

Chính sách mã nhận diện thường trực (PID) cho Đám mây Khoa học Mở châu Âu (EOSC)

Báo cáo từ các Nhóm Làm việc về Kiến trúc và FAIR của Đám mây Khoa học Mở châu Âu

Dịch sang tiếng Việt: Lê Trung Nghĩa

Dịch xong: 03/07/2023

Bản gốc tiếng Anh: <https://op.europa.eu/da/publication-detail/-/publication/35c5ca10-1417-11eb-b57e-01aa75ed71a1/language-en>

A Persistent Identifier (PID) policy for the European Open Science Cloud

Report from the European Open Science Cloud FAIR and Architecture Working Groups

Ban Điều hành EOSC
Nhóm làm việc về FAIR và Kiến trúc
Tháng 10/2020
Nghiên cứu & Đổi mới sáng tạo

EOSC Executive Board
WG FAIR and Architecture
October 2020
Research and Innovation

Chính sách mã nhận diện thường trực (PID) cho Đám mây Khoa học Mở châu Âu (EOSC)

Ủy ban châu Âu

Ban Tổng giám đốc về Nghiên cứu và Đổi mới

Ban G - Tiếp cận Nghiên cứu và Đổi mới

Đơn vị G.4 - Khoa học Mở

Liên hệ Corina Pascu

Thư điện tử Corina.PASCU@ec.europa.eu

RTD-EOSC@ec.europa.eu

RTD-PUBLICATIONS@ec.europa.eu

Ủy ban châu Âu

B-1049 Brussels

Bản thảo hoàn thành vào tháng 10/2020.

Ủy ban châu Âu không chịu trách nhiệm về bất kỳ hệ quả nào bắt nguồn từ việc sử dụng lại xuất bản phẩm này.

Các quan điểm được thể hiện trong xuất bản phẩm này là trách nhiệm chỉ của tác giả và không nhất thiết phản ánh quan điểm của Ủy ban châu Âu.

Nhiều thông tin hơn về Liên minh châu Âu là sẵn có trên Internet (<http://europa.eu>).

PDF ISBN 978-92-76-22780-9 doi: 10.2777/926037 KI-04-20-576-EN-N

Luxembourg: Văn phòng Xuất bản của Liên minh châu Âu, