

BỘ THÔNG TIN VÀ TRUYỀN THÔNG

Sổ tay Dữ liệu mở

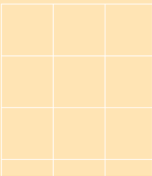
Phiên bản 1.0





Lời nói đầu

Dữ liệu mở là một trong những nội dung thể hiện sự minh bạch, tinh thần kiến tạo của Chính phủ đối với đổi mới sáng tạo, thúc đẩy phát triển chính phủ số, kinh tế số, xã hội số. Đây cũng là nội dung được nhiều tổ chức quốc tế đánh giá để nhìn nhận về sự cởi mở, minh bạch của cơ quan nhà nước. Vậy dữ liệu mở là gì? Tại sao lại cần quan tâm? và làm thế nào để mở dữ liệu?. Hãy cùng nhau tìm hiểu các nội dung liên quan đến dữ liệu mở.



Nội dung

- 01 Dữ liệu mở là gì?
- 02 Tại sao cần cung cấp dữ liệu mở?
- 03 Việt Nam đang ở đâu trên thế giới về dữ liệu mở?
- 04 Tìm và khai thác dữ liệu mở ở đâu?
- 05 Khi sử dụng dữ liệu mở phải tuân thủ yêu cầu gì?
- 06 Giấy phép sử dụng dữ liệu mở là gì?
- 07 Các cơ quan nhà nước phải làm gì?
- 08 Dữ liệu gì nên là dữ liệu mở?
- 09 Nguyên tắc cung cấp dữ liệu mở?
- 10 Thuật ngữ.



Dữ liệu mở là gì?

Giống như ghế đá trong công viên, bạn tự do ngồi trên ghế, ngồi trên ghế để bán hàng của bạn, mời người khác ngồi cùng. Tuy nhiên, bạn không được giữ chỗ để bán chỗ đó cho người khác hay biến nó thành chỗ sử dụng riêng hay thậm chí làm cho người khác hiểu lầm rằng ghế đó là của bạn.

Dữ liệu mở cũng tương tự thế, dữ liệu mở là dữ liệu mà mọi người có thể tự do sử dụng, tái sử dụng hay chia sẻ cho người khác. Điều hạn chế cơ bản đối với dữ liệu mở là người sử dụng dữ liệu mở phải ghi nhận nguồn gốc của dữ liệu mở khi sử dụng cũng như cung cấp lại dữ liệu cho người khác.

Dữ liệu mở của cơ quan nhà nước là dữ liệu mở được cơ quan nhà nước có thẩm quyền công bố rộng rãi cho cơ quan, tổ chức, cá nhân tự do sử dụng, tái sử dụng, chia sẻ.

Dữ liệu mở của cơ quan nhà nước được quy định tại Khoản 4, Điều 3 của Nghị định số 47/2020/NĐ-CP ngày 9/4/2020.

Tại sao cơ quan nhà nước lại cần cung cấp dữ liệu mở?



MINH BẠCH

Nâng cao tính công khai, minh bạch của cơ quan nhà nước; tạo điều kiện thuận lợi cho việc đo lường tác động của các chính sách; nâng cao hiệu quả các dịch vụ công.

HUY ĐỘNG NGUỒN LỰC

Tạo điều kiện thuận lợi cho sự đóng góp của của người dân vào hoạt động của cơ quan nhà nước

KIẾN TẠO

Cung cấp tài nguyên để cộng đồng tạo ra tri thức mới, dịch vụ và sản phẩm mới

VÍ DỤ

- Dữ liệu về triệu chứng bệnh có thể giúp các bác sỹ nghiên cứu để tìm ra phương pháp điều trị bệnh, tiết kiệm rất nhiều thời gian và công sức, chi phí; doanh nghiệp sử dụng để nghiên cứu các loại thuốc điều trị bệnh
- Dữ liệu camera giao thông, quan trắc môi trường được cung cấp dữ liệu mở cho người dân, doanh nghiệp sử dụng dữ liệu mở cũng có tác dụng như nhà nước huy động thêm hàng ngàn cộng tác viên giám sát giao thông, môi trường trong khi nguồn lực của nhà nước còn hạn chế.

Cung cấp dữ liệu mở, chính là cùng nhau xây dựng một xã hội tốt đẹp hơn, sáng tạo hơn.

Việt Nam đang ở đâu trong thế giới về cung cấp dữ liệu mở?

Hơn 80% các nước trên thế giới có cổng dữ liệu quốc gia để cung cấp dữ liệu mở.

Việt Nam đang tích cực triển khai cung cấp dữ liệu mở. Năm 2022, Việt Nam xếp thứ 81 trên 187 quốc gia về cung cấp dữ liệu mở. Tăng 10 bậc so với năm 2020.

MỘT SỐ HOẠT ĐỘNG VÀ LỢI ÍCH ĐIỂN HÌNH TRÊN THẾ GIỚI

- Các dự án liên quan được phát triển từ dữ liệu mở của Phần Lan, Vương quốc Anh đã thể hiện trực quan được ngân sách thu từ thuế, lệ phí của người dân được chính phủ chi tiêu như thế nào.
- Một dự án sử dụng dữ liệu mở đã giúp cho Canada tiết kiệm được 3.2 tỷ đô la gian lận từ tiền thuế từ thiện.
- Website như folketsting.dk của Đan Mạch giúp người dân giám sát được các hoạt động của quốc hội và quá trình làm luật cũng như những đại biểu quốc hội có liên quan.
- Website findtoilet.dk được phát triển từ dữ liệu mở bởi một cá nhân đã giúp cho mọi người dễ dàng tìm kiếm các nhà vệ sinh công cộng của Đan Mạch.
- Một dịch vụ có tên vervuilingsalarm.nl tại Hà Lan sẵn sàng cung cấp các cảnh báo về chất lượng không khí xấu nếu vượt ngưỡng cho phép quanh khu vực lân cận chúng ta sống sử dụng dữ liệu mở.
- Và rất nhiều ví dụ khác



Chúng ta có thể tìm và khai thác dữ liệu mở ở đâu?



Trên thế giới

- Cổng dữ liệu các nước. (thường tại địa chỉ `data.gov.[xx]` trong đó xx là tên miền nước đó).
- Cổng dữ liệu Châu Á: `dataportal.asia`
- Cổng dữ liệu của Ủy hội sông Mêlong: `portal.mrcmekong.org`
- Và nhiều cổng khác có thể dễ dàng tìm kiếm bằng từ khóa `opendata portal`.

Tại Việt Nam

o Dữ liệu mở của cơ quan nhà nước

- Cổng dữ liệu quốc gia `data.gov.vn`
- Cổng dữ liệu mở của các bộ, ngành, địa phương thông thường có địa chỉ tại “`data.[ten].gov.vn`” hoặc “`opendata.[ten].gov.vn`” trong đó [ten] là tên của bộ, ngành, tỉnh, thành đó.
- Trên cổng thông tin điện tử của các bộ, ngành, địa phương nếu trên cổng có mục cung cấp dữ liệu mở.

o Cổng dữ liệu Itrithuc: `dulieu.itrithuc.vn` và một số tổ chức khác cũng đang xúc tiến cung cấp dữ liệu mở.

Khi sử dụng dữ liệu mở phải tuân thủ yêu cầu gì?



Phải trích dẫn nguồn gốc

Ghi nhận thông tin sử dụng dữ liệu mở trong các sản phẩm, tài liệu liên quan có sử dụng dữ liệu mở. Điều đó có nghĩa, bạn phải ghi nguồn gốc cơ quan nhà nước cung cấp, địa chỉ trang web của cơ quan nhà nước nơi bạn khai thác nơi bạn sử dụng dữ liệu mở để bên thứ ba biết được điều đó.

Không được bán

Không được bán lại dữ liệu mở đã được khai thác nguyên trạng cho tổ chức, cá nhân khác. Khi sử dụng dữ liệu mở trong sản phẩm, dịch vụ thương mại của mình phải cung cấp miễn phí dữ liệu mở kèm theo sản phẩm, dịch vụ thương mại đó.

Tuân thủ quy định/giấy phép sử dụng

Dữ liệu mở có điều khoản sử dụng đi kèm gọi là giấy phép sử dụng. Bạn phải tuân thủ các điều khoản của giấy phép khi bạn sử dụng dữ liệu đó.

Giấy phép sử dụng dữ liệu mở là gì?



Mặc dù dữ liệu mở rất tự do trong sử dụng, chia sẻ nhưng vẫn bị ràng buộc một số điều kiện nhất định. Những ràng buộc này được gọi là giấy phép sử dụng dữ liệu mở.

Tư tưởng của giấy phép muốn nhắc nhở bạn sử dụng dữ liệu mở một cách công bằng, minh bạch và có trách nhiệm đối với người cung cấp cho bạn và người được bạn chia sẻ sử dụng lại dữ liệu mở.

Có một số giấy phép nổi tiếng được nhiều người chấp nhận như Creative Commons và ngay cả bạn cũng có thể tạo ra một giấy phép riêng cho mình khi cung cấp dữ liệu mở nhưng đều theo tư tưởng trên.

Đối với dữ liệu mở của cơ quan nhà nước: Quy định Yêu cầu sử dụng dữ liệu mở của cơ quan nhà nước tại Điều 18 Nghị định 47/2020/NĐ-CP cũng có thể coi là một dạng giấy phép sử dụng dữ liệu mở



Các cơ quan nhà nước phải làm gì?

- Xây dựng kế hoạch cung cấp dữ liệu mở, rà soát điều chỉnh 6 tháng một lần.
- Cung cấp dữ liệu mở trên cổng dữ liệu.
- Tiếp nhận ý kiến phản hồi của cộng đồng để nâng cao chất lượng dữ liệu mở.
- Khuyến khích tổ chức, cá nhân sử dụng, đóng góp dữ liệu mở.

Quy định trách nhiệm của cơ quan nhà nước cung cấp dữ liệu mở tại Nghị định số 47/2020/NĐ-CP ngày 09/4/2020

Xây dựng kế hoạch cung cấp dữ liệu mở thực hiện theo hướng dẫn của Bộ TTTT tại văn bản 5752/BTTTT-CĐSQG ngày 26/11/2022

Dữ liệu gì nên là dữ liệu mở ?

- Tất cả các dữ liệu khi chia sẻ không vi phạm: bí mật nhà nước, bí mật kinh doanh, bảo vệ thông tin cá nhân, bảo vệ quyền sở hữu trí tuệ và bị giới hạn chia sẻ theo pháp luật đều có thể được xem xét để cung cấp dưới dạng dữ liệu mở.
- Dữ liệu quan trắc, dữ liệu giám sát công cộng, danh mục dùng chung, dữ liệu cảm biến, dữ liệu thời tiết, dữ liệu hành vi đã được bỏ định danh...là những dữ liệu thường được quan tâm để cung cấp dưới dạng dữ liệu mở nhiều nhất.
- Dữ liệu mà cộng đồng cần để nghiên cứu khoa học và phát huy tính sáng tạo.

Danh mục dữ liệu gợi ý tại Phụ lục văn bản 5752/BTTTT-CDSQG ngày 26/11/2022 của Bộ Thông tin và Truyền thông





Nguyên tắc cung cấp dữ liệu mở

- Dữ liệu phải toàn vẹn và phản ánh đầy đủ nội dung thông tin cung cấp
- Dữ liệu được cập nhật mới nhất
- Dữ liệu phải có khả năng truy cập sử dụng được trên môi trường mạng
- Đảm bảo khả năng thiết bị số có thể gửi, nhận, lưu trữ và xử lý được dữ liệu
- Tự do truy cập sử dụng dữ liệu mở, không yêu cầu khai báo định danh khi khai thác, sử dụng
- Dữ liệu ở định dạng mở, thường là định dạng có cấu trúc như xls; json, xml, cvs...
- Dữ liệu được sử dụng miễn phí
- Ưu tiên cung cấp dữ liệu mở có nhu cầu sử dụng cao.



Một số thuật ngữ

Cổng dữ liệu mở

Là một nền tảng số trực tuyến hỗ trợ các cơ quan, tổ chức công bố, chia sẻ dữ liệu mở; hỗ trợ người sử dụng truy cập các bộ dữ liệu mở. Nền tảng số này bao gồm danh mục dữ liệu cùng với các thông tin và dịch vụ khác tạo thành một hệ sinh thái dữ liệu mở (như diễn đàn trao đổi, tương tác với người dùng, hỗ trợ việc sử dụng...)

Tập dữ liệu (Dataset)

Một tập hợp dữ liệu có cùng dữ liệu đặc tả được cơ quan, tổ chức cung cấp dưới dạng một tệp dữ liệu hoặc thông điệp dữ liệu qua dịch vụ chia sẻ dữ liệu dữ liệu

Dữ liệu đặc tả (Metadata)

Là những thông tin mô tả dữ liệu về tên, nguồn gốc, nội dung sơ bộ, chất lượng, định dạng, thời điểm thu thập, thời điểm công bố, cập nhật và các thông tin khác nhằm tạo thuận lợi cho quá trình tìm kiếm, truy cập, quản lý, lưu trữ và trao đổi tập dữ liệu mở.

Cùng chung tay thúc đẩy dữ liệu mở Việt Nam

"Ẩn sau một khối lượng lớn dữ liệu là tri thức
mà chúng ta có thể cứu sống một bệnh nhân,
hoặc là thay đổi cả thế giới."

Atul Butte, Đại học Y khoa Stanford

