



SECURITY BOOTCAMP 2022

HẠ LONG, QUẢNG NINH 9-11/09/2022

DỮ LIỆU MỞ VÀ VIỆC BẢO VỆ DỮ LIỆU CÁ NHÂN

LÊ TRUNG NGHĨA

HIỆP HỘI CÁC TRƯỜNG ĐẠI HỌC, CAO ĐẲNG VIỆT NAM

Email: letrungnghia.foss@gmail.com

Blogs: <http://vnfoss.blogspot.com/>

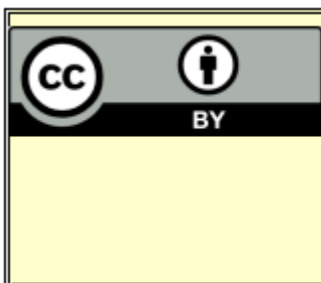
<http://letrungnghia.mangvn.org/>

Trang web CLB PMTDNM Việt Nam:

<http://vfossa.vn/vi/>

Nội dung

1. Về Nghị định Bảo vệ dữ liệu cá nhân
2. Dữ liệu mở, dữ liệu cá nhân và ẩn danh
3. Kỹ thuật ẩn danh
4. Vài phần mềm nguồn mở (PMNM) đáp ứng GDPR
5. Vài gợi ý

	<p>Tài liệu này mang giấy phép Creative Commons Attribution v4.0. Bạn được trao các quyền để sử dụng, chia sẻ, sao chép, phân phối, phân phối lại, áp dụng, pha trộn, tùy biến và xây dựng dựa trên các tư liệu của nó, kể cả thương mại hóa, miễn là bạn ghi nhận công (các) tác giả gốc ban đầu của tài liệu. Một bản sao giấy phép này có tại: http://creativecommons.org/licenses/by/4.0/.</p>
--	--



ANTV is funded in whole or in part by the Vietnamese government. [Wikipedia](#)

Hội thảo "Chính sách bảo vệ dữ liệu cá nhân trên không gian mạng" | ANTV

1,309 views Apr 20, 2019 ANTV | Để các yêu cầu của Luật an ninh mạng đáp ứng yêu cầu thực tiễn, tạo hành lang pháp lý bảo vệ người dùng khi sử dụng các dịch vụ trên không gian mạ ...more



ANTV is funded in whole or in part by the Vietnamese government. [Wikipedia](#)

#antv #truyenhinhcongannhandan

Cần Xây Dựng Nghị Định Bảo Vệ Dữ Liệu Cá Nhân | Cư Dân Mạng | ANTV

1,037 views Premiered Aug 11, 2022 ANTV | Cư dân mạng | Trong thời gian qua trước nhì ...more

- Hoàn thiện pháp luật về bảo vệ dữ liệu cá nhân của ThS. Bạch Thị Nhã Nam, 13/05/2022



**NGHIÊN CỨU
LẬP PHÁP**
VIỆN NGHIÊN CỨU LẬP PHÁP THUỘC ỦY BAN THƯỜNG VỤ QUỐC HỘI

- Nghị quyết số 27/NQ-CP ngày 07/3/2022 của Chính phủ thông qua hồ sơ xây dựng Nghị định bảo vệ dữ liệu cá nhân

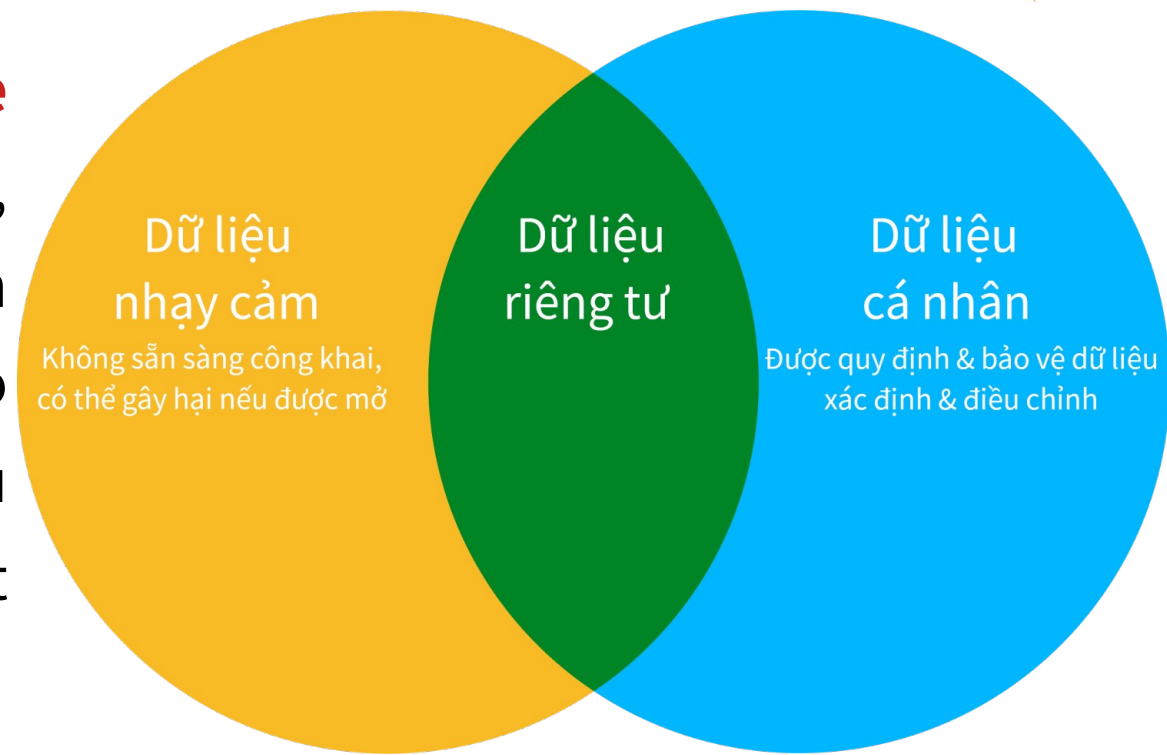
→ **Xây dựng Nghị định bảo vệ dữ liệu cá nhân là rất cần thiết!**

- **Dữ liệu Mở (ODI)**: (1) Dữ liệu bất kỳ ai cũng có thể sử dụng và chia sẻ; (2) Ở định dạng **máy đọc được**; (3) Phải được cấp phép mở, phải **cho phép mọi người sử dụng dữ liệu đó theo bất kỳ cách gì họ muốn**, bao gồm biến đổi, kết hợp và chia sẻ nó với những người khác, ngay **cả với mục đích thương mại**.
- **Dữ liệu cá nhân (Personal Data) (GDPR)**: Bất kỳ thông tin nào liên quan đến việc xác định hoặc có thể xác định một thể nhân (‘chủ thể dữ liệu’). Một thể nhân là một cá nhân cụ thể **có thể được nhận dạng trực tiếp hoặc gián tiếp thông qua một mã định dạng** chẳng hạn như tên, số căn cước công dân, dữ liệu về địa điểm, mã định danh số hoặc **thông qua một hoặc vài yếu tố đặc thù khác** về về thể chất, sinh lý, di truyền, tinh thần, kinh tế, văn hóa hoặc xã hội của cá nhân đó.
- **GDPR** (General Data Protection Regulation): Quy định Bảo vệ Dữ liệu Chung

Hỏi: Có hay không dữ liệu mở có nguồn gốc từ dữ liệu cá nhân?

