

**QUYẾT ĐỊNH**

**Ban hành Quy trình thu thập, khai thác, sử dụng  
và lưu trữ dữ liệu hành chính**

**CỤC TRƯỞNG CỤC THU THẬP DỮ LIỆU  
VÀ ỨNG DỤNG CÔNG NGHỆ THÔNG TIN THỐNG KÊ**

Căn cứ Luật Thống kê ngày 23 tháng 11 năm 2015; Luật sửa đổi, bổ sung một số điều và Phụ lục Danh mục chỉ tiêu thống kê quốc gia của Luật Thống kê ngày 12 tháng 11 năm 2021;

Căn cứ Quyết định số 1083/QĐ-BKHĐT ngày 14 tháng 6 năm 2023 của Bộ trưởng Bộ Kế hoạch và Đầu tư ban hành Quy trình sản xuất thông tin thống kê chung;

Căn cứ Quyết định số 1001/QĐ-TCTK ngày 20 tháng 5 năm 2020 của Tổng cục trưởng Tổng cục Thống kê quy định chức năng, nhiệm vụ, quyền hạn và cơ cấu tổ chức của Cục Thu thập dữ liệu và Ứng dụng công nghệ thông tin thống kê;

Xét đề nghị của Trưởng phòng Phòng Thu thập và Quản lý dữ liệu hành chính.

**QUYẾT ĐỊNH:**

**Điều 1.** Ban hành kèm theo Quyết định này Quy trình thu thập, khai thác, sử dụng và lưu trữ dữ liệu hành chính gồm 8 bước: (1) Xác định nhu cầu sử dụng dữ liệu hành chính; (2) Chuẩn bị thu thập dữ liệu hành chính; (3) Thu thập dữ liệu hành chính; (4) Chuyển đổi dữ liệu hành chính; (5) Kiểm tra, tích hợp và hoàn thiện dữ liệu hành chính; (6) Khai thác, sử dụng dữ liệu hành chính; (7) Lưu trữ dữ liệu hành chính; (8) Quản lý khai thác và bảo mật dữ liệu hành chính.

**Điều 2.** Quyết định này có hiệu lực thi hành kể từ ngày ký.

**Điều 3.** Lãnh đạo, công chức, viên chức của các đơn vị thuộc Cục Thu thập dữ liệu và Ứng dụng công nghệ thông tin thống kê và các cá nhân, đơn vị liên quan trong quá trình thực hiện thu thập, khai thác, sử dụng và lưu trữ dữ liệu hành chính chịu trách nhiệm thi hành Quyết định này.

**Nơi nhận:**

- Như Điều 3;
- Lãnh đạo Tổng cục Thống kê (để báo cáo);
- Lãnh đạo Cục TTDL;
- Các đơn vị thuộc Cục TTDL;
- Lưu: VT, TTQL (5b).



## QUY TRÌNH

**Thu thập, khai thác, sử dụng và lưu trữ dữ liệu hành chính**  
(Ban hành kèm theo Quyết định số /QĐ-TTDL ngày tháng năm 2024 của  
Cục trưởng Cục Thu thập dữ liệu và Ứng dụng công nghệ thông tin thống kê)

### I. MỘT SỐ VẤN ĐỀ CHUNG

#### 1. Phạm vi và đối tượng điều chỉnh

##### 1.1. Phạm vi áp dụng

Quy trình này được áp dụng đối với dữ liệu hành chính.

Cá nhân, đơn vị có liên quan trong quá trình thực hiện thu thập, khai thác, sử dụng và lưu trữ dữ liệu hành chính.

##### 1.2. Đối tượng áp dụng

Lãnh đạo Cục Thu thập dữ liệu và ứng dụng công nghệ thông tin thống kê (Cục TTDL).

Lãnh đạo, công chức, viên chức (viết gọn là công chức) của các đơn vị thuộc Cục TTDL.

#### 2. Mục đích

Dữ liệu hành chính là dữ liệu của cơ quan nhà nước được ghi chép, lưu giữ, cập nhật trong các hồ sơ hành chính dạng giấy hoặc dạng điện tử phục vụ công tác chỉ đạo, điều hành và thực thi nhiệm vụ của các cơ quan nhà nước. Dữ liệu hành chính là một trong các nguồn dữ liệu dùng để tổng hợp số liệu báo cáo thống kê.

Quy trình thu thập, khai thác, sử dụng và lưu trữ dữ liệu hành chính được xây dựng với mục đích sau:

- Thực hiện việc thu thập, khai thác, sử dụng và lưu trữ dữ liệu hành chính theo trình tự thống nhất, rõ ràng với nội dung theo quy định cụ thể;
- Đảm bảo tính thống nhất trong thực hiện thu thập, khai thác, sử dụng và lưu trữ các nguồn dữ liệu hành chính phục vụ công tác thống kê;
- Nâng cao chất lượng, hiệu quả quá trình thu thập, khai thác, sử dụng và lưu trữ dữ liệu hành chính.

## II. QUY TRÌNH THỰC HIỆN

Quy trình thu thập, khai thác, sử dụng và lưu trữ dữ liệu hành chính gồm 8 bước: (1) Xác định nhu cầu sử dụng dữ liệu hành chính; (2) Chuẩn bị thu thập dữ liệu hành chính; (3) Thu thập dữ liệu hành chính; (4) Chuyển đổi dữ liệu hành chính; (5) Kiểm tra, tích hợp và hoàn thiện dữ liệu hành chính; (6) Khai thác, sử dụng dữ liệu hành chính; (7) Lưu trữ dữ liệu hành chính; (8) Quản lý khai thác và bảo mật dữ liệu hành chính.

### 1. Xác định nhu cầu sử dụng dữ liệu hành chính

#### 1.1 Xác định nhu cầu thông tin

Tổng hợp nhu cầu thu thập thông tin của các đơn vị trong Tổng cục Thống kê. Các nội dung tổng hợp chỉ tiêu thống kê, thông tin bao gồm: Tên chỉ tiêu, thông tin cần thu thập để tính chỉ tiêu, phương pháp tính, phân tổ, tần suất thu thập.

#### 1.2 Tìm hiểu dữ liệu hành chính

##### 1.2.1. Nguồn dữ liệu hành chính

Tìm hiểu, xác định các nguồn dữ liệu hành chính hiện có chứa các thông tin cần thu thập để tổng hợp chỉ tiêu.

Xác định cơ quan quản lý trực tiếp nguồn dữ liệu hành chính.

##### 1.2.2. Xác định các thông tin hiện có trong bộ dữ liệu

Tìm hiểu các chỉ tiêu trong bộ dữ liệu hành chính, nội dung cụ thể cần phải tìm hiểu rõ đối với dữ liệu hành chính:

- Bộ dữ liệu hành chính có những chỉ tiêu nào.
- Tên của các chỉ tiêu, mã số hoặc ký hiệu của các chỉ tiêu trong bộ dữ liệu.
- Khái niệm, định nghĩa, nội hàm của các chỉ tiêu.
- Phương pháp tính toán chỉ tiêu.
- Phân tổ của các chỉ tiêu: Xác định chỉ tiêu theo các phân tổ hành chính (tỉnh, huyện, xã), phân tổ ngành kinh tế (ngành cấp 5, cấp 4, cấp 3, cấp 2, cấp 1), phân tổ loại hình (loại hình doanh nghiệp, loại hình kinh tế)...

##### 1.2.3. Tần suất dữ liệu

Xác định tần suất dữ liệu đối với mỗi cơ sở dữ liệu hành chính và khả năng đáp ứng nhu cầu thông tin phục vụ biên soạn các chỉ tiêu thống kê quốc gia. Tần suất dữ liệu cần nêu rõ loại thông tin nào thu thập vào thời gian nào, thời kỳ của số liệu và thời gian chia sẻ dữ liệu cho cơ quan thống kê phục vụ biên soạn chỉ tiêu thống kê.

### **1.3 Tham vấn, khẳng định nhu cầu thu thập dữ liệu hành chính**

Tham vấn các đơn vị liên quan để xác nhận và khẳng định nhu cầu thông tin có thể khai thác từ dữ liệu hành chính.

Việc tham vấn, khẳng định nhu cầu thông tin cần xác định rõ:

- Các thông tin thu thập từ nguồn dữ liệu hành chính phải đảm bảo để thực hiện tính toán các chỉ tiêu thống kê theo nhu cầu của đơn vị.

