

BÁO CÁO CHUYÊN ĐỀ 267

Quy trình quản lý dữ liệu cho các nghiên cứu xã hội học

Hoàng Tuấn Long
Phạm Thu Thủy



RESEARCH
PROGRAM ON
Forests, Trees and
Agroforestry

Báo cáo chuyên đề 267

Quy trình quản lý dữ liệu cho các nghiên cứu xã hội học

Hoàng Tuấn Long

Phạm Thu Thủy

Báo cáo chuyên đề 267

© 2020 Trung tâm Nghiên cứu Lâm nghiệp Quốc tế (CIFOR)



Nội dung trong ấn phẩm này được cấp quyền bởi Giấy phép bản quyền Ghi nhận công của tác giả - Phi thương mại, không chỉnh sửa, thay đổi hay phát triển - Không phát sinh 4.0. <http://creativecommons.org/licenses/by-nc-nd/4.0/>

DOI: 10.17528/cifor/007888

Hoàng TL và Phạm TT. 2020. *Quy trình quản lý dữ liệu cho các nghiên cứu xã hội học*. Báo cáo chuyên đề 267. Bogor, Indonesia: CIFOR.

CIFOR
Jl. CIFOR, Situ Gede
Bogor Barat 16115
Indonesia

T +62 (251) 8622-622
F +62 (251) 8622-100
E cifor@cgiar.org

cifor.org

Chúng tôi xin cảm ơn các nhà tài trợ đã hỗ trợ cho nghiên cứu này thông qua việc đóng góp vào quỹ của CGIAR. Xin xem danh sách các nhà tài trợ: <http://www.cgiar.org/about-us/our-funders/>

Tất cả các quan điểm thể hiện trong ấn phẩm này là của các tác giả. Chúng không nhất thiết đại diện cho quan điểm của CIFOR, các cơ quan chủ quản của tác giả hay của các nhà tài trợ cho ấn phẩm này.

Mục lục

Lời cảm ơn	v
Tóm tắt	vi
1 Giới thiệu	1
2 Người dùng của hướng dẫn này	2
3 Thế nào là quản lý số liệu? Nguyên tắc chung và chiến lược quản lý số liệu	3
4 Xác định mục tiêu của tổ chức	4
5 Xây dựng quy trình cho việc thu thập, xử lý, làm sạch, lưu trữ, bảo mật, quản lý và chia sẻ dữ liệu	5
5.1 Xác định số liệu cần thu thập	5
5.2 Xác định phương pháp thu thập số liệu	5
5.3 Xây dựng khung nhập liệu	6
5.4 Đào tạo và thống nhất giữa người thu thập số liệu, người nhập số liệu và người thiết lập khung nhập số liệu	9
6 Làm sạch số liệu	13
6.1 Bước 1: Rà soát, kiểm tra từng phiếu	13
6.2 Bước 2: Kiểm tra tổng thể bộ dữ liệu	14
6.3 Bước 3: Kiểm tra ngẫu nhiên	18
7 Đảm bảo có bộ phận chuyên trách về quản lý dữ liệu và ngân sách cho công việc quản lý số liệu	20
8 Đào tạo và nâng cao năng lực cho cán bộ	21
9 Kết luận	22
10 Tài liệu tham khảo	23

Danh sách bảng và hình

Bảng

1	So sánh ưu nhược điểm của thu thập số liệu bằng giấy và trực tuyến	6
2	Một số vấn thường gặp khi thiết kế bảng phỏng vấn và khung nhập liệu	7
3	Quy trình phối hợp giữa người thiết kế và người phỏng vấn	9
4	Một số điều kiện ngoại cảnh có thể ảnh hưởng đến người trả lời và cách phòng tránh sự ảnh hưởng	10
5	Dấu hiệu của việc người phỏng vấn đang bị ảnh hưởng bởi thiên kiến	10
6	Một số lỗi thường gặp phải khi thu thập số liệu	11
7	Một số lỗi thường phát sinh cũng như cách khắc phục khi nhập liệu	12

Hình

1	Năm bước để xây dựng quy trình và chiến lược quản lý số liệu	3
2	Quy trình quản lý số liệu	5
3	Ví dụ về thiết kế phiếu bản khảo sát và khung nhập liệu	7
4	Ví dụ về một số lỗi khi nhập liệu	12
5	Quy trình kiểm tra và làm sạch dữ liệu	13
6	Quy trình kiểm tra chéo giữa các thành viên	14
7	Kiểm tra sự đầy đủ của dữ liệu bằng hàm 'counta' trong Excel	15
8	Kiểm tra sự chính xác của dữ liệu bằng 'countif' trong Excel	15
9	Lưu ý khi sử dụng công cụ Filter trong Excel	16
10	Sử dụng Filter để liệt kê dữ liệu sẵn có	17
11	Sử dụng Filter để lọc những dữ liệu sai và thiếu	17
12	Sử dụng Filter để kiểm tra câu hỏi dạng follow-up	18

Lời cảm ơn

Bài viết này nằm trong một hợp phần của Nghiên cứu so sánh toàn cầu về REDD+ mà CIFOR đang tiến hành (www.cifor.org/gcs). Chúng tôi xin trân trọng cảm ơn hỗ trợ tài chính từ các nhà tài trợ bao gồm Cơ quan Hợp tác Phát triển Na Uy (NORAD) và Chương trình nghiên cứu CGIAR về rừng, cây và nông lâm kết hợp (CRP-FTA) đã hỗ trợ nghiên cứu này.

Chúng tôi xin gửi lời cảm ơn chân thành tới Maria Brockhaus, Grace Wong, Moira Moeliono đã cung cấp các hướng dẫn cho báo cáo này.

Tóm tắt

Tài liệu này được xây dựng nhằm chia sẻ kinh nghiệm với các nhà nghiên cứu và các nhà quản lý dự án về quy trình lập kế hoạch, thiết kế, thu thập và quản lý số liệu cho các nghiên cứu xã hội học. Tài liệu không chỉ đưa ra các nguyên tắc và các yếu tố cần xem xét mà còn cung cấp các ví dụ thực tiễn để giúp các nhà nghiên cứu và các nhà quản lý dự án tránh khỏi các sai lầm trong quá trình thu thập, quản lý và xử lý số liệu.

1 Giới thiệu

Số liệu là tài sản quý giá trong thế kỉ 21 bởi số liệu sẽ giúp các bên có liên quan bao gồm các nhà hoạch định chính sách, các tổ chức phi chính phủ, các tổ chức nghiên cứu, doanh nghiệp, cộng đồng, cá nhân có căn cứ để ra các quyết định quan trọng (Dallemand 2019). Tuy nhiên, việc thu thập và sử dụng số liệu trong thực tế gặp nhiều khó khăn bởi số liệu thường không được thu thập chính xác và đầy đủ, quản lý không có hệ thống do vậy thường không đảm bảo chất lượng và độ tin cậy. Sử dụng số liệu thiếu chính xác có thể dẫn tới các quyết định sai lầm, trong khi đó nếu số liệu thu thập chính xác và lưu trữ có hệ thống sẽ giúp các bên tăng năng suất công việc, tiết kiệm và giảm các chi phí làm việc, khi mà tổ chức không tốn thời gian và công sức để tìm kiếm những dữ liệu sẵn có, tránh được các rủi ro về bảo mật, và đưa ra các quyết định chính xác và kịp thời hơn (Ainsworth 2020). Chính bởi vậy, xây dựng một quy trình chuẩn từ thu thập, làm sạch, lưu trữ và quản lý số liệu là rất cần thiết.

Hiện nay có rất nhiều tài liệu hướng dẫn về các bước thu thập số liệu nhưng lại có rất ít các tài liệu hướng dẫn việc làm sạch và lưu trữ số liệu hiệu quả. Các tổ chức nghiên cứu, các tổ chức phi chính phủ và các cơ quan nhà nước đã tiến hành nhiều nghiên cứu phục vụ cho quá trình ra quyết định. Tuy nhiên, vấn đề đặt ra với các tổ chức này là: 1) làm thế nào để kiểm soát và đảm bảo chất lượng số liệu thu được? và 2) làm thế nào để kiểm tra số liệu đã hoàn chỉnh, đầy đủ và lưu trữ có hệ thống?

Hướng dẫn này nhằm mục đích trả lời câu hỏi trên dựa trên tổng hợp các kinh nghiệm của tác giả trong quá trình làm việc với dữ liệu tại Tổ chức Nghiên cứu Lâm nghiệp Quốc tế (CIFOR). Tài liệu này dựa trên kinh nghiệm thực tế của tác giả trong quá trình tiến hành thu thập và quản lý một số lượng lớn số liệu lớn được thu thập qua các nghiên cứu so sánh toàn cầu như dự án Nghiên cứu so sánh toàn cầu về REDD+ (<https://www.cifor.org/gcs/>) và dự án ASEAN-Swiss partnership on Social Forestry and Climate Change (<https://www.cifor.org/asfcc/about/>)

Báo cáo hi vọng có thể giúp người đọc xây dựng kĩ năng và quy trình về quản lý dữ liệu cơ bản, đảm bảo tính chính xác và hạn chế tối đa những sai sót phát sinh trong quá trình làm việc với dữ liệu. Hướng dẫn này được chia làm 8 phần:

- Người sử dụng của hướng dẫn này
- Thế nào là quản lý số liệu?
- Xác định mục tiêu của tổ chức
- Xây dựng quy trình cho việc thu thập, xử lý, làm sạch, lưu trữ, bảo mật, quản lý và chia sẻ dữ liệu
- Làm sạch số liệu
- Đảm bảo có bộ phận chuyên trách về quản lý dữ liệu và ngân sách cho công việc quản lý số liệu
- Đào tạo và nâng cao năng lực

2 Người dùng của hướng dẫn này

Hướng dẫn này được viết cho nhóm người dùng dưới đây:

