possible for a

[2] https://github.com/SNStatComp/awesome-official-statistics-software

process step to add and/or remove records or introduce new fields (for example, a process step calculating a derived variable), there are no restrictions on the overall size of the output statistical data relative to the input statistical data, other than the fact that the names of existing fields should not change, and individual records should maintain the same Record ID throughout the SDE process.

B. Status File

24. Alongside statistical data, the Banff modules must also produce standardized metadata outputs, for reasons outlined earlier in this section. Within the Banff system, this metadata is coded as status flags and stored in a Banff status file. During the SDE process, Banff status flags are typically altered and updated after each step, while a copy of all flags generated throughout the overall process are saved in a separate file for the purposes of auditing and diagnostics.

25. Existing Banff users will already be familiar with status flags such as FTI (field to impute), FTE (field to exclude) as well as various imputation flags (IDE – deterministic imputation, IDN – donor imputation, etc.) associated with specified values in the statistical data. The redeveloped version of Banff will increase the scope of metadata that can be coded into the status file by introducing status flags at three different levels: the process level (identified by a user-specified Process ID), the record level (identified by Process ID and Record ID) and the field level (Process ID, Record ID, and Field ID). Examples of the type of metadata that could be encoded in the status flags, and how they might be used in subsequent process steps, are given in the table below.

Table 1: Potential uses for Banff status flags

Flag level	Reason for flag (examples)	Potential uses in subsequent steps
Process level	Process failure	Trigger manual intervention, or a back-up automated process
Record level	Exclusion flag	Record excluded from treatment steps such as donor or model-based imputation
	Records that fail a high number of edits	Manual review Mass imputation Excluded from donor imputation
	Suspicious records	Manual review Excluded from donor imputation
	Records that fail a process step, such as error localization or pro-rating	Sent to another stage of automated or manual treatment
	Imputation flag (indicating treatment is complete)	Excluded from subsequent treatment steps
Field level	Missing value	Manual review Automated imputation
	Field to impute: non-missing value selected by error localization for imputation	Manual review Automated imputation
	Field to exclude: non-missing value identified as an outlier	Excluded from donor imputation Excluded from parameter calculations in model-based imputation steps
	Imputation flag (indicating treatment is complete)	Excluded from subsequent treatment steps

26. We note that the production of imputation flags is important not just for auditing and record-keeping purposes, but can also be used by subsequent procedures: for example, Banff's Prorate procedure allows users to specify whether previously imputed records should be included in the pro-rating process or not.

27. As part of a process flow, all Banff modules must also include an option to produce an updated version of the current status file, regardless of functions performed. For modules performing review and selection, it should include selection flags identifying which records and/or fields require further treatment. For modules performing treatment, it should include flags for any records and/or fields that have been imputed. Other output flags will depend on the nature of the module and needs of the user. We note that some modules might not read in status flags; the input status file must nonetheless be specified so that it may be updated before the next process step.

28. It is possible, in certain circumstances, that a certain process step does not generate new status flags; for example, if an outlier detection module detects no outliers, or a donor imputation module fails to find any suitable donors. In this case, the status file would remain unchanged. However, a process step that would *never* produce new status flags, regardless of inputs, would not be considered a Banff module under the current criteria.

29. The status flags generated by each module can either be predetermined, or user-specified, as long as they meet the formatting requirements. While custom flags can be useful in certain circumstances, the following status flags will be reserved in the Banff system for specific generic purposes, and are recommended as the default in most cases where they apply:

(a) FTI: Field to impute (field level)
(b) FTE: Field to exclude (field level)
(c) RTI: Record to impute (record level)
(d) RTE: Record to exclude (record level)
(e) IMP: Imputation flag (field or record level)
(f) FAIL: Process failed (process level)

C. Limitations and Exclusions

30. The target of this standardization project is on methods and tools that alter the data, or interact in some way with other process steps, such as review and selection tools. Process steps that run independently within the process flow, which neither alter the data nor influence later steps, do not require standardization. Examples would include the current Banff procedures VerifyEdits (a design tool) and EditStats (a review tool) which produce meaningful diagnostics but do not participate interactively in an automated process.

31. There are no plans at this point to standardize the outputs of review-only methods. (Examples would include outlier methods that produce outlier scores, but do not perform selection.) There are two reasons for this choice. First, the metadata output of review functions is much more diverse in nature than selection or treatment tools, the latter two typically producing binary selection or imputation flags. Second, review and selection functions are often packaged in a single tool, where users specify both the review method to be used and associated selection thresholds.

32. The goal of the Banff modules structure is to communicate as much metadata information as possible through the Banff status flags, both for the purposes of linking modules together and as an audit trail. It is not expected to record *all* relevant metadata. Nothing will restrict modules from producing additional metadata outputs, such as donor-recipient maps (for donor imputation) or estimated model parameters (for model-based imputation).

33. In some cases, the link between two process steps may require intermediate metadata that does not fit into the Banff status flag framework. Consider a process step that identifies suspicious aggregate domains, whose contributing records are then sent through a custom imputation process. In this case, the intermediate metadata consists of flagged domains, which on the surface would not appear to fit into the Banff status file framework. In this scenario, adding a step that converts the domain-level flags into record-level flags would effectively convert the review step into a Banff module, but would not be required by the system.

IV. Banff Processor

34. The Banff Processor will undergo a complete redesign. To use the current processor, users specify individual process step parameters and overall process flow sequencing in a Microsoft Excel file, which is then converted into XML data. To run the process flow, users call the Banff Processor in SAS, specifying the XML data and input statistical data. The Banff Processor generates SAS code from the XML data, calling the built-in Banff procedures, as well as optional user-supplied custom procedures, as required.

35. The new Banff Processor will no longer rely on SAS, or any other proprietary software. While the design has not been finalized, plans are to improve accessibility and user-friendliness through a point-and-click interface, likely as a web application. This interface will allow users to run any of the included Banff modules, alongside user-selected programs that have been modified for Banff. Modules can be run individually, or as part of a larger process flow.

36. The changes to the Banff modules described in the previous section remove the need for intermediate data management in a linear process flow, making the intermediate steps trivial. Instead of having to manage all the intermediate steps associated with a simple linear process flow, the Banff Processor will instead focus on new features of process flow design, convenience and efficiency.

A. Process blocks

37. The first new feature is one of convenience, to help users organize a complex SDE process flow into separate blocks, comprised of one or more process steps. This allows users to specify a large-scale process flow with main components, and then break those components up into individual process steps. Users will also be able to create nested blocks, to organize the process steps into multiple levels. This approach facilitates the structural organization of SDE process flows, and when combined with a point-and-click interface and process flow visualizations, should vastly simplify and improve the design process.

38. Additionally, any tools of the Banff Processor that can be applied to an individual process step can also be applied to a block as a whole, affecting all elements of the block. This would include process controls (described in the next section), but also generic process step specifications, such as edit groups or imputation classes. Specifying these features at the block level eliminates duplication, further increasing convenience and reducing the risk of specification errors.

B. High-level process controls

39. The GSDEM described process controls as follows: *"The navigation between process steps is managed by rule-based process controls. A process control is called trivial when a process step is followed by the same process step under all circumstances and non-trivial when a step can be followed by several alternative steps, depending on the conditions of the process control."* Example given include specific treatments for influential units, suspicious aggregates, or different imputation approaches based on variable type. These process controls act as filters to determine which data undergoes certain process steps, and are based on conditional logic. The criteria for the process controls are often data-driven, and

may even depend on the state of the statistical data at the specific point when the process control is exercised.

40. To a certain degree, the Banff status flags act as process controls, as they can be used to determine which records or fields undergo subsequent treatment. Within the Banff Processor, we focus instead on high-level process controls, determining when a process step (or block) should be executed, and which data (including metadata) should be processed. Some common categories of high-level process controls are given below:

(a) Filter by record: For each process step, users will be able to specify which records are processed, and which are excluded. Excluded records would be re-integrated with the statistical dataset before any subsequent process steps. Some examples:
- Only send companies meeting specific criteria (size, geography, etc.) through a specialized block of process steps.
- Select records with unexpected responses (e.g., negative sales) for manual review.

(b) Filter by variable: Similar to the record filter, users could limit process steps to only a subset of variables. Some examples:
- Categorical and numerical variables are sent to different process blocks.
- A single process block is applied to different sets of variables.

(c) Filter by status flag: Users could also determine which status flags enter a process step. Some examples:
- Users could exclude all previous status flags, essentially "starting fresh" in a process block.
- Through the use of status flags generated in previous steps, such as record-level flags for influential units or members of a suspicious domain.
- By calling certain process steps only called if a previous process step failed, indicated by a process-level status flag.

Combined with the process blocks, users would be able to apply a process control to large sections of the process flow. These process controls would be visible and editable at the highest levels.

C. Parallel Processing

41. To improve efficiency, the Banff Processor will also introduce the concept of parallel process steps. Two sequential process steps can be run in parallel if the order in which they are run has no impact on the process flow, regardless of input statistical data. In such cases, users would be able to specify sets of two or more parallel process steps (or blocks) in the process flow. When executed, the Banff Processor would then take advantage of any available parallel processing infrastructure. This feature would also apply to data editing classes (i.e., imputation classes, or BY groups in SAS) which are assumed to be independent by nature. (It would be the user's responsibility to ensure that the identified steps fit the described criteria.)

D. Manual Breaks

42. Finally, the Banff Processor would facilitate the integration of manual breaks in the automated process flow. Flagging a process step as "manual" would trigger a pause in the process flow until the outputs of the manual process step have been loaded in the system. Building manual breaks directly into the system not only improves user convenience, but also helps ensure that any human decisions (such as selection or subject matter imputation) are appropriately documented in the process.

V. Conclusion

43. This paper serves as an announcement for Banff's redevelopment, which is at the beginning of its cycle with a planned completion in early to mid-2024. Important changes include a departure from the SAS architecture, standardization of the Banff modules that should simplify process flows and improve inter-module compatibility, and a completely revamped Banff Processor.

44. These changes, while not introducing any new SDE methods or functions, address a number of user-identified issues, both in the individual procedures and the existing Banff processor. The improved user interface will greatly reduce the amount of time spent setting up Banff process flows while reducing the risk of human error. The new module standards will offer users greater control over metadata management and documentation, while the new processor features will improve not just user convenience, but add new functionality and computational efficiency.

45. Upon release, these changes will significantly improve the Banff user experience. But they also lay the foundation for future opportunities in research, innovation, and collaboration, through the integration of external methods and tools into the Banff production environment. The emergence of new and promising SDE tools, in particular based on advanced ML and AI methods, necessitates this shift. A key tenet of the redevelopment is that this integration is as easy and straightforward as possible.

46. Building a catalogue of Banff-compatible SDE modules will require care and diligence, though the rewards should be well worth it. Once the Banff redevelopment is complete, the Banff team hopes to work with internal stakeholders and external partners to develop and expand this catalogue, into possible tiers along the following lines:

(a) **Tier 1:** Banff modules developed and supported within Statistics Canada. These would be available and recommended to all users, with full methodological and technical support. This tier would include all existing Banff procedures, and any future ones developed internally. Possible candidates would include software such as CANCEIS.
(b) **Tier 2:** External programs modified and vetted by the Banff team. These methods would need to be thoroughly vetted, and extensively tested, to achieve integration and support. Only those programs with the highest reputations within the SDE community, with adequate theoretical and technical documentation, would be considered. Though vetted internally, limited support would be offered, with a focus on the integration within the Banff framework. Possible candidates include modified versions of open-source packages such as *missForest*.
(c) **Tier 3:** Community-vetted and maintained programs. This tier would include methods suggested by Banff's user community, appropriately modified to fit the Banff framework. The Banff team would aid in the integration of such modules, but support would rest with the original developers. Possible candidates include programs developed and maintained by other NSOs, such as the R package *Validate*[3].
(d) **Tier 4:** Custom user programs: This tier would consist of external programs, developed or obtained by the user, and modified to fit the Banff module framework. The Banff team would provide tools and guidance for the integration of custom programs, but no additional support; users would be responsible for their own vetting and testing.

47. The full potential for Banff's future rests largely on the level of community buy-in for the third tier of Banff modules, which is why the redevelopment is focused on making the system as accessible, diverse, and flexible as possible. The Banff team looks forward to working closely with the international SDE community in our quest to build a robust, powerful data editing platform with access to the widest and most advanced tools available. We encourage any users or developers interested in collaborating on these developments to reach out to the Banff team at the following email address: statcan.banff-banff.statcan@statcan.gc.ca

[3] https://cran.r-project.org/web/packages/validate/index.html

VI. Acknowledgements

48. The author would like to thank the following people to their contribution to this project: members of the Banff team Steve Matthews, Joël Bissonnette, Danny Finch, and Jonathan Baillargeon, as well as our IT colleagues Marie-Claude Mathieu, Lihsin Hwang and Stephen Arsenault. Thanks as well to previous team members whose work went into this development project, Banff users for incredibly helpful and informative feedback, and Statistics Canada management for supporting this ambitious project. Finally, special thanks to reviewers Étienne Rassart, Mark Stinner, and Matei Mireuta for their time and thoughtful comments.

VII. References

Dagdoug, M., Goga, C., & Haziza, D. (2020). Imputation procedures in surveys using nonparametric and machine learning methods: an empirical comparison. *arXiv preprint arXiv:2007.06298*.

Gray, D. (2018). The evolution of BANFF in the context of modernization. UNECE Workshop on Statistical Data Editing.

Gray, D. and Matthews, S. (2020). Rethinking the role of Generalized Systems for economic statistics. Internal Statistics Canada document presented at the Advisory Committee on Statistical Methods meeting held June 4, 2020.

Kovar, J.G., MacMillan, J. and Whitridge, P. (1988). Overview and strategy for the Generalized Edit and Imputation System. (Updated February 1991). Statistics Canada, Methodology Branch Working Paper No. BSMD-88-007E/F.

Kozak, R. (2005). The Banff System for Automated Editing and Imputation, Proceeding of the SSC Annual Meeting, Survey Methods Section.

Statistics Canada (2017a). Functional Description of the Banff System for Edit and Imputation, Technical Report.

Statistics Canada (2017b). Banff 2.07 Processor User Guide, Technical Report.

Thomas, S. (2017). Future Development of Statistics Canada's Edit and Imputation System Banff. UNECE Work Session on Statistical Data Editing.

United Nations Economic Commission for Europe (UNECE) (2019). *Generic Statistical Data Editing Model (GSDEM)*. [https://statswiki.unece.org/display/sde/GSDEM].