- Nội hàm của các chỉ tiêu đảm bảo đồng nhất với nhu cầu tổng hợp của các đơn vị.

- Tần suất thông tin phù hợp với nhu cầu sử dụng.

## **2. Chuẩn bị thu thập dữ liệu hành chính**

Sau khi tham vấn, khẳng định nhu cầu thu thập dữ liệu hành chính từ các đơn vị, cần tiến hành công tác chuẩn bị để thu thập dữ liệu hành chính. Các nội dung chuẩn bị cụ thể:

### **2.1 Xây dựng hành lang pháp lý để thu thập dữ liệu hành chính**

Xây dựng quy chế phối hợp chia sẻ dữ liệu giữa Tổng cục Thống kê với cơ quan quản lý dữ liệu hành chính, quy chế phải bao gồm các chỉ tiêu chia sẻ cụ thể, phân tổ, tần suất, hình thức chia sẻ.

Thống nhất bằng các văn bản khác như quyết định, công văn, cam kết,... về việc cung cấp dữ liệu hành chính cho Tổng cục Thống kê.

### **2.2 Tìm hiểu định dạng, cấu trúc dữ liệu**

Xác định định dạng dữ liệu đối với mỗi cơ sở dữ liệu hành chính và các yêu cầu về hạ tầng công nghệ thông tin, yêu cầu kỹ thuật của các cơ quan khi chia sẻ dữ liệu để nghiên cứu xây dựng công cụ tiếp nhận dữ liệu.

Xác định cấu trúc dữ liệu thuộc dạng nào: dạng mảng (arrays), danh sách liên kết (linked list), cây (tree)... để đề xuất xây dựng công cụ chuyển đổi định dạng và thiết kế cấu trúc dữ liệu phù hợp phục vụ nhu cầu khai thác tại Tổng cục Thống kê.

## **2.3 Xây dựng hệ thống thu thập dữ liệu hành chính**

### **2.3.1 Hệ thống cầu nối dữ liệu trực tuyến (online)**

a) Xây dựng hệ thống nhận dữ liệu trực tiếp từ cơ quan chia sẻ dữ liệu tới Tổng cục Thống kê thông qua đường truyền mạng đảm bảo 4 yếu tố sau:

- Tăng cường bảo mật:

- + Hệ thống bảo đảm các yêu cầu về bảo mật trong quá trình truyền dữ liệu trực tuyến bao gồm: mã hoá dữ liệu trong quá trình truyền dữ liệu; xác thực 2 (nhiều) lớp để

kiểm soát truy cập; sử dụng giao thức bảo mật (HTTPS, VPN ...) để bảo đảm đường truyền an toàn.

+ Hệ thống có cơ chế phát hiện và ứng phó với tấn công mạng (tấn công DDoS, đánh cắp dữ liệu ...).

- Đồng bộ và theo dõi thời gian thực:

+ Thiết lập hệ thống đồng bộ dữ liệu theo thời gian thực để đảm bảo dữ liệu luôn được cập nhật.

+ Xây dựng công cụ giám sát và nhật ký (log) quá trình truyền dữ liệu nhằm phát hiện lỗi kịp thời.

- Định kỳ đánh giá và tối ưu hóa đường truyền mạng để tránh tắc nghẽn hoặc mất dữ liệu.

- Khả năng mở rộng: thiết kế hệ thống với kiến trúc linh hoạt, dễ dàng nâng cấp khi cần mở rộng để tích hợp thêm nguồn dữ liệu mới hoặc khối lượng dữ liệu lớn hơn.

b) Đối với việc thu thập dữ liệu hành chính thường xuyên, giai đoạn này thường thực hiện ở lần đầu tiên, các lần lặp lại tiếp theo chỉ rà soát, cập nhật, nâng cấp và mở rộng khi cần thiết.

### **2.3.2 Truyền dữ liệu ngoại tuyến (offline)**

Thực hiện tiếp nhận dữ liệu hành chính để sao chép dữ liệu của cơ quan chia sẻ dữ liệu bằng các thiết bị như: USB, ổ cứng di động, đĩa CD,... Việc truyền dữ liệu ngoại tuyến áp dụng đối với các loại dữ liệu khi chưa xây dựng được cầu nối dữ liệu trực tuyến (online).

Truyền dữ liệu ngoại tuyến (offline) bảo đảm các yếu tố sau:

- Quy định về thiết bị lưu trữ: bảo đảm các thiết bị sử dụng (USB, ổ cứng di động, đĩa CD,...) được mã hóa và bảo mật bằng mật khẩu hoặc phần mềm chuyên dụng.

- Hạn chế rủi ro mất dữ liệu:

+ Bảo đảm lưu trữ dữ liệu sao chép trên ít nhất hai thiết bị khác nhau để tránh rủi ro thiết bị hỏng hoặc thất lạc.

+ Quy định rõ cách bảo vệ thiết bị lưu trữ trong quá trình vận chuyển, bao gồm việc sử dụng hộp chống sốc và môi trường bảo quản an toàn.

## **3. Thu thập dữ liệu hành chính**

### **3.1 Thực hiện thu thập dữ liệu hành chính**

Thực hiện thu thập dữ liệu trực tuyến: Phối hợp với cơ quan chia sẻ dữ liệu hành chính trong thực hiện chia sẻ dữ liệu. Dữ liệu hành chính được chia sẻ trực

tuyển trên các hệ thống công nghệ thông tin do Tổng cục Thống kê và cơ quan chia sẻ dữ liệu hành chính thống nhất thực hiện.

Thực hiện thu thập dữ liệu ngoại tuyến: sử dụng các thiết bị như: USB, ổ cứng di động, đĩa CD... để tiến hành sao chép dữ liệu trực tiếp từ cơ quan cung cấp dữ liệu hành chính và chuyển vào hệ thống lưu trữ dữ liệu hành chính của Tổng cục Thống kê.

### **3.2 Đối soát dữ liệu hành chính được truyền nhận**

Kiểm soát số lượng bản ghi thu thập. Đối soát dữ liệu hành chính được truyền nhận bao gồm việc:

- So sánh và đối chiếu số lượng gói tin được cơ quan chia sẻ dữ liệu truyền đi thành công với số lượng gói tin Tổng cục Thống kê nhận được.

- Kiểm tra tính đầy đủ về số lượng các bản ghi đã nhận được: căn cứ vào số lượng bản ghi trong danh sách nền của các cuộc điều tra và các nguồn thông tin khác để đối chiếu với số lượng bản ghi thu thập từ nguồn dữ liệu hành chính.

- Trao đổi kỹ thuật với cơ quan chia sẻ để tìm hiểu nguyên nhân thiếu gói tin, thiếu bản ghi (nếu có) và khắc phục tình trạng thiếu gói tin, thiếu bản ghi dữ liệu.

### **3.3 Chuyển đổi cấu trúc dữ liệu**

Đối với dữ liệu được thu thập từ hệ thống cầu nối dữ liệu trực tuyến: Hệ thống thu thập dữ liệu tích hợp chức năng chuyển đổi cấu trúc dữ liệu từ cơ quan chia sẻ về cấu trúc dữ liệu lưu trữ của Tổng cục Thống kê.

Đối với dữ liệu được thu thập bằng hình thức ngoại tuyến: thực hiện xây dựng công cụ lưu trữ và chuyển đổi dữ liệu thu thập sang cấu trúc dữ liệu của Tổng cục Thống kê.

## **4. Chuyển đổi dữ liệu hành chính**

Xây dựng thuật toán ánh xạ, chuẩn hóa các danh mục đang sử dụng trong dữ liệu hành chính để có thể kết nối với danh mục của Tổng cục Thống kê: danh mục hành chính, danh mục ngành, danh mục loại hình kinh tế...

Các danh mục chuẩn hóa này được tích hợp vào phần mềm thu thập dữ liệu hành chính để phục vụ việc chuyển đổi dữ liệu hành chính một cách tự động.

Thực hiện chuyển đổi dữ liệu hành chính sang dữ liệu thống kê (dữ liệu lưu trữ theo các trường, các loại mã theo quy định của Tổng cục Thống kê) trên hệ thống hoặc bằng các phần mềm phụ trợ.

## **5. Kiểm tra, tích hợp và hoàn thiện dữ liệu hành chính**

### **5.1 Kiểm tra dữ liệu thu thập ban đầu**

#### **5.1.1 Kiểm tra và loại bỏ dữ liệu trùng lặp**

Dữ liệu thu thập về có thể có nhiều bản ghi của cùng một đối tượng do các cơ quan lưu lại lịch sử các lần cập nhật, thay đổi thông tin của đối tượng, cần kiểm tra sự trùng lặp của các bản ghi để loại bỏ các bản ghi cũ.

Xây dựng thuật toán để loại các bản ghi trùng lặp.

Thực hiện tin học hóa các thuật toán để loại bản ghi trùng để tích hợp vào hệ thống phần mềm thu thập, xử lý dữ liệu hành chính.

#### **5.1.2 Kiểm tra tính đầy đủ của các trường dữ liệu**

Kiểm tra số lượng trường dữ liệu và số lượng chỉ tiêu thu thập so với các nguồn dữ liệu khác để xác định mức độ đầy đủ của dữ liệu và chỉ tiêu thu thập.

### **5.2 Tích hợp dữ liệu hành chính**

Quá trình tích hợp dữ liệu hành chính với dữ liệu lịch sử và các nguồn dữ liệu từ điều tra đòi hỏi sự chặt chẽ trong việc xử lý, liên kết và phân tích để đảm bảo tính chính xác và ý nghĩa của thông tin tích hợp. Các nội dung cần thực hiện tích hợp dữ liệu hành chính:

- Chuẩn hóa dữ liệu, đảm bảo cùng 01 loại dữ liệu từ các nguồn khác nhau có cùng định dạng và đơn vị tính.
- Xác định các trường dữ liệu cần tích hợp.
- Xây dựng thuật toán để tích hợp dữ liệu hành chính với dữ liệu lịch sử và dữ liệu các cuộc điều tra. Thuật toán phải đảm bảo:
  - + Xác định trường dữ liệu nào dùng để liên kết giữa các nguồn dữ liệu.
  - + Xử lý dữ liệu trùng lặp và thiếu hụt giữa các nguồn dữ liệu.
- Xây dựng công cụ phần mềm hoặc ngôn ngữ lập trình để kết hợp dữ liệu từ các nguồn khác nhau dựa trên trường dữ liệu chung.

### **5.3 Hoàn thiện dữ liệu hành chính**

Hoàn thiện dữ liệu hành chính là quá trình kiểm tra, làm sạch, hiệu chỉnh dữ liệu hành chính để giảm thiểu lỗi và nâng cao chất lượng của dữ liệu, phục vụ phân tích và sử dụng dữ liệu một cách hiệu quả hơn.

#### **5.3.1 Kiểm tra logic giữa các biến**

- Căn cứ vào các chỉ tiêu thu thập được từ nguồn dữ liệu hành chính thực hiện xây dựng quy luật và thuật toán logic phục vụ rà soát, làm sạch, hiệu chỉnh những

dữ liệu chưa chính xác hoặc cần kiểm tra lại. Kiểm tra logic giữa các biến là một bước quan trọng để đảm bảo tính chính xác và đáng tin cậy của dữ liệu thống kê. Quy trình kiểm tra logic phụ thuộc vào mục tiêu của việc kiểm tra. Tuy nhiên cần xác định chính xác các biến cần kiểm tra và đảm bảo rằng chúng có ý nghĩa và mối tương quan với nhau trong nguồn dữ liệu. Có nhiều cách tiến hành kiểm tra, tùy thuộc vào loại dữ liệu đang làm việc (dữ liệu định lượng hoặc định tính) và mục tiêu cụ thể.

- Xây dựng các công cụ phần mềm hoặc ngôn ngữ lập trình để kiểm tra logic giữa các biến theo thuật toán.

- Thực hiện kiểm tra các logic giữa các biến đối với dữ liệu hành chính đã thu thập để xử lý các lỗi logic.

### ***5.3.2 Kiểm tra logic theo không gian và thời gian***

Xây dựng quy luật và thuật toán logic để kiểm tra dữ liệu theo không gian và thời gian:

- + Kiểm tra logic theo không gian: Xác định giới hạn của các biến thống kê để đảm bảo chúng phản ánh đúng không gian nghiên cứu. Tiến hành so sánh cùng chỉ tiêu với các không gian khác nhau, ví dụ: so sánh chỉ tiêu giữa các đơn vị hành chính, giữa các ngành với nhau...

- + Kiểm tra logic theo thời gian: Kiểm tra dữ liệu các biến có tính liên tục qua thời gian, so sánh theo một khoảng thời gian nhất định để phát hiện những biến động bất thường, đặc thù.

- + Kiểm tra logic đan xen giữa không gian và thời gian để xác định mức độ hợp lý giữa không gian và thời gian.

- Xây dựng các công cụ phần mềm hoặc ngôn ngữ lập trình để kiểm tra logic giữa các biến theo thuật toán để phục vụ việc kiểm tra dữ liệu theo không gian và thời gian.

- Thực hiện kiểm tra dữ liệu theo không gian và theo thời gian để phát hiện những bất thường của số liệu để làm sạch.

### ***5.3.3 Kiểm tra đối chiếu giữa các nguồn dữ liệu***

- Xây dựng quy luật và thuật toán kiểm tra đối chiếu dữ liệu giữa các nguồn (hành chính; điều tra; Bộ, ngành...)

- Xây dựng các công cụ phần mềm hoặc ngôn ngữ lập trình để kiểm tra kiểm tra đối chiếu giữa các nguồn dữ liệu theo thuật toán.



- Thực hiện kiểm tra đối chiếu các biến trong dữ liệu đã được tích hợp với các nguồn dữ liệu (hành chính; điều tra; Bộ, ngành...) bằng công cụ sẵn có trên phần mềm, hoặc tự kiểm tra bằng công cụ khác. Đây là một quá trình quan trọng để đảm bảo tính chính xác và nhất quán của thông tin, các bước thực hiện:

+ Xác định các chỉ tiêu, tiêu chí cần đối chiếu giữa các nguồn dữ liệu. Chỉ tiêu có thể là các chỉ số, thông tin chung, hay bất kỳ chỉ tiêu nào khác.

+ Sử dụng các phần mềm hoặc công cụ thống kê để tiến hành kiểm tra và so sánh dữ liệu. Đối với những khác biệt phát sinh, xác định nguyên nhân và thực hiện xử lý, hiệu chỉnh.

+ So sánh tính đầy đủ, tính chính xác của chỉ tiêu điều tra và các nguồn dữ liệu khác. Ví dụ: so sánh chỉ tiêu doanh thu trong báo cáo tài chính từ Tổng cục Thuế và chỉ tiêu doanh thu trong phiếu điều tra; so sánh để xác định những đơn vị có trong nguồn dữ liệu hành chính nhưng không có trong điều tra hoặc ngược lại...

#### ***5.3.4 Trao đổi, thống nhất hiệu chỉnh dữ liệu với cơ quan chia sẻ dữ liệu***

Thực hiện trao đổi với cơ quan chủ quản dữ liệu hành chính để thống nhất sửa và hoàn thiện dữ liệu hành chính theo các nội dung đã kiểm tra tại mục 5.3.1, 5.3.2 và 5.3.3.

Trong trường hợp không đủ thông tin để xác nhận và hoàn thiện số liệu cần thống nhất với cơ quan chủ quản dữ liệu hành chính phương pháp và cách thức xử lý dữ liệu để loại bỏ những sai sót trước khi kết nối với các nguồn dữ liệu điều tra để tổng hợp và sử dụng dữ liệu.

#### ***5.3.5 Tạo biến trung gian phục vụ tính toán, hiệu chỉnh dữ liệu***

Xây dựng thuật toán tạo biến trung gian để phục vụ tính toán, hiệu chỉnh dữ liệu, cụ thể:

- Biến trung gian giúp chuẩn hóa dữ liệu một cách dễ dàng hơn bằng cách tách các phần của logic xử lý ra thành các bước nhỏ hơn và đơn giản hóa quy trình xử lý dữ liệu.

- Sử dụng các biến trung gian để kiểm tra dữ liệu, xác định các yếu tố bất thường của dữ liệu và phục vụ tổng hợp dữ liệu.

- Sử dụng biến trung gian tính toán các giá trị tạm thời trong quá trình xử lý dữ liệu.