- Nghiên cứu viên thiết kế và thực hiện thu thập số liệu
- Người nhập số liệu
- Người quản lý số liệu
- Người quản lý dự án muốn kiểm tra chất lượng số liệu



Lưu trữ và số hóa các loài cây gia vị của nhà hàng tại Yangambi - DRC. Ảnh chụp bởi Axel Fassio/CIFOR

3 Thế nào là quản lý số liệu? Nguyên tắc chung và chiến lược quản lý số liệu

Quản lý dữ liệu là một quá trình bao gồm các quy trình thu thập, lưu trữ, tổ chức, bảo mật, chia sẻ dữ liệu được tạo ra và thu thập bởi một tổ chức nhằm đảm bảo dữ liệu chính xác, đầy đủ, nhất quán và có thể tiếp cận dễ dàng (Rouse 2019, Tableau 2020). Để quản lý số liệu, một tổ chức cần phải chủ động xây dựng quy trình và chiến lược quản lý số liệu thay vì bị động và giải quyết các vấn đề đột xuất nảy sinh trong một thời gian gấp gáp. Nguyên tắc của việc quản lý tốt số liệu (IGGI 2005) bao gồm: Tránh thu thập số liệu trùng lặp, kiểm soát vòng đời dữ liệu, xây dựng và thực thi chính sách dữ liệu, xác định rõ quyền sở hữu dữ liệu và tập trung vào quản lý và đảm bảo chất lượng dữ liệu

Chiến lược quản lý số liệu: xác định quy trình và lộ trình của tổ chức để thu thập, quản lý, và sử dụng dữ liệu nhằm đạt được mục tiêu cụ thể của tổ chức. Lộ trình này đảm bảo rằng tất cả các hoạt động xoay quanh việc quản lý số liệu – từ thu thập, làm sạch số liệu đến phối hợp – làm việc cùng nhau có hiệu lực và hiệu quả để trở nên hữu dụng nhất có thể và dễ dàng quản lý. Để xây dựng chiến lược quản lý số liệu, tổ chức cần xem xét 5 bước (Hình 1).



Hình 1. Năm bước để xây dựng quy trình và chiến lược quản lý số liệu

Từng bước trong quy trình này sẽ được thảo luận chi tiết trong các phần dưới đây.

4 Xác định mục tiêu của tổ chức

Xác định mục tiêu quản lý số liệu là bước đầu tiên quan trọng mà mỗi tổ chức cần thực hiện bởi cơ cấu, tổ chức và quy trình quản lý số liệu phụ thuộc vào từng mục tiêu cụ thể. Ví dụ, đối với nhiều chương trình dự án của NGOs, mục tiêu chính của họ là xây dựng quy trình và hệ thống có thể kiểm tra chất lượng của tư vấn họ thuê làm nghiên cứu và đảm bảo số liệu này được thu thập đầy đủ, chính xác. Đối với mục tiêu này, các NGOs cần xây dựng các biểu mẫu, nguyên tắc, quy định thu thập số liệu, nhập số liệu chuẩn của mình và yêu cầu các bên tư vấn nộp trả lại sản phẩm theo đúng định dạng yêu cầu. Đối với các tổ chức nghiên cứu, số liệu của các dự án khác nhau có thể được sử dụng để tiến hành các nghiên cứu so sánh do vậy đòi hỏi phải có một quy trình và hệ thống lưu trữ số liệu đồng nhất giữa các chương trình dự án theo cùng một định dạng chuẩn để có thể tận dụng tối đa số liệu sẵn có. Đối với các cơ quan quản lý nhà nước, số liệu thống kê được thu thập hàng năm cần được lưu trữ và quản lý có hệ thống để có thể xây dựng những bộ số liệu dài hạn có tính thống nhất và có giá trị sử dụng lâu dài.



Tham gia hội thảo lấy ý kiến về sản bản bền vững tại Aishalton, Guyana, ngày 8-10 tháng 9 năm 2019.
Ảnh chụp bởi Barbara Fraser/CIFOR

5 Xây dựng quy trình cho việc thu thập, xử lý, làm sạch, lưu trữ, bảo mật, quản lý và chia sẻ dữ liệu

Để đảm bảo sự nhất quán giữa các thành viên của tổ chức trong việc thu thập, xử lý làm sạch, quản lý, lưu trữ, bảo mật và chia sẻ số liệu cần có một quy trình chuẩn của tổ chức để các bên áp dụng. Hình 2 miêu tả các bước cả tổ chức và người quản lý dự án cần xem xét trong quá trình lập kế hoạch, thiết kế, thu thập, và quản lý số liệu.



Hình 2. Quy trình quản lý số liệu

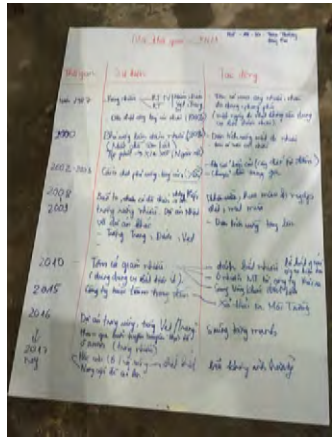
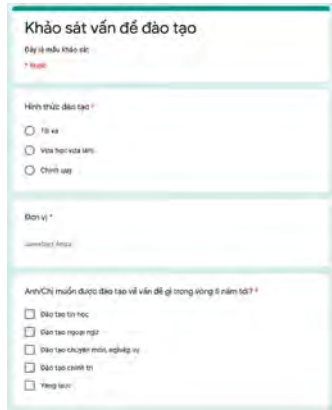
5.1 Xác định số liệu cần thu thập

Tùy vào mục tiêu của tổ chức, số liệu cần thu thập có thể ở nhiều dạng khác nhau như dạng hình ảnh, dạng âm thanh, dạng văn bản. Với mỗi loại số liệu sẽ có các cách thức và quy trình để thu thập, quản lý, lưu trữ và làm sạch khác nhau. Trong tài liệu này, nhóm tác giả chỉ tập trung vào dạng số liệu văn bản thu thập qua quá trình tiến hành khảo sát bằng giấy hoặc các công cụ online.

5.2 Xác định phương pháp thu thập số liệu

Để thu thập số liệu dạng văn bản, có hai phương pháp có thể áp dụng: trực tuyến sử dụng internet và trên giấy. Mỗi phương pháp đều có ưu và nhược điểm riêng (Bảng 1).

Bảng 1. So sánh ưu nhược điểm của thu thập số liệu bằng giấy và trực tuyến

Định dạng bản khảo sát	Ưu điểm	Nhược điểm	Ví dụ
Giấy	<p>Lưu trữ bản gốc chuẩn của thông tin để đối chiếu khi sử dụng sau này</p> <p>Phục vụ ở nhiều tình huống không cần điện hay internet</p> <p>Với các thông lệ quốc tế hiện nay và theo các yêu cầu của dự án, các tài liệu và bản khảo sát giấy phải giữ ít nhất tối thiểu 5 năm</p>	<p>Cần thiết kế thủ công</p> <p>Cần số hoá để lưu trữ và phân tích</p> <p>Chi phí cao để in ấn</p> <p>Cần có không gian để lưu trữ số liệu</p> <p>Cần nhân lực để nhập lại số liệu, làm sạch số liệu để phân tích</p>	
Điện tử online qua điện thoại, máy tính bảng, internet và các phần mềm chuyên dụng khảo sát như googleform, surveymonkey	<p>Giảm công việc thiết kế</p> <p>Tiết kiệm chi phí in ấn</p> <p>Lưu trữ và số hoá thông tin ngay khi thu thập số liệu</p> <p>Không cần không gian để lưu trữ</p> <p>Các ứng dụng thu thập và khảo sát online đều có chức năng phân tích và tổng hợp thông tin do vậy sẽ tiết kiệm được chi phí để thu thập, tổng hợp và phân tích số liệu</p>	<p>Cần có điện, một số trường hợp cần internet</p> <p>Cần chú ý đến quy trình nhập và chỉnh sửa số liệu</p> <p>Số liệu có thể bị chỉnh sửa nếu như bị hack</p> <p>Chỉ phù hợp với những bên khảo sát có kĩ năng sử dụng internet và máy tính thành thạo</p> <p>Do không có sự tương tác trực tiếp nên người trả lời khảo sát có thể trả lời vội vàng nhanh chóng hoặc điền sai</p>	

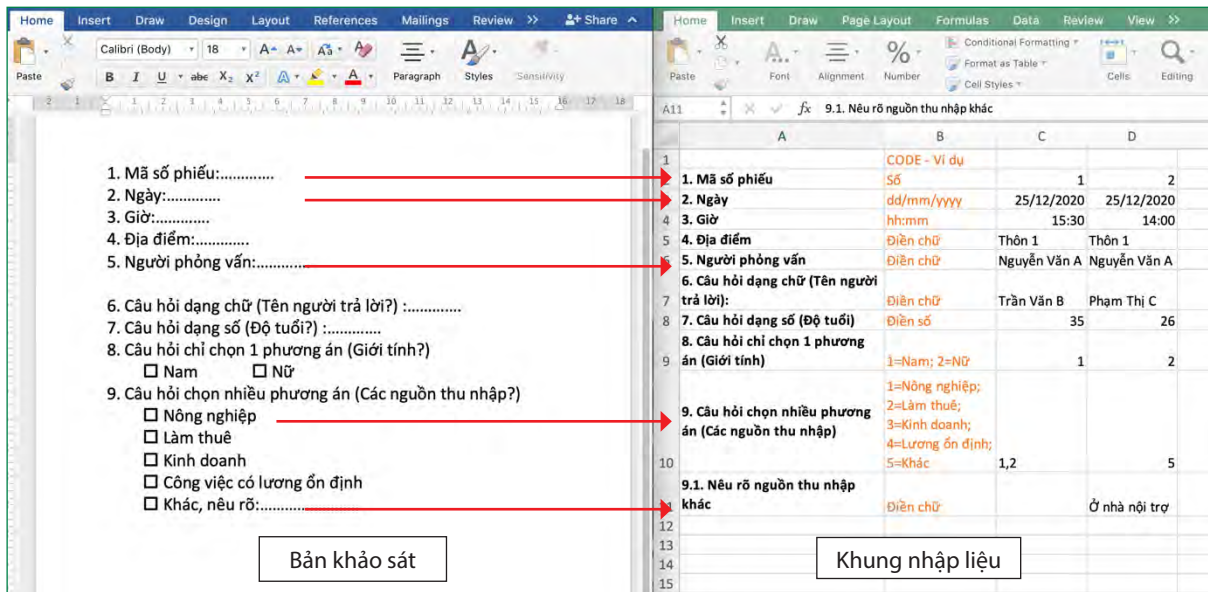
Hiện nay, tuy đã có nhiều công cụ khảo sát và thu thập thông tin online (ví dụ Survey Monkey hay Google Form), phần lớn các nghiên cứu và khảo sát thu thập số liệu đều được tiến hành ở dạng giấy. Dù là thu thập số liệu bằng giấy hay bằng các phần mềm online thì bảng khảo sát là phương tiện quan trọng và cần được thiết kế đầy đủ, dễ dùng để giúp người khảo sát thu thập thông tin. Tùy vào mục đích và đối tượng khảo sát, bản khảo sát có thể được thiết kế với độ dài ngắn khác nhau và dưới hình thức khác nhau. Đối với những nghiên cứu xã hội, thông thường để đảm bảo chất lượng thông tin đầy đủ và chính xác, bản khảo sát bao gồm khảo sát thông tin định tính và định lượng.

5.3 Xây dựng khung nhập liệu

Khung nhập liệu là nơi lưu trữ các thông tin và dữ liệu đã thu thập được cho các mục đích sử dụng sau này. Bản khảo sát và khung nhập liệu cần được thiết kế tương đồng như ví dụ trong Hình 3, trong đó:

- Câu hỏi trong bản khảo sát và khung nhập liệu cần trùng khớp.
- Trong khung nhập liệu cần có đủ vị trí để người nhập liệu có thể nhập đầy đủ dữ liệu từ bản khảo sát.
- Thứ tự các câu hỏi và thứ tự các phương án trả lời nên được thiết kế đồng nhất giữa phiếu hỏi và khung nhập liệu để dễ dàng hơn cho quá trình nhập liệu sau này.
- Khung nhập liệu nên được thiết kế để có thể dễ dàng sử dụng và có quy tắc chung khi nhập liệu như định dạng, cách nhập, mã hoá cho các câu trả lời như một cách hướng dẫn cho người phỏng vấn trong quá trình nhập liệu.

Một số vấn đề thường gặp và giải pháp khắc phục khi thiết kế bảng hỏi và khung nhập số liệu có thể kể đến trong Bảng 2.





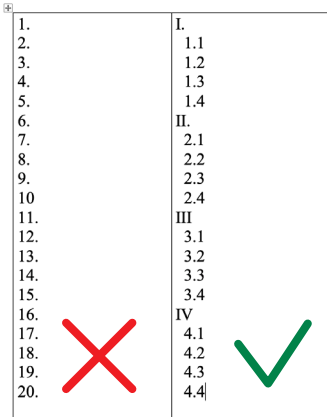
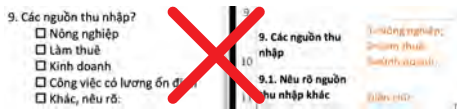
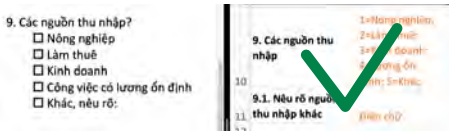
Hình 3. Ví dụ về thiết kế phiếu bản khảo sát và khung nhập liệu

Bảng 2. Một số vấn đề thường gặp khi thiết kế bảng phỏng vấn và khung nhập liệu

Vấn đề thường gặp	Ví dụ cụ thể	Giải pháp
Bố trí bảng hỏi không thuận tiện	Câu hỏi hoặc các phương án trả lời dài bị chia cắt thành 2 trang dẫn đến việc người phỏng vấn vô tình bỏ sót câu hỏi hoặc phương án trả lời	Dàn lại trang, nội dung của phiếu hỏi. Đảm bảo câu hỏi và các phương án trả lời nằm trên cùng một trang

Tiếp tục đến trang tiếp theo

Bảng 2. Tiếp tục

Vấn đề thường gặp	Ví dụ cụ thể	Giải pháp
<p>Bố trí khung nhập liệu không thuận tiện</p>	<p>Thứ tự câu hỏi giữa bản khảo sát và khung nhập liệu khác nhau hoặc thứ tự các phương án trả lời giữa bản hỏi và khung nhập liệu khác nhau</p> 	<p>Sắp xếp thứ tự câu hỏi, phương án trả lời đồng nhất giữa phiếu hỏi và khung nhập liệu</p> 
	<p>Chia phần cho các bảng hỏi dài để có các quãng nghỉ trong cả lúc phỏng vấn cũng như lúc nhập liệu</p>	<p>Thay vì câu hỏi từ 1 đến 50 thì có thể chia ra 5 phần với 10 câu hỏi cho mỗi phần và đánh dấu số lượng câu hỏi theo từng phần</p> 
	<p>Khung nhập liệu cần thân thiện với người nhập liệu, tránh nhiều thao tác</p>	<p>Khi sử dụng Excel làm khung nhập liệu cách nhập liệu theo chiều dọc (như Hình 2) sẽ thuận tiện hơn cho người nhập liệu thay vì nhập liệu theo chiều ngang</p>
<p>Thiếu vị trí nhập liệu</p>	<p>Trong bản hỏi có câu hỏi mà trong khung nhập liệu không có hoặc ngược lại</p> <p>Khi thiết kế bản khảo sát, đặc biệt là trong những nghiên cứu xã hội có những thông tin nền hoặc câu trả lời không có sẵn trong bản khảo sát</p>	<p>Bổ sung câu hỏi hoặc câu trả lời bị thiếu</p> <p>Trong bản khảo sát và khung nhập liệu nên có những khu vực để người phỏng vấn ghi chú những thông tin bổ sung, và trong mỗi câu hỏi cũng nên có phương án "Khác" để ghi chú cho những phương án trả lời chưa được liệt kê sẵn</p>
<p>Thừa hoặc thiếu code</p>	<p>Trong câu trả lời có 4 phương án để chọn, nhưng trong khung nhập liệu chỉ có 3 code để nhập</p> 	<p>Có thể bổ sung ngay từ khi thiết kế. Trong trường hợp khi đi phỏng vấn thực tế có nhiều phương án chưa có trong thiết kế ban đầu, có thể thống nhất với người nhập liệu để có code mới cho phương án đó.</p> 

5.4 Đào tạo và thống nhất giữa người thu thập số liệu, người nhập số liệu và người thiết lập khung nhập số liệu

Thu thập số liệu là công việc thường được thực hiện bởi một nhóm thay vì cá nhân. Vì trong một nhóm là tập hợp của nhiều người khác nhau với nền tảng khác nhau nên trước khi tiến hành thu thập số liệu, việc phối hợp và trao đổi giữa những người phỏng vấn rất quan trọng để giảm thiểu sai sót về mặt số liệu. Để quản lý số liệu tốt, cần có sự phối hợp chặt chẽ giữa người thiết kế nghiên cứu và bảng hỏi, người xây dựng khung nhập số liệu và người tiến hành phỏng vấn (Bảng 3).

Bảng 3. Quy trình phối hợp giữa người thiết kế và người phỏng vấn

Các bước tiến hành	Chi tiết
Bước 1. Thống nhất về khung số liệu giữa người thiết kế nghiên cứu và người thiết kế khung nhập số liệu	<p>Người thiết kế nghiên cứu và câu hỏi khảo sát trao đổi về mục đích và cách hỏi của nghiên cứu với người xây dựng khung nhập liệu</p> <p>Người thiết kế khung nhập liệu dựa trên các trao đổi ban đầu này sẽ thiết kế Bản thảo nhập 1 khung nhập liệu để lưu giữ các thông tin khảo sát theo hướng tiện dụng cho cả người nhập số liệu lẫn người phân tích dữ liệu</p> <p>Cả người thiết kế nghiên cứu và người xây dựng khung số liệu cùng xem xét lại Bản thảo nhập 1 bảng khung nhập liệu xây dựng xem có đã phù hợp chưa và điều chỉnh sau khi cần thiết</p>
Bước 2. Thống nhất về cách hỏi và cách nhập thông tin giữa người thiết kế khảo sát, người đi thu thập số liệu, người nhập số liệu và người thiết kế khung nhập số liệu	<p>Trước khi tiến hành khảo sát, người thiết kế nghiên cứu và khung nhập liệu cần trao đổi cả về nội dung hỏi, cách hỏi và cách nhập số liệu vào khung nhập liệu với người tiến hành khảo sát và người nhập số liệu</p> <p>Đối với một nhóm khảo sát có nhiều thành viên, việc thống nhất về nội dung hỏi, cách hỏi, cách nhập số liệu là vô cùng quan trọng, cả nhóm cần trao đổi để đảm bảo tất cả các thành viên đều hiểu bản khảo sát, các câu hỏi và các phương án trả lời theo cùng một cách để khi đi thu thập số liệu việc đưa ra câu hỏi cho người trả lời đảm bảo sự đồng nhất giữa những người phỏng vấn.</p>
Bước 3. Phỏng vấn thực tế vài trường hợp để thử nghiệm thu thập và nhập liệu để điều chỉnh lại cả bảng hỏi và khung nhập liệu	<p>Quá trình khảo sát thực tế luôn có nhiều điều không nằm trong dự kiến. Do vậy cần phỏng vấn thực tế một vài trường hợp và yêu cầu nhóm khảo sát nhập thử số liệu bằng khung thiết kế</p> <p>Nếu có bất cứ vấn đề phát sinh trong quá trình hỏi thực tế và nhập số liệu thử (ví dụ cần bổ sung, chỉnh sửa câu hỏi và phương án trả lời) thì sẽ điều chỉnh bản hỏi và khung nhập liệu với sự phối hợp chặt chẽ giữa người thiết kế và nhóm khảo sát.</p>

Bên cạnh quy trình phối hợp giữa người thiết kế và người phỏng vấn cũng có một số lưu ý trong suốt quá trình thu thập số liệu để tránh phát sinh những sai sót của dữ liệu đến từ những nguyên nhân như: (1) điều kiện ngoại cảnh, (2) tiếp nhận thông tin sai lệch và (3) ghi nhận và truyền tải thông tin không đầy đủ, chính xác.

Điều kiện ngoại cảnh

Người trả lời có thể bị ảnh hưởng từ ngoại cảnh khiến câu trả lời không còn chính xác, người phỏng vấn cần hiểu và tránh những điều kiện ngoại cảnh có thể tác động đến người trả lời (Bảng 4).

Bảng 4. Một số điều kiện ngoại cảnh có thể ảnh hưởng đến người trả lời và cách phòng tránh sự ảnh hưởng

Điều kiện ngoại cảnh	Ví dụ một số ảnh hưởng	Gợi ý phương pháp
Tụ tập người trả lời đến cùng một địa điểm gây ảnh hưởng thông tin bởi hiệu ứng đám đông	Người trả sau có thể bắt chước câu trả lời của người trước khi họ không biết phải trả lời như thế nào hoặc trả lời theo số đông những người khác	Đến tận nhà người trả lời để phỏng vấn, thu thập thông tin thay vì yêu cầu họ đến cùng một địa điểm
Ảnh hưởng bởi cán bộ địa phương (cán bộ thôn, xã)	Khi đi thu thập số liệu, luôn cần thiết có sự giúp đỡ của cán bộ địa phương. Trong nhiều trường hợp, sự hiện diện của cán bộ địa phương giúp người trả lời cởi mở hơn. Tuy nhiên, trong một vài trường hợp có thể tạo sự căng thẳng, nỗi lo cho người trả lời, khiến họ không thoải mái trả lời	Nhạy bén trong tình hình thực tế để có thể mời cán bộ địa phương hiện diện hoặc không hiện diện trong buổi phỏng vấn.
Ảnh hưởng bởi chính người phỏng vấn	Người trả lời không hiểu rõ về mục đích thu thập thông tin khiến họ nghi vấn Người trả lời sợ thông tin mình cung cấp có thể ảnh hưởng đến họ sau này Ảnh hưởng bởi thời gian phỏng vấn, ví dụ như đang phỏng vấn nhưng sắp phải đi nấu cơm hoặc đi làm đồng nên trả lời đại khái, cho xong.	Cần giải thích rõ mục đích thu thập thông tin để người trả lời thoải mái, cởi mở Cần giải thích rõ nguyên tắc bảo mật, chia sẻ thông tin và quyền của người trả lời Người phỏng vấn cần cung cấp thông tin về thời gian từ trước khi phỏng vấn. Trong trường hợp chưa thể hoàn thành và có dấu hiệu trả lời đại khái, có thể tạm dừng và quay lại vào một khoảng thời gian khác mà người trả lời thoải mái hơn để tiếp tục

Tiếp nhận thông tin sai lệch

Nhiều trường hợp, người phỏng vấn làm việc ở địa bàn đã quen thuộc, khiến cho suy nghĩ và những hiểu biết của họ ảnh hưởng đến kết quả của bản khảo sát, vì vậy khi đi thu thập số liệu, người phỏng vấn cần khách quan, trung lập, không để quan điểm, hiểu biết của mình ảnh hưởng đến kết quả phỏng vấn và cần thật sự cởi mở trong việc tiếp nhận thông tin đến từ người trả lời. Bảng 5 đưa ra một số dấu hiệu của việc người phỏng vấn đang bị ảnh hưởng bởi thiên kiến và giải pháp khắc phục.

Bảng 5. Dấu hiệu của việc người phỏng vấn đang bị ảnh hưởng bởi thiên kiến

Dấu hiệu	Hành vi có thể gây sai lệch	Giải pháp khắc phục
Biết chắc câu trả lời	Bỏ qua câu hỏi và tự điền phương án mình cho là đúng	Hỏi đầy đủ các câu hỏi trong bản khảo sát.
Phỏng đoán câu trả lời: "Hầu hết mọi người đã trả lời như vậy nên chắc trường hợp này cũng thế"	Sau một thời gian và số lượng phỏng vấn nhất định, người phỏng vấn có thể đã thấy những điểm chung của các phương án nên bỏ qua câu hỏi và điền phương án giống với các trường hợp khác	Kể cả có biết rằng họ sẽ trả lời như vậy nhưng vẫn phải đưa ra câu hỏi. Không tự ý đoán, giả định và tự điền câu trả lời.
Cho rằng người trả lời sai	Vi bị ảnh hưởng bởi quan điểm cá nhân, nên nhận định người trả lời đang sai và đưa ra định hướng đến điều mình cho là đúng	Mọi câu trả lời đều cần được ghi nhận khách quan, không có tính định hướng và định kiến. Mọi câu trả lời đều là kết quả và cần được tôn trọng.

Ghi nhận và truyền tải thông tin không đầy đủ, chính xác

Khi ghi chép thông tin đã tiếp nhận từ người trả lời vào bản khảo sát, cần đảm bảo điền đầy đủ thông tin kể cả với những trường hợp câu trả lời của người được phỏng vấn là “không biết” “không có” hay “từ chối trả lời”; người phỏng vấn vẫn cần ghi chép lại những thông tin đó và cũng nên có diễn giải cho lý do vì sao họ lại “không biết” hoặc “từ chối trả lời” vì những lý do này rất quan trọng cho việc diễn giải kết quả sau này.

Thông tin ghi chép trong bản khảo sát nên rõ ràng và đầy đủ để sau này khi người phỏng vấn hoặc thành viên khác trong nhóm đọc lại vẫn có thể hiểu được nội dung của buổi phỏng vấn.

Ngoài ra sau khi hoàn thành phiếu hỏi, người phỏng vấn cần rà soát từ đầu đến cuối để kiểm tra thông tin đã đầy đủ, chính xác hay cần phải bổ sung, sửa đổi để có thể thực hiện ngay tại địa bàn làm việc tránh trường hợp cần phải sửa đổi sau khi ra khỏi địa bàn.

Trong trường hợp khi đi thu thập số liệu thực tế có những vấn đề phát sinh cần thay đổi nội dung bản khảo sát, người thu thập số liệu cần phối hợp với cả nhóm cũng như người thiết kế bản khảo sát để có phương án bổ sung, chỉnh sửa phù hợp.

Đồng thời, sau mỗi ngày thu thập số liệu, cả nhóm nên cùng tổng kết những trường hợp đặc biệt hoặc những điều cần lưu ý trước khi tiến hành thu thập số liệu vào ngày hôm sau.

Một số lỗi thường gặp phải khi thu thập số liệu có thể kể đến trong Bảng 6.

Bảng 6. Một số lỗi thường gặp phải khi thu thập số liệu

Lỗi thường gặp	Ví dụ cụ thể	Cách khắc phục
Câu hỏi bị để trống	Người phỏng vấn vô tình bỏ qua câu hỏi hoặc vô tình không điền đủ đáp án trong phiếu hỏi.	Đọc lại từ đầu đến cuối ngay sau khi phỏng vấn để đảm bảo không bỏ sót câu hỏi
	Với những câu trả lời không biết/không có/không trả lời, người phỏng vấn có thể sẽ coi đó không phải là một phương án nên có thể sẽ quên không ghi chép lại phương án trả lời đó	Ghi chép tất cả những gì người trả lời phản hồi sau mỗi câu hỏi. Mọi câu trả lời đều là kết quả

Thông tin cũng có thể bị sai lệch từ việc truyền tải dữ liệu không đầy đủ và chính xác. Nguyên nhân của việc nhập sai dữ liệu có thể đến từ việc người nhập liệu chưa hiểu rõ về khung nhập liệu, cũng có thể do trong quá trình nhập liệu bị phân tán, mất tập trung hoặc một số yếu tố khách quan khác v.v...

Để tối thiểu hoá việc nhập sai số liệu, người nhập số liệu cần tập trung, có thể tự kiểm tra từng câu sau mỗi phiếu nhập để đảm bảo mình đã nhập chính xác. Nên nghỉ ngơi thư giãn sau một số phiếu nhất định để đầu óc tỉnh táo, không nên thực hiện công việc này khi tâm trạng mệt mỏi.

Nếu việc nhập liệu được thực hiện theo nhóm, có thể kiểm tra chéo lẫn cho nhau sau khi đã nhập một số lượng phiếu nhất định (Việc kiểm tra chéo sẽ được trình bày trong phần tiếp theo).

Tuy là một bước đơn giản khi đã có sẵn khung nhập liệu được thiết kế phù hợp với bản khảo sát, nhưng quá trình này lại có thể phát sinh rất nhiều sai lệch về thông tin. Một số lỗi thường phát sinh khi nhập liệu cũng như các khắc phục được trình bày trong Bảng 7.

Bảng 7. Một số lỗi thường phát sinh cũng như cách khắc phục khi nhập liệu

Lỗi thường gặp	Ví dụ cụ thể/Nguyên nhân	Cách khắc phục
Nhập sai mã	Mã quy định 1=Nam, 2=Nữ mà lại nhập 3.	Cần thống nhất mã số cho các phương án trả lời giữa các thành viên trong nhóm Kiểm tra, rà soát dữ liệu, sẽ được trình bày trong phần tiếp theo
Nhập sai định dạng	Thừa một dấu cách hoặc dấu chấm có thể làm định dạng số thành định dạng chữ	Excel có tính năng phân loại định dạng ngay sau khi nhập, nếu nhập số sẽ căn lề bên phải ô dữ liệu, nếu nhập chữ sẽ là bên trái ô dữ liệu, nếu nhập số mà lại thấy căn lề trái thì tức là đang sai định dạng và cần kiểm tra lại
Nhập nhầm ô	Trong lúc nhập liệu vô tình có thể khiến con trỏ nhảy sang ô bên cạnh hoặc câu hỏi khác	Có thể sử dụng tính năng freeze panel của Excel để luôn nhìn thấy vị trí nhập liệu có đang nằm đúng câu hỏi không
Nhập thiếu thông tin	Trong bảng hỏi có mà không nhập vào khung dữ liệu, có thể do câu trả lời nằm ngoài các phương án đã có sẵn	Cần nhập đầy đủ thông tin, nếu trong khung dữ liệu không có chỗ để nhập, cần thảo luận với người thiết kế để bổ sung

Một ví dụ về việc nhập sai dữ liệu được thể hiện ở Hình 4, khi người nhập có thể nhập sai mã hoặc nhập sai định dạng đã được quy định từ trước.

3. Giờ: **14:00**
 4. Địa điểm: **Thôn 1**
 5. Người phỏng vấn: **Nguyễn Văn A**
 6. Tên người trả lời?: **Phạm Thị C**
 7. Độ tuổi?: **26**
 8. Giới tính?
 Nam Nữ
 9. Các nguồn thu nhập?
 Nông nghiệp
 Làm thuê
 Kinh doanh
 Công việc
 Khác, nếu có
 9.1. Nếu rõ nhập khác

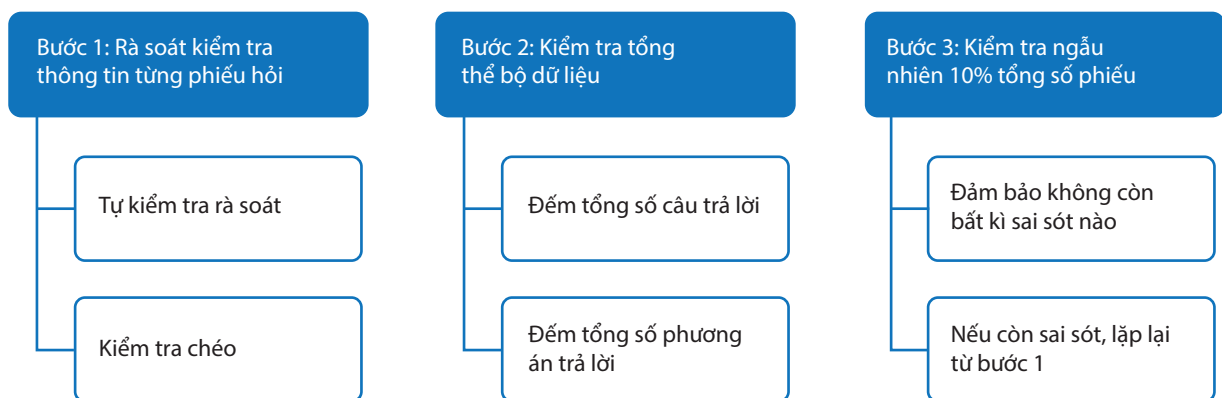
	A	B	C	D
2. Ngày	dd/mm/yyyy		23/12/2020	
3. Giờ	hh:mm		14:00	
4. Địa điểm	Điền chữ		Thôn 1	
5. Người phỏng vấn	Điền chữ		Nguyễn Văn A	
6. Tên người trả lời	Điền chữ		Phạm Thị C	
7. Độ tuổi	Điền số		26	
8. Giới tính	1=Nam; 2=Nữ		2	ĐÚNG
8. Giới tính	1=Nam; 2=Nữ		1	SAI
8. Giới tính	1=Nam; 2=Nữ		NỮ	SAI
9. Các nguồn thu nhập?	1=Nông nghiệp; 2=Làm thuê; 3=Kinh doanh; 4=Lương, ổn định;			
9.1. Nếu rõ nhập khác				5

Hình 4. Ví dụ về một số lỗi khi nhập liệu

Quy trình tổng hợp dữ liệu cũng cần đảm bảo độ chính xác. Lỗi thường xảy ra trong quá trình tổng hợp dữ liệu đó là: Copy, Paste nhầm, lệch vị trí tạo ra những khoảng trống hoặc chèn lên dữ liệu sẵn có gây mất dữ liệu. Để tránh gặp tình trạng này, chúng ta có thể kiểm tra tổng số dữ liệu từ các file riêng lẻ trước khi được tổng hợp và so sánh với dữ liệu đã được tổng hợp.

6 Làm sạch số liệu

Làm sạch dữ liệu giúp chúng ta xử lý các trường hợp trùng lặp số liệu, số liệu bị thiếu, số liệu bị nhập sai, số liệu không mang tính đại diện và những dạng số liệu không phù hợp, không được xác định từ trước. Sau khi nhập liệu, nên kiểm tra dữ liệu theo quy trình (1) Rà soát, kiểm tra từng phiếu, (2) Kiểm tra tổng thể bộ dữ liệu và (3) Kiểm tra ngẫu nhiên để đảm bảo thông tin được chuyển tiếp từ bản khảo sát sang khung dữ liệu đầy đủ và chính xác (Hình 5).



Hình 5. Quy trình kiểm tra và làm sạch dữ liệu

Trong cuốn cẩm nang này, chúng tôi sẽ hướng dẫn những kĩ năng và quy trình cơ bản một cách có hệ thống để có thể áp dụng cho nhiều trường hợp cũng như các công cụ nhập liệu khác nhau. Tuy nhiên những ví dụ cụ thể sẽ được trình bày qua phần mềm Microsoft Word và Excel, là những công cụ cơ bản, dễ sử dụng và phổ biến với đa số người sử dụng. Bên cạnh đó, chúng tôi sẽ nêu ra một vài sai sót về thông tin có thể xảy ra khi làm việc với số liệu từ đó đưa ra các phương pháp để phòng tránh và sửa chữa những sai sót đó.

6.1 Bước 1: Rà soát, kiểm tra từng phiếu

Cách tốt nhất để thực hiện bước kiểm tra này là ngay sau khi nhập liệu, người nhập liệu tự rà soát các thông tin của từng câu trả lời trong phiếu đã nhập. Quy trình kiểm tra đó là so sánh từng câu hỏi giữa bản khảo sát và khung nhập liệu để xem thông tin và dữ liệu đã được truyền tải đầy đủ và chính xác hay chưa. Trong trường hợp có sai sót, có thể sửa chữa bổ sung ngay. Sau khi người nhập liệu đã tự tin với dữ liệu đã nhập, có thể tiến hành quy trình kiểm tra chéo trong nhóm để có thể phát hiện những lỗi mà chính bản thân người nhập liệu có thể không nhận thức được.

Quy trình kiểm tra chéo có thể sử dụng khi có 2 người trở lên. Việc tổ chức kiểm tra chéo nhằm đảm bảo tất cả bản khảo sát của mỗi thành viên đều được kiểm tra một thành viên khác và hình thành 1 vòng lặp giữa tất cả thành viên (Hình 6).

Trong quá trình này, mỗi thành viên sẽ rà soát, kiểm tra từng phiếu trong bản phỏng vấn và khung nhập liệu của một thành viên khác. Tuy nhiên nếu phát hiện sai sót trong một phiếu khảo sát, người kiểm tra sẽ chỉ đánh dấu, ghi chép lại vị trí bị sai và sau đó yêu cầu người thu thập thông tin của phiếu khảo sát đó sửa lại những sai sót và lặp lại quy trình cho đến khi sửa toàn bộ lỗi và được người kiểm tra xác nhận rằng phiếu đã hoàn toàn đầy đủ và chính xác.



Hình 6. Quy trình kiểm tra chéo giữa các thành viên

6.2 Bước 2: Kiểm tra tổng thể bộ dữ liệu

Bên cạnh việc rà soát kiểm tra từng phiếu, chúng ta có thể kiểm tra tổng thể bộ dữ liệu để phát hiện nhanh những trường hợp thiếu thông tin hoặc những trường hợp thông tin thiếu chính xác. Việc kiểm tra tổng thể này có thể thực hiện thủ công bằng cách quan sát các giá trị của bộ dữ liệu, tuy nhiên đối với những bộ dữ liệu có nhiều giá trị, đồng thời để tiết kiệm thời gian, chúng tôi gợi ý một số quy trình kiểm tra như sau:

Kiểm tổng giá trị có trong mỗi dải dữ liệu (đếm tổng số câu trả lời ứng với mỗi câu hỏi): với mỗi một câu hỏi trong bản khảo sát, khi nhập vào khung dữ liệu sẽ được lưu lại thành một dải dữ liệu vì vậy thông thường số lượng câu trả lời cần bằng với tổng số lượng người được phỏng vấn. Vì vậy chúng ta có thể sử dụng phương pháp này để có thể kiểm tra sự đầy đủ của dữ liệu cho mỗi câu hỏi.

Ví dụ: Trong phần mềm Excel, chúng ta có thể dùng hàm 'count' hoặc 'counta' để có thể kiểm tra nhanh những câu trả lời nào bị thiếu. Ví dụ trong Hình 7, khi sử dụng hàm 'counta' để đếm tổng số các câu trả lời của các câu hỏi, thì câu hỏi số '7. Độ tuổi' và câu hỏi số '8. Giới tính' đang thiếu câu trả lời vì giá trị của hàm 'counta' trả về chỉ là 9 thay vì 10. Vì vậy ta có thể biết được hai câu hỏi này đang thiếu giá trị, và có thể nhanh chóng tìm ra giá trị bị thiếu ở mỗi câu, ví dụ phiếu số 2 bị thiếu giá trị ở câu hỏi độ tuổi và phiếu số 7 bị thiếu giá trị ở câu hỏi giới tính.

Lưu ý: trong Excel hàm 'count' đếm những dải dữ liệu ở định dạng số, nên nếu trong dải dữ liệu có những câu trả lời dạng chữ cần phải dùng đến hàm 'counta'. Ngoài ra có thể dùng hàm 'countblank' để đếm những ô đang bỏ trống.

Phương pháp kiểm tra này có thể giúp chúng ta kiểm tra sự đầy đủ thông tin của bộ số liệu, có thể sử dụng cho cả dữ liệu định lượng và định tính. Tuy nhiên, phương pháp này chưa thể kiểm tra được độ chính xác của thông tin.

Đếm tổng giá trị ứng với mỗi phương án trong dải dữ liệu (đếm số từng phương án trả lời trong mỗi câu hỏi): trong mỗi câu hỏi với nhiều phương án trả lời, tổng số các câu trả lời của các phương án phải bằng với tổng số câu trả lời của câu hỏi. Phương pháp này sẽ giúp chúng ta tìm ra những giá trị bị sai định dạng hoặc nhập sai mã để kiểm tra tính chính xác của những thông tin đã được mã hoá hoặc có chung một định dạng.

Ví dụ: trong 8 thì câu hỏi số '8. Giới tính' đã được mã hoá các phương án Nam là 1 và Nữ là 2. Tuy nhiên trong quá trình nhập liệu, có thành viên nhập Nữ vào ô số liệu hoặc một thành viên khác nhập số 3, nếu chỉ dùng phương pháp đếm tổng số phương án như 'counta', giá trị trả về vẫn sẽ là 10.

	A	B	C	D	E	F	G	H	I	J
1	1. Mã số phiếu	2. Ngày	3. Giờ	4. Địa điểm	5. Người phỏng vấn	6. Tên người trả lời	7. Độ tuổi	8. Giới tính	9. Các nguồn thu nhập	
2		dd/mm/yyyy	hh:mm	Điện	Điện chữ	Điện chữ	Điện số	1=Nam; 2=Nữ	1=Nông nghiệp; 2=Làm thuê; 3=Kinh doanh; 4=Lương ổn định; 5=Khác	
3	Số	y		chữ	Điện chữ	Điện chữ	Điện số			
3	1	25/12/2020	14:00	Thôn 1	Nguyễn Văn A	Trần Thị B	20	2	1	
4	2	26/12/2020	15:00	Thôn 1	Nguyễn Văn A	Trần Thị C		2	2	
5	3	27/12/2020	16:00	Thôn 1	Nguyễn Văn A	Trần Thị D	23	2	1	
6	4	28/12/2020	17:00	Thôn 1	Nguyễn Văn A	Nguyễn Văn E	25	1	1	
7	5	29/12/2020	18:00	Thôn 1	Nguyễn Văn A	Nguyễn Văn F	48	1	1	
8	6	30/12/2020	19:00	Thôn 1	Nguyễn Văn A	Nguyễn Văn G	63	2	1	
9	7	31/12/2020	20:00	Thôn 1	Nguyễn Văn A	Hoàng Diệu X	18	1	3	
10	8	01/01/2021	21:00	Thôn 1	Nguyễn Văn A	Hoàng Diệu Y	42		4	
11	9	02/01/2021	22:00	Thôn 1	Nguyễn Văn A	Hoàng Diệu Z	59	2	3	
12	10	03/01/2021	23:00	Thôn 1	Nguyễn Văn A	Trần Thị A	55	2	1	
13										
14	10	10	10	10	10	10	9	9	10	
15						COUNTA(F3:F12)				
16										

Hình 7. Kiểm tra sự đầy đủ của dữ liệu bằng hàm 'counta' trong Excel

Trong Excel ta có thể dùng hàm 'countif' hoặc 'countifs' để đếm số lượng của từng phương án như là có bao nhiêu phương án 1 và bao nhiêu phương án 2 và sau đó xem tổng số câu trả lời của các phương án này có trùng với tổng số 10 câu trả lời của câu hỏi về giới tính này không.

Như ví dụ trong Hình 9, sử dụng hàm 'countif' sẽ trả về 4 phương án cho giá trị 1 và 4 phương án cho giá trị 2, tổng số là 8 phương án, không trùng khớp với 10 phương án của câu hỏi, như vậy có thể xác định câu hỏi về giới tính này đang có 2 giá trị bị sai định dạng hoặc số liệu và có thể tiến hành rà soát các giá trị trong câu hỏi này để tìm ra phiếu số 3 nhập 'Nữ' và phiếu số 6 nhập '3'

	A	B	C	D	E	F	G	H	I	J
		dd/mm/yyyy	hh:mm	Điện	Điện chữ	Điện chữ	Điện số	1=Nam; 2=Nữ	1=Nông nghiệp; 2=Làm thuê; 3=Kinh doanh; 4=Lương ổn định; 5=Khác	
	Số	y		chữ	Điện chữ	Điện chữ	Điện số			
	1	25/12/2020	14:00	Thôn 1	Nguyễn Văn A	Trần Thị B	20	2	1	
	2	26/12/2020	15:00	Thôn 1	Nguyễn Văn A	Trần Thị C	45	2	2	
	3	27/12/2020	16:00	Thôn 1	Nguyễn Văn A	Trần Thị D	23	Nữ	1	
	4	28/12/2020	17:00	Thôn 1	Nguyễn Văn A	Nguyễn Văn E	25	1	1	
	5	29/12/2020	18:00	Thôn 1	Nguyễn Văn A	Nguyễn Văn F	48	1	1	
	6	30/12/2020	19:00	Thôn 1	Nguyễn Văn A	Nguyễn Văn G	63	3	1	
	7	31/12/2020	20:00	Thôn 1	Nguyễn Văn A	Hoàng Diệu X	18	1	3	
	8	01/01/2021	21:00	Thôn 1	Nguyễn Văn A	Hoàng Diệu Y	42	1	4	
	9	02/01/2021	22:00	Thôn 1	Nguyễn Văn A	Hoàng Diệu Z	59	2	3	
	10	03/01/2021	23:00	Thôn 1	Nguyễn Văn A	Trần Thị A	55	2	1	
	10	10	10	10	10	10	10	10	10	
								4	COUNTIF(H3:H12,1)	
								4	COUNTIF(H3:H12,2)	
									H12,2)	

Hình 8. Kiểm tra sự chính xác của dữ liệu bằng 'countif' trong Excel

Lưu ý: phương pháp này chỉ phát huy hiệu quả đối với những câu hỏi mà câu trả lời được giới hạn với số lượng nhóm nhất định hoặc đã được mã hoá từ trước. Với những câu hỏi định tính hoặc các phương án trả lời mở và không thể mã hoá trước, phương pháp này sẽ không hiệu quả. Thay vào đó chúng ta nên sử dụng công cụ filter, một công cụ rất mạnh trong Excel hỗ trợ quản lý số liệu.

Sử dụng công cụ filter của Excel để kiểm tra dữ liệu:

Công cụ Filter của Excel có rất nhiều chức năng để quản lý, sắp xếp và làm việc với dữ liệu, trong phạm vi của cẩm nang, ta có thể sử dụng Filter để hỗ trợ cho quá trình kiểm tra tổng thể toàn bộ dữ liệu.

Một điểm lưu ý khi sử dụng Filter: Công cụ Filter có khả năng tự nhận biết vùng dữ liệu khi ta bôi đen chọn 1 ô hoặc 1 hàng của bộ dữ liệu nhưng nếu có cột hoặc hàng trống trong dữ liệu thì Filter sẽ nhận diện sai vì vậy chúng ta nên bôi đen chọn toàn bộ vùng dữ liệu sẵn có trước khi ấn vào công cụ filter thay vì chỉ bôi đen 1 ô hay 1 dòng đầu tiên của bộ dữ liệu (Hình 9).

1. Mã số phiếu	2. Ngày	3. Giờ	4. Địa điểm	5. Người phỏng vấn	6. Tên người trả lời	7. Độ tuổi	8. Giới tính	9. Các nguồn thu nhập	9.1. Nội nguồn thu nhập khác
	dd/mm/yy	hh:mm	Diễn chữ	Diễn chữ	Diễn chữ	Diễn số	1=Nam; 2=Nữ	1=Nông nghiệp; 2=Làm thuê; 3=Kinh doanh; 4=Lương ổn định; 5=Khác	Diễn chữ
Số	y								
3	1	25/12/2020	14:00	Thôn 1	Nguyễn Văn Trần Thị B	20	2	1	
4	2	26/12/2020	15:00	Thôn 1	Nguyễn Văn Trần Thị C	45	2	2	
5	3	27/12/2020	16:00	Thôn 1	Nguyễn Văn Trần Thị D	23	2	1	
6	4	28/12/2020	17:00	Thôn 1	Nguyễn Văn Nguyễn Văn	25	1	5	Thất nghiệp
7	5	29/12/2020	18:00	Thôn 1	Nguyễn Văn Nguyễn Văn	48	1	1	
8	6	30/12/2020	19:00	Thôn 1	Nguyễn Văn Nguyễn Văn	63	1	5	Về hưu, con nuôi
9	7	31/12/2020	20:00	Thôn 1	Nguyễn Văn Hoàng Diệu \	18	1	3	
10	8	01/01/2021	21:00	Thôn 1	Nguyễn Văn Hoàng Diệu \	42	1	5	
11	9	02/01/2021	22:00	Thôn 1	Nguyễn Văn Hoàng Diệu z	59	2	3	
12	10	03/01/2021	23:00	Thôn 1	Nguyễn Văn Trần Thị A	55	2	1	

Hình 9. Lưu ý khi sử dụng công cụ Filter trong Excel

Đối với những bộ dữ liệu nhiều và dài, ta có thể sử dụng Filter để quản lý các giá trị trong từng câu hỏi. Ví dụ như ở câu hỏi '8. Giới tính' Filter sẽ có thể liệt kê ra các phương án đang có trong dải dữ liệu, ta có thể nhìn thấy dễ dàng các phương án sai và lọc ra những phương án đó như Hình 10 và Hình 11.

Ngoài ra, Filter còn đặc biệt hữu dụng khi có thể kiểm tra các câu hỏi ở dạng nối tiếp (follow-up question). Ví dụ khi được hỏi về nguồn thu nhập, sau khi người trả lời chọn phương án 'khác', họ sẽ nêu rõ nguồn thu nhập khác đó, ta có thể sử dụng Filter để kiểm tra tổng giá trị có trong cột nêu rõ đã đủ chưa. Ví dụ như trong Hình 13, với câu '9. Các nguồn thu nhập' sau khi dùng Filter để chọn phương án '5=Khác' và ẩn những kết quả khác đi, ta có thể sử dụng Filter để kiểm tra câu '9.1 Nêu rõ nguồn thu nhập khác' đã đủ hay chưa, và với trường hợp Filter liệt kê ra kết quả (Blanks) như Hình 12, có thể thấy Phiếu số 8 đã bỏ trống câu trả lời 9.1

1. Mã số phiếu	2. Ngày	3. Giờ	4. Địa điểm	5. Người phỏng vấn	6. Tên người trả lời	7. Độ tuổi	8. Giới tính
1	25/12/2020	14:00	Thôn 1	Nguyễn Văn	Trần Thị B	20	2
2	26/12/2020	15:00	Thôn 1	Nguyễn Văn	Trần Thị C	45	2
3	27/12/2020	16:00	Thôn 1	Nguyễn Văn	Trần Thị D	23	NỮ
4	28/12/2020	17:00	Thôn 1	Nguyễn Văn	Nguyễn Văn	25	1
5	29/12/2020	18:00	Thôn 1	Nguyễn Văn	Nguyễn Văn	48	1
6	30/12/2020	19:00	Thôn 1	Nguyễn Văn	Nguyễn Văn	63	3
7	31/12/2020	20:00	Thôn 1	Nguyễn Văn	Hoàng Diệu Y	18	1
8	01/01/2021	21:00	Thôn 1	Nguyễn Văn	Hoàng Diệu Y	42	
9	02/01/2021	22:00	Thôn 1	Nguyễn Văn	Hoàng Diệu Z	59	2
10	03/01/2021	23:00	Thôn 1	Nguyễn Văn	Trần Thị A	55	2

Hình 10. Sử dụng Filter để liệt kê dữ liệu sẵn có

1. Mã số phiếu	2. Ngày	3. Giờ	4. Địa điểm	5. Người phỏng vấn	6. Tên người trả lời	7. Độ tuổi	8. Giới tính
3	27/12/2020	16:00	Thôn 1	Nguyễn Văn	Trần Thị D	23	NỮ
6	30/12/2020	19:00	Thôn 1	Nguyễn Văn	Nguyễn Văn	63	3
8	01/01/2021	21:00	Thôn 1	Nguyễn Văn	Hoàng Diệu Y	42	

Hình 11. Sử dụng Filter để lọc những dữ liệu sai và thiếu

Số	dd/mm/yyyy	hh:mm	Địa chỉ	Địa chỉ	Địa chỉ	Địa chỉ	1=Nam; 2=Nữ	1=Nông nghiệp; 2=Làm thue; 3=Kinh doanh; 4=Lương lớn; 5=Khác định;	Địa chỉ
4	28/12/2020	17:00	Thôn 1	Nguyễn Văn	Nguyễn Văn	Nguyễn Văn	25	1	5 Thất nghiệp
6	30/12/2020	19:00	Thôn 1	Nguyễn Văn	Nguyễn Văn	Nguyễn Văn	63	1	5 Về hưu, con n
8	01/01/2021	21:00	Thôn 1	Nguyễn Văn	Hoàng Diệu \	Nguyễn Văn	42	1	5

Hình 12. Sử dụng Filter để kiểm tra câu hỏi dạng follow-up

Quá trình kiểm tra tổng thể bộ số liệu có thể giúp chúng ta xác định nhanh một số trường hợp thiếu và sai lệch thông tin nhằm hỗ trợ cho quá trình rà soát và kiểm tra từng phiếu. Tuy nhiên, để xác thực những quá trình kiểm tra trên cũng như kiểm chứng mức độ đáng tin cậy của bộ dữ liệu, chúng ta cần tiến hành quá trình kiểm tra ngẫu nhiên.

6.3 Bước 3: Kiểm tra ngẫu nhiên

Đây là bước cuối cùng trong quá trình kiểm tra và làm sạch số liệu để đảm bảo dữ liệu đã chính xác và đáng tin cậy.

Với phương pháp này, chúng ta sẽ đánh số thứ tự và lấy mẫu ngẫu nhiên ít nhất 10% tổng số phiếu (hoặc sai số tối thiểu có thể chấp nhận được) trong tất cả phiếu phỏng vấn để rà soát đối chứng từng câu trả lời giữa bản khảo sát và khung nhập liệu như Bước 1 đồng thời ghi chép lại quá trình kiểm tra và không thực hiện bất cứ chỉnh sửa gì đối với bộ dữ liệu. Nếu trong số 10% phiếu được kiểm tra ngẫu nhiên vẫn có lỗi sai, bộ dữ liệu chưa đủ độ tin cậy và sẽ phải lặp lại quá trình kiểm tra từ Bước 1, đó là trả về cho người phỏng vấn tự rà soát lại số liệu và chỉnh sửa lại những lỗi sai.

Việc kiểm tra ngẫu nhiên này cũng cần thực hiện kiểm tra chéo, tốt nhất là được thực hiện bởi một nhóm khác, có thể là những thành viên không tham gia vào quá trình thu thập và nhập số liệu để đảm bảo tính khách quan và trung lập. Trong trường hợp không có đủ nhân sự, vẫn cần đảm bảo việc kiểm tra ngẫu nhiên được thực hiện dưới hình thức kiểm tra chéo trong nhóm, tức là người phỏng vấn sẽ không kiểm tra phiếu của chính mình mà sẽ để một người khác trong nhóm kiểm tra.

Nếu dữ liệu đã vượt qua toàn bộ ba bước của quá trình kiểm tra và không còn phát sinh bất kỳ một lỗi nào, chúng ta có thể tự tin khẳng định bộ dữ liệu đã được truyền tải đầy đủ và chính xác từ bản khảo sát sang khung nhập liệu.

Sau khi đã đảm bảo bộ số liệu đã chính xác và đáng tin cậy, chúng ta có thể dùng số liệu cho việc phân tích. Tuy nhiên, để tránh mọi tác động lên dữ liệu gốc, chúng ta nên lưu trữ dữ liệu này ở một nơi an toàn và tạo nhiều bản copy từ dữ liệu gốc này và chỉ làm việc trên các bản copy đó.

Ngoài ra, với những bản khảo sát bằng giấy, ngoài việc photo lại để lưu trữ cũng nên được số hoá bằng cách scan lại thành bản mềm PDF để lưu trữ tránh mất mát dữ liệu. Vì một bộ số liệu có thể đóng góp cho những công việc kéo dài rất nhiều năm, nên việc lưu trữ cũng nên được thực hiện cẩn thận để tránh nhiều trường hợp nhất có thể xảy ra. Ví dụ như khi có sự cố hoả hoạn, lũ lụt làm

ảnh hưởng đến toàn bộ dữ liệu giấy bản gốc, ta vẫn có 1 bản scan để có thể khôi phục lại dữ liệu giấy, hoặc ngược lại, khi dữ liệu scan hoặc bộ dữ liệu trong ổ cứng bị hỏng hoặc mất, ta vẫn có thể khôi phục lại từ dữ liệu giấy.

Trong một số trường hợp phải nhập/xuất dữ liệu để phân tích (ví dụ như nhập dữ liệu từ Excel vào một phần mềm tính toán chuyên dụng khác), cần phải kiểm tra xem quá trình nhập/xuất dữ liệu có phát sinh vấn đề gây thiếu hoặc sai sót dữ liệu không. Chúng ta có thể áp dụng cách kiểm tra tổng số và kiểm tra ngẫu nhiên để làm việc này.

7 Đảm bảo có bộ phận chuyên trách về quản lý dữ liệu và ngân sách cho công việc quản lý số liệu

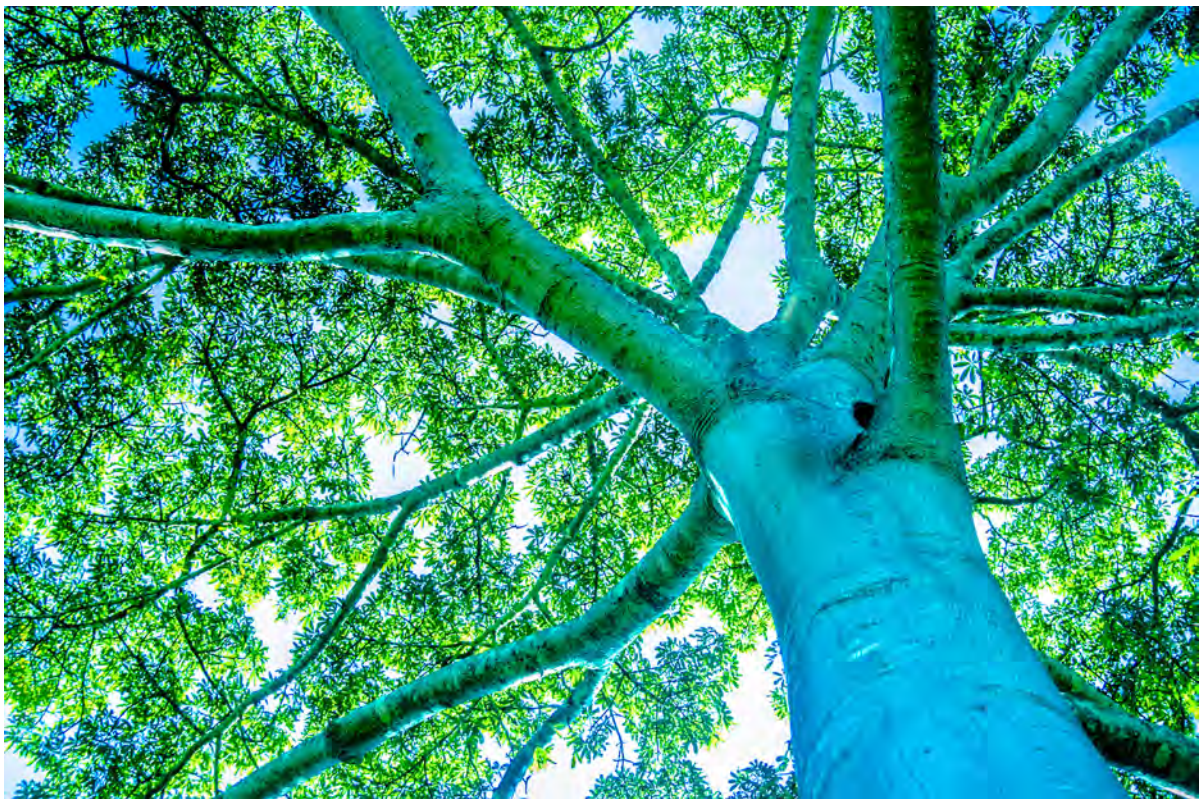
Để tổ chức có một hệ thống quản lý dữ liệu hiệu quả và an toàn, cần có một đội ngũ hoặc nhân viên chuyên trách về quản lý số liệu cũng như ngân sách phân bổ cho hoạt động này. Nhân viên chuyên trách về quản lý số liệu sẽ xây dựng quy trình chuẩn của thu thập, quản lý và lưu trữ số liệu mà các thành viên khác của tổ chức phải tuân theo, đồng thời chịu trách nhiệm đào tạo, hướng dẫn và giám sát các bên có liên quan trong việc thực hiện quy trình này. Điều này cũng tránh việc số liệu của tổ chức thường được lưu giữ tản mạn, không có hệ thống và chỉ được lưu trữ của một cá nhân đơn lẻ dẫn đến mất mát về mặt số liệu. Việc có một bộ phận hay nhân viên chuyên trách cũng sẽ giúp việc tổng hợp số liệu được nhanh chóng và thuận tiện hơn thay vì phải trao đổi với từng phòng ban và cán bộ mỗi người giữa số liệu riêng lẻ.

Việc phân bổ ngân sách thường xuyên và ổn định để quản lý dữ liệu thường bị xem nhẹ trong quá trình xây dựng dự án và xây dựng kế hoạch hàng năm của các tổ chức. Việc này dẫn đến sự đứt quãng trong việc thu thập và quản lý số liệu và đồng thời đẩy chi phí giao dịch và thực hiện của dự án cao hơn so với dự định (ví dụ nếu có hệ thống quản lý số liệu tốt, tổ chức chỉ mất có 1 ngày để thu thập và phân tích số liệu cần thiết, nhưng nếu không có, các bên phải mất nhiều thời gian và kinh phí để thu thập và tổng hợp lại số liệu mới có thể dùng được).

Đừng tiết kiệm khi thiết lập quản trị dữ liệu bởi cần có ngân sách phù hợp để đảm bảo khả năng lưu trữ, chất lượng, tính bảo mật của dữ liệu cũng như đào tạo cho cán bộ trong tổ chức về các kỹ năng quản lý dữ liệu này.

8 Đào tạo và nâng cao năng lực cho cán bộ

Phần lớn cán bộ trong các tổ chức nghiên cứu và NGOs thường không có kĩ năng về quản lý số liệu. Tổ chức cần xây dựng các chương trình đào tạo ngay từ khi các cán bộ được tuyển dụng giúp các cán bộ hiểu rõ quy trình, nguyên tắc và quy định thu thập, quản lý, lưu trữ số liệu mà tổ chức đã đề ra. Ngoài ra, nâng cao năng lực cho cán bộ trong việc cập nhật ứng dụng khoa học công nghệ và công nghệ thông tin trong việc thu thập, xử lý và làm sạch, quản lý và lưu trữ số liệu cũng giúp tiết kiệm chi phí quản lý và nâng cao tính hiệu quả của việc quản lý số liệu.



Cây rừng tại Entre Rios. Ảnh chụp bởi Cooke Vieira/CIFOR

9 Kết luận

Quản lý số liệu hiệu quả đóng vai trò quan trọng trong việc vận hành và hoạt động của tổ chức. Để xây dựng và vận hành một hệ thống quản lý số liệu hiệu quả cần có tầm nhìn và mục tiêu quản lý rõ ràng, quy trình và hệ thống toàn diện, cũng như đầu tư cả về tài chính lẫn nguồn lực lâu dài và ổn định. Việc thu thập, nhập, lưu trữ và quản lý số liệu có nhiều rủi ro sai sót và do vậy cần có các biện pháp và chiến lược phòng tránh rõ ràng. Tài liệu này hi vọng sẽ cung cấp một phần kiến thức cơ bản để giúp các bên trong việc quản lý số liệu có hệ thống và hiệu quả.

10 Tài liệu tham khảo

- Ainsworth Q. 2020. *Quản lý dữ liệu là gì*. California: JotForm. Truy cập ngày 16 tháng 12 năm 2020: <https://www.jotform.com/what-is-data-management/>
- Dallemand J. 2019. *Năm nguyên tắc quản lý số liệu bạn không thể từ chối năm 2019*. Tây Ban Nha: Datumize. Truy cập ngày 16 tháng 12 năm 2020: <https://blog.datumize.com/5-data-management-principles-you-cant-ignore-in-2019>
- [IGGI] Intra-governmental Group on Geographic Information. 2005. *Nguyên tắc quản lý số liệu tốt*. London: Văn phòng phó thủ tướng. Truy cập ngày 16 tháng 12 năm 2020: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/14867/Good_dataMan.pdf
- Rouse M. 2019. *Quản lý dữ liệu là gì và tại sao nó quan trọng?*. Truy cập ngày 16 tháng 12 năm 2020: <https://searchdatamanagement.techtarget.com/definition/data-management>
- Tableau. 2020. *Năm bước chính để tạo ra một chiến lược quản lý số liệu*. Truy cập ngày 16 tháng 12 năm 2020: <https://www.tableau.com/learn/articles/data-management-strategy>

DOI: 10.17528/cifor/007888

Các báo cáo chuyên đề của CIFOR bao gồm các kết quả nghiên cứu sơ bộ hoặc nâng cao về các vấn đề về rừng ở các khu vực nhiệt đới và cận được công bố vào thời điểm thích hợp để tạo ra và thúc đẩy các cuộc thảo luận. Nội dung báo cáo đã được rà soát nội bộ nhưng chưa trải qua quá trình bình duyệt từ các chuyên gia bên ngoài tổ chức.

Tài liệu này được xây dựng nhằm chia sẻ kinh nghiệm với các nhà nghiên cứu và các nhà quản lý dự án về quy trình lập kế hoạch, thiết kế, thu thập và quản lý số liệu cho các nghiên cứu xã hội học. Tài liệu không chỉ đưa ra các nguyên tắc và các yếu tố cần xem xét mà còn cung cấp các ví dụ thực tiễn để giúp các nhà nghiên cứu và các nhà quản lý dự án tránh khỏi các sai lầm trong quá trình thu thập, quản lý và xử lý số liệu.



RESEARCH
PROGRAM ON
Forests, Trees and
Agroforestry

Chương trình nghiên cứu của CGIAR về Rừng, Cây gỗ và Nông lâm kết hợp (FTA) là chương trình phát triển nghiên cứu lớn nhất thế giới nhằm nâng cao vai trò của rừng, cây gỗ và nông lâm kết hợp với mục tiêu phát triển bền vững và đảm bảo lương thực để ứng phó với biến đổi khí hậu. CIFOR chủ trì các nghiên cứu FTA trong mối quan hệ đối tác chiến lược với Bioversity International, CATIE, CIRAD, INBAR, ICRAF và TBI.

Nghiên cứu này được hỗ trợ bởi Quỹ đối tác CGIAR: cigar.org/funders/

cifor.org | forestsnews.cifor.org



Norad



Federal Ministry
for the Environment, Nature Conservation
and Nuclear Safety



Trung tâm Nghiên cứu Lâm nghiệp Quốc tế (CIFOR)

CIFOR thúc đẩy sự phồn vinh của nhân loại, cải thiện bảo vệ môi trường và thúc đẩy sự bình đẳng thông qua tiến hành các nghiên cứu sáng tạo, nâng cao năng lực của các bên đối tác, tích cực tham gia đối thoại với các bên liên quan để hỗ trợ định hình chính sách và thực tiễn tác động tới rừng và con người. CIFOR là tổ chức nghiên cứu thuộc liên minh CGIAR và chủ trì các chương trình nghiên cứu của CGIAR về Rừng, Cây gỗ và Nông lâm kết hợp (FTA). Trụ sở chính của CIFOR đặt tại Bogor, Indonesia và các văn phòng của CIFOR có mặt tại Nairobi, Kenya; Yaounde, Cameroon; Lima, Peru và Bonn, Germany.