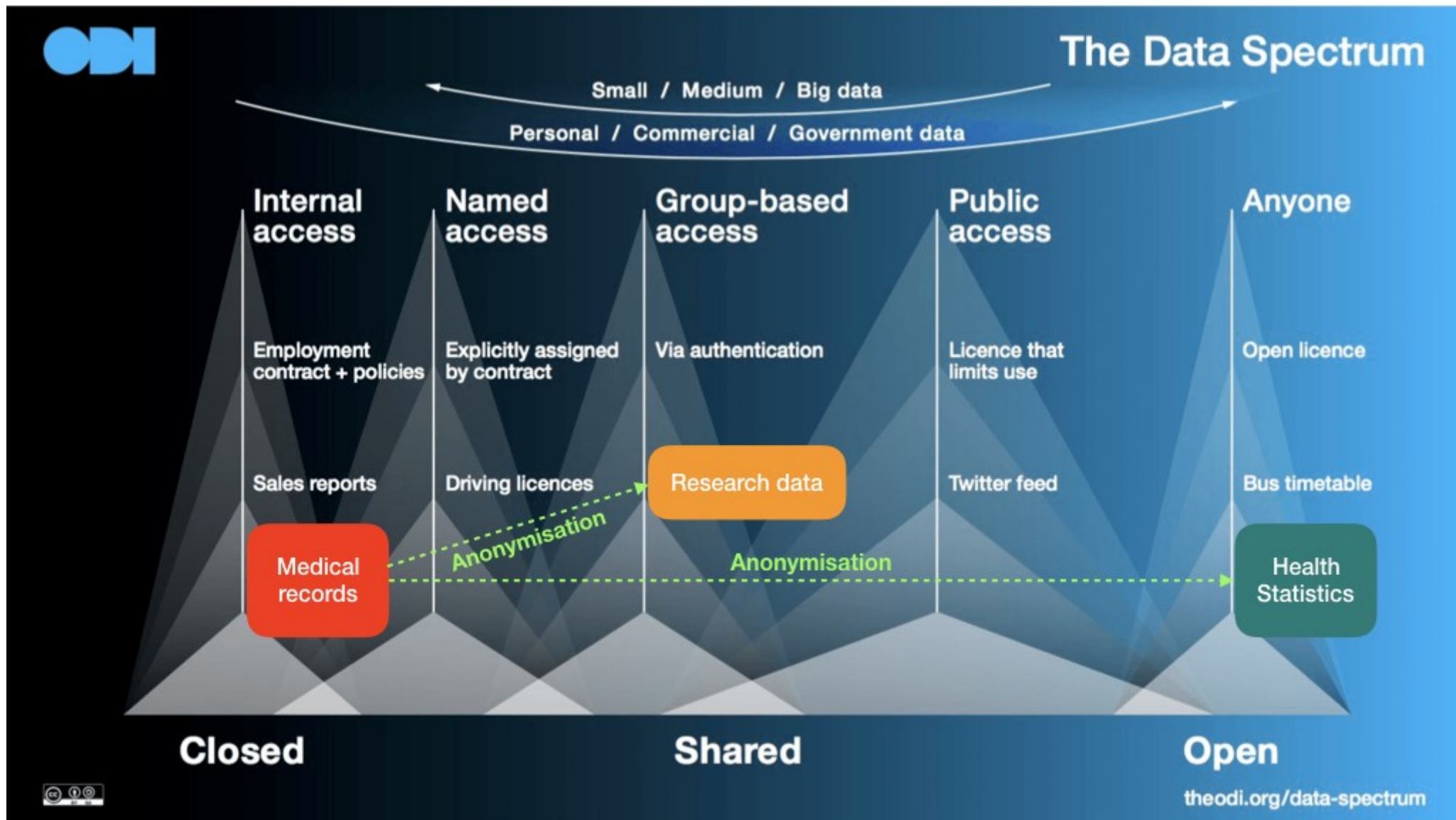
Dữ liệu mở, dữ liệu cá nhân và ẩn danh -2

- **Dữ liệu nhạy cảm (Sensitive Data)**: là dữ liệu có thông tin mật, thông tin lưu hành nội bộ của đơn vị hoặc do đơn vị quản lý, nếu lộ lọt ra ngoài sẽ gây ảnh hưởng xấu đến danh tiếng, tài chính và hoạt động của đơn vị[01] → Cá nhân.



- **Thông tin riêng tư (Private Information)**[02]: là thông tin liên quan tới một pháp nhân hoặc thể nhân có thể nhận dạng được; không là trong phạm vi công cộng hoặc kiến thức chung; và nếu để lộ có thể gây cho họ thiệt hại, tổn thất hoặc đau khổ cho họ. Định nghĩa này là rất giống với những gì GDPR gọi là dữ liệu danh mục đặc biệt[03].

- **Ẩn danh (Anonimisation)**: là quy trình sửa một tập hợp dữ liệu để làm giảm rủi ro tái nhận dạng nó càng nhiều càng tốt[04, tr. 8].



Viện Dữ liệu Mở - ODI (Open Data Institute): Ví dụ về Dữ liệu đóng, dữ liệu chia sẻ, và dữ liệu mở với kỹ thuật ẩn danh theo Phổ Dữ liệu[05, tr. 9].

Kỹ thuật ẩn danh -1

Vài kỹ thuật ẩn danh: (1) Ngăn chặn - Loại bỏ thông tin có thể nhận dạng;
(2) Thường hóa - Tạo mức chi tiết thô hơn để dấu đi dữ liệu của từng cá nhân;
(3) Làm gián đoạn - Tạo nhiễu để không suy luận ra thông tin từng cá nhân.

Họ và tên	Ngày sinh	Số buổi làm việc trực tuyến trong tuần
Nguyễn Văn A	16/03/1969	5
Trần Thị H	02/12/1985	1
...
Phạm Thị T	26/05/2000	2

Họ và tên	Ngày sinh	Số buổi làm việc trực tuyến trong tuần
1	16/03/1969	5
2	02/12/1985	1
...
753	26/05/2000	2

Ví dụ: Số buổi làm việc trực tuyến theo tuần? → (1) Bỏ họ tên, thay bằng số thứ tự

ID	Độ tuổi đến tháng 2/2019	Số buổi làm việc trực tuyến trong tuần
1	40-50	5
2	30-40	1
...
753	10-20	2

ID	Độ tuổi đến tháng 2/2019	Số buổi làm việc trực tuyến trong tuần (tối 10% các giá trị đã được hoán đổi giữa các hàng)
1	40-50	1
2	30-40	5
...
753	10-20	2

→ (2) Bỏ ngày sinh, thay bằng độ tuổi

→ (3) Hoán đổi các giá trị với tỷ lệ nào đó

“Dữ liệu có thể hoặc hữu dụng hoặc được ẩn danh tuyệt vời, nhưng không bao giờ có cả hai” → RỦI RO = 0 THÌ LỢI ÍCH CŨNG = 0

Số buổi làm việc trực tuyến trong tuần (được ẩn danh)
Vài người đôi lúc làm việc trực tuyến.



Kỹ thuật ẩn danh -2

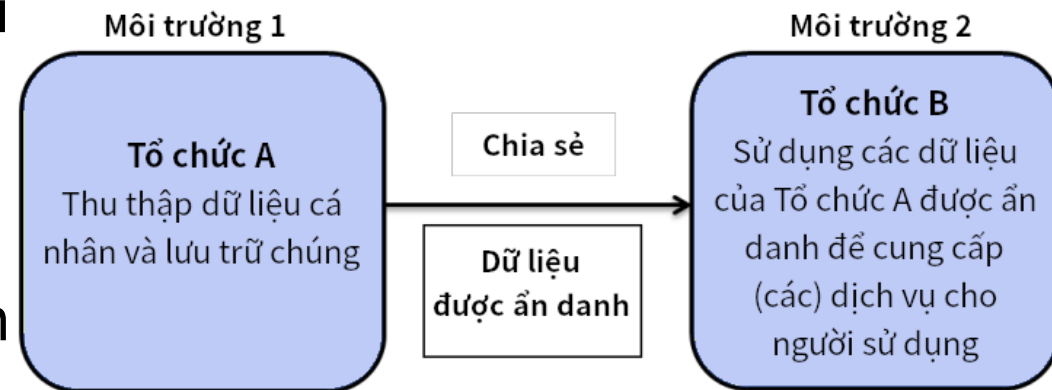
Ẩn danh là một quy trình gồm nhiều hành động. Để ẩn danh hiệu quả, quy trình đó nên bao gồm suy nghĩ nghiêm túc về **điều gì xảy ra với dữ liệu sau khi ẩn danh**; dữ liệu sẽ được chia sẻ với ai; liệu có những sai lầm nào có thể diễn ra và liệu có biện pháp giảm nhẹ nào có thể tránh được điều đó không?

Ví dụ về dòng chảy của dữ liệu và vai trò của bạn

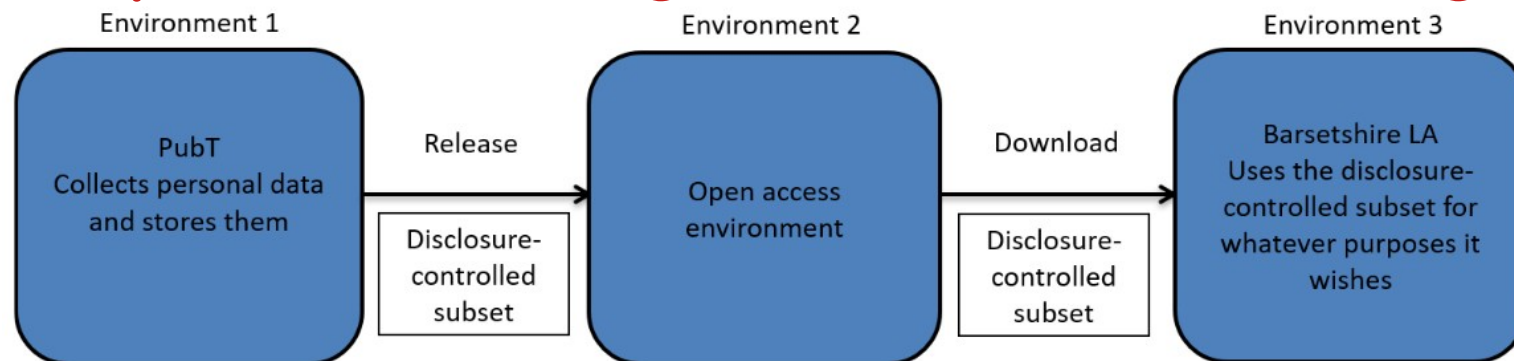
Bạn là: (1) Người kiểm soát; (2) Người xử lý; hay (3) Người sử dụng dữ liệu?

Gốc dữ liệu: từ đâu tới và đi về đâu?

Tình trạng dữ liệu: dữ liệu cá nhân hay thông tin ẩn danh - với từng bên?



Ví dụ về các môi trường phát hành mở chia sẻ đơn giản

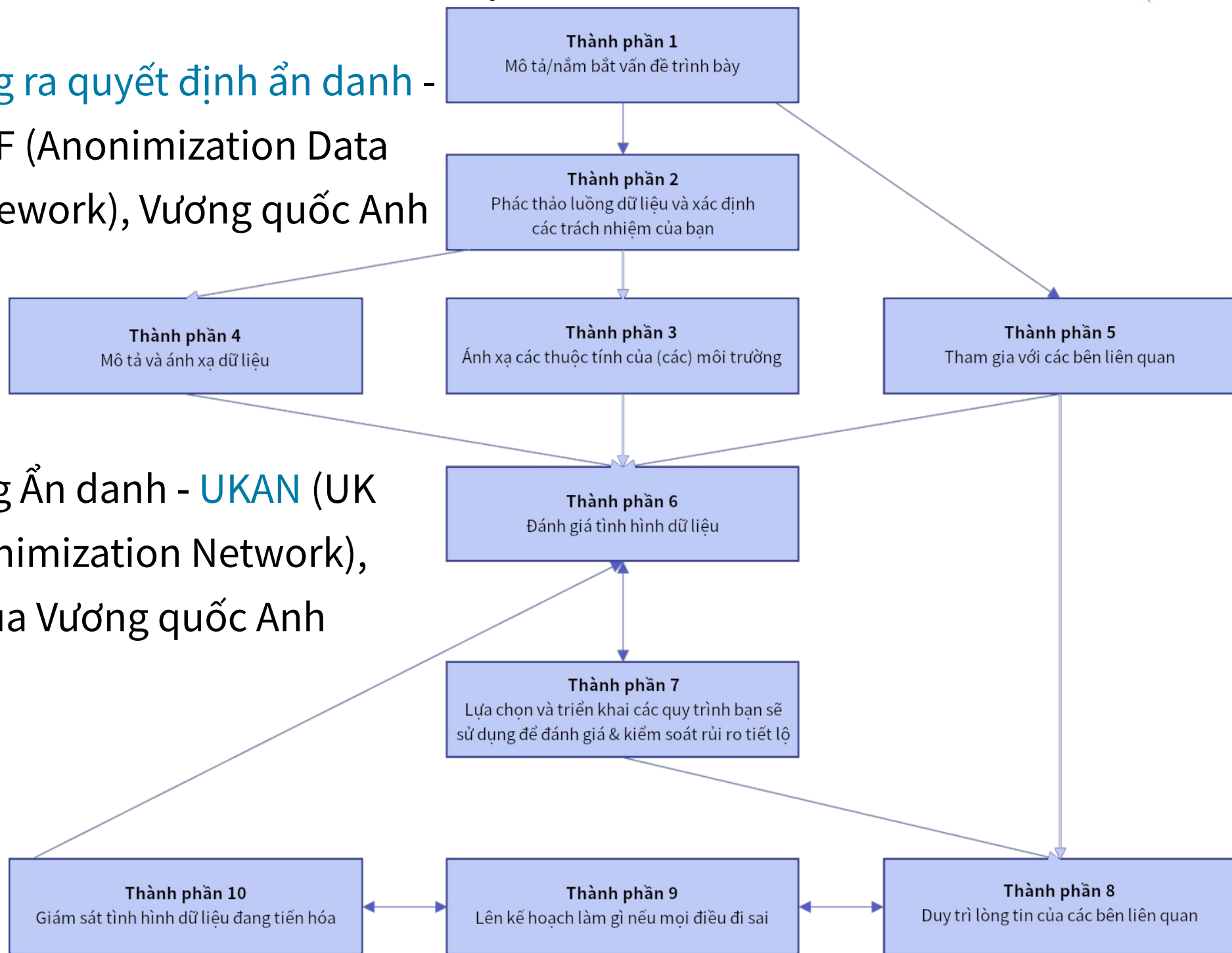


Kỹ thuật ẩn danh -3

Khung ra quyết định ẩn danh -

ADF (Anonimization Data Framework), Vương quốc Anh

Mạng Ẩn danh - **UKAN** (UK Anonimization Network), của Vương quốc Anh



Vài phần mềm nguồn mở đáp ứng GDPR

Bảng 1: 17 dự án nguồn mở sẵn sàng với GDPR cho Doanh nghiệp (ERP, CRM, CMS, CHAT, Đám mây, Phân tích)[9]

PMNM	Magento	EspoCRM	SuiteCRM	Vtiger CRM	NextCloud
Giấy phép mở	OSL v3, AFLv3	GPLv3	AGPLv3	SUGARCRM PL v1.1.2	AGPLv3
PMNM	OwnCloud	RocketChat	ERPNext	Axelor ERP	Dolibarr ERP/ CRM
Giấy phép mở	AGPLv3	MIT	GPL	AGPLv3	GPLv3+
PMNM	Matomo	OWA	GrandNode	0 A.D.	Wordpress
Giấy phép mở	AGPLv3	MIT	GPLv3	GPLv2	GPLv2+
PMNM	Zenario CMS	Jahia CMS			
Giấy phép mở	BSD	GPLv3			

Bảng 2: 5 công cụ phân tích tuân thủ với GDPR là PMNM[10]

PMNM	PostHog	Plausible	Countly	GoAccess	Matomo
Giấy phép mở	MIT	AGPLv3	AGPLv3	MIT	AGPLv3

Vài phần mềm nguồn mở và công cụ phân tích đáp ứng GDPR là PMNM

1. Câu hỏi: Có hay không dữ liệu mở có nguồn gốc từ các dữ liệu cá nhân?

→ Câu trả lời: Chắc chắn CÓ! bằng việc sử dụng kỹ thuật ẩn danh!

2. Ẩn danh là việc đánh đổi và cân nhắc giữa LỢI ÍCH & RỦI RO!

RỦI RO = 0 THÌ LỢI ÍCH CŨNG = 0!

3. Ẩn danh là phức tạp, không chỉ là khía cạnh công nghệ dữ liệu, mà còn phụ thuộc vào nhiều yếu tố khác, như môi trường & pháp luật từng quốc gia!

4. Tuân thủ các hiệp định thương mại với nước ngoài, các công ty Việt Nam cần chú ý các quy định bảo vệ dữ liệu riêng tư, như GDPR với EVFTA, để tránh bị phạt (2% doanh thu toàn cầu của công ty) → ví dụ, **học cách thống kê!**

5. Nghị định bảo vệ quyền riêng tư nên cân nhắc tất cả các khía cạnh để tránh các rủi ro đồng thời mang lại nhiều lợi ích cho quốc gia, doanh nghiệp!
→ tham khảo kinh nghiệm của nước ngoài, ví dụ, của Vương quốc Anh.

Xem thêm: Các bài viết liên quan: [01](#), [02](#), [03](#) và các đường liên kết ở đó.

Cảm ơn!

LOD.VN: <https://www.facebook.com/groups/511059912615187/>
oervn@googlegroups.com
oer-vn@googlegroups.com

Hỏi đáp

LÊ TRUNG NGHĨA

Email: letrungnghia.foss@gmail.com

Blogs: <http://vnfoss.blogspot.com/>

<http://letrungnghia.mangvn.org/>

Trang web CLB PMTDNM Việt Nam: <http://vfossa.vn/vi/>