#### ***5.3.6 Hiệu chỉnh và gán dữ liệu***

Xây dựng thuật toán hiệu chỉnh và gán dữ liệu để phục vụ tính toán, hiệu chỉnh dữ liệu, nội dung cần thực hiện:

- Chuẩn hóa logic dữ liệu.
- Hiệu chỉnh dữ liệu sai đơn vị tính.
- Thực hiện điều chỉnh hoặc sửa đổi dữ liệu sau khi xác minh để loại bỏ hoặc giảm thiểu các ảnh hưởng của các yếu tố không mong muốn có thể ảnh hưởng đến tính chính xác của dữ liệu.
- Loại các quan sát nằm ngoài khoảng do biến động bất thường, ví dụ: giá trị ngoại lai, đột biến làm ảnh hưởng đến dữ liệu.
- Gán dữ liệu trống (nếu cần): thực hiện gán giá trị cho các ô dữ liệu thiếu hoặc thiếu sót trong bộ dữ liệu tuân theo phương pháp luận thống kê. Khi có dữ liệu bị thiếu, gán dữ liệu có thể được thực hiện để giúp hoàn thiện thông tin và tạo cho bộ dữ liệu hoàn chỉnh mà không ảnh hưởng đến dữ liệu gốc.

### **5.3.7 Xây dựng từ điển cho các tập dữ liệu vi mô (codebook)**

Xây dựng từ điển cho các tập dữ liệu vi mô (còn được gọi là codebook). Đây là quá trình tạo ra một danh sách các bảng mã hóa cho các biến trong bộ dữ liệu. Mục tiêu chính của việc này là giảm kích thước của dữ liệu và tăng hiệu suất trong quá trình xử lý và phân tích dữ liệu.

Các nội dung từ điển của dữ liệu vi mô:

- Mã hóa: Danh sách các biến được sử dụng trong dữ liệu.
- Mô tả: Giải thích ý nghĩa của các biến
- Kiểu dữ liệu: Thông tin về kiểu dữ liệu của từng biến số hoặc giá trị trong dữ liệu.
- Thông tin bổ sung: Thông tin bổ sung như đơn vị tính, hạn chế, hoặc bất kỳ chi tiết quan trọng nào khác về dữ liệu.
- Ghi chú và tài liệu: Các ghi chú chi tiết về cách xây dựng từ điển, một số giả định hay quyết định quan trọng trong quá trình xử lý dữ liệu.

### **5.4 Tạo lập bộ dữ liệu hành chính hoàn chỉnh**

Sau khi thực hiện xong các bước trên, tiến hành tạo ra tập dữ liệu hoàn chỉnh để sử dụng trong các phân tích, báo cáo hoặc ứng dụng khác, ghi chú về các giá trị đã được hoàn thiện và tạo báo cáo về quá trình hoàn thiện dữ liệu.

## **6. Khai thác, sử dụng dữ liệu hành chính**

Dữ liệu hành chính được sử dụng phục vụ cho các mục đích khác nhau:

- Kết hợp với các chỉ tiêu đã thu thập để hình thành bộ dữ liệu phục vụ tổng hợp báo cáo;

- Đối chiếu dữ liệu thu thập với dữ liệu hành chính để phục vụ kiểm tra các chỉ tiêu đã thu thập;

- Sử dụng trực tiếp dữ liệu hành chính để tổng hợp và viết báo cáo phân tích.

### **6.1 Xây dựng sổ tay hướng dẫn sử dụng dữ liệu hành chính vi mô**

Sổ tay hướng dẫn sử dụng dữ liệu hành chính vi mô bao gồm một số nội dung sau:

#### **(1) Giới thiệu**

- Mục đích của dữ liệu vi mô.

- Đối tượng sử dụng dữ liệu.

#### **(2) Nguồn dữ liệu**

- Danh sách các nguồn dữ liệu vi mô.

- Mô tả chi tiết về mỗi nguồn dữ liệu.

- Kỳ số liệu, tính cập nhật của dữ liệu.

- Định nghĩa của các chỉ tiêu (nếu có).

#### **(3) Xây dựng codebook**

#### **(4) Hướng dẫn truy cập dữ liệu**

- Hướng dẫn cách truy cập mỗi nguồn dữ liệu.

- Điều kiện và quyền truy cập.

#### **(5) Giới thiệu về công cụ khai thác dữ liệu**

- Hướng dẫn cách thực hiện các truy vấn cơ bản.

- Mô tả cách lọc, sắp xếp, và tận dụng dữ liệu.

#### **(6) Bảo mật thông tin**

- Quy tắc và biện pháp an ninh đối với dữ liệu.

- Hướng dẫn về quyền truy cập và phân quyền.

#### **(7) Thông tin hỗ trợ và liên hệ**

### **6.2 Xây dựng các công cụ tổng hợp dữ liệu**

Căn cứ vào phương pháp khai thác dữ liệu, phương pháp tổng hợp dữ liệu để xây dựng các công cụ phù hợp.

#### **6.2.1 Xây dựng hệ biểu tổng hợp đầu ra**

Căn cứ biểu mẫu tổng hợp đầu ra của các đơn vị sử dụng dữ liệu hành chính để rà soát hệ biểu tổng hợp đầu ra, rà soát thuật toán tổng hợp và xây dựng hệ biểu đầu ra và thuật toán tổng hợp.

Xây dựng phần mềm tổng hợp các mẫu biểu tổng hợp để tổng hợp dữ liệu hành chính theo các mẫu biểu quy định.

### **6.2.2 Xây dựng công cụ tổng hợp tùy biến**

Căn cứ vào nhu cầu tổng hợp dữ liệu theo các phân tổ khác nhau để đề xuất các chỉ tiêu tổng hợp và các phân tổ cần thiết.

Xây dựng công cụ để thực hiện việc tổng hợp các chỉ tiêu theo phân tổ tùy biến.

### **7. Lưu trữ dữ liệu hành chính**

Xác định các loại dữ liệu, tần suất sao lưu dữ liệu của từng loại dữ liệu hành:

- Loại dữ liệu hành chính:
- + Dữ liệu thu thập ban đầu;
- + Dữ liệu hành chính đã hoàn thiện sau kiểm tra, làm sạch và hiệu chỉnh;
- + Dữ liệu hành chính đã tích hợp với các loại dữ liệu khác;
- Xác định tần suất sao lưu dữ liệu của từng loại, số lượng phiên bản cần sao lưu, thời hạn lưu trữ.

Thực hiện tổ chức lưu trữ dữ liệu theo từng loại và từng phiên bản chức lưu trữ dữ liệu theo từng loại và từng ượng phiên bản cần sao lưu.

### **8. Quản lý khai thác và bảo mật dữ liệu hành chính**

#### **8.1 Quản lý khai thác dữ liệu hành chính**

- Cấp quyền cho các tài khoản khai thác dữ liệu hành chính theo chức năng nhiệm vụ và phân công của đơn vị.
- Quản lý việc sử dụng, khai thác dữ liệu đảm bảo dữ liệu được khai thác, sử dụng đáp ứng mục đích của các đơn vị.
- Quản lý mức độ sử dụng, kiểm soát dữ liệu.
- Quản lý nhật ký truy cập, kiểm soát các tác động của tài khoản người dùng tới dữ liệu.

#### **8.2 Bảo mật dữ liệu hành chính**

Trong quá trình thu thập, khai thác, sử dụng dữ liệu hành chính cần bảo đảm bảo mật thông tin, đặc biệt là các thông tin cá nhân theo quy định của Nghị định 13/2023/NĐ-CP, các biện pháp cần thực hiện bảo mật thông tin tuân thủ theo các quy định tại khoản 2.2 Quy trình này.

Một số biện pháp cần thực hiện bảo mật thông tin:

- Mã hóa dữ liệu và đặt nhiều lớp bảo vệ trong quá trình truyền dữ liệu trên hệ thống thu thập dữ liệu hành chính;
- Có cơ chế phát hiện và ứng phó với tấn công mạng (tấn công DDoS, đánh cắp dữ liệu ...).

- Mã hóa thông tin cá nhân trong quá trình truyền nhận dữ liệu và giải mã khi nhận dữ liệu về hệ thống của Tổng cục Thống kê;

- Mã hóa thông tin cá nhân hoặc loại bỏ thông tin cá nhân trong quá trình khai thác, sử dụng dữ liệu;

- Hệ thống thu thập dữ liệu bảo đảm phân cấp, phân quyền chặt chẽ đối với người dùng và có lưu vết dữ liệu lịch sử để quản lý việc thu thập dữ liệu.

Quy trình này được rà soát và cập nhật khi cần thiết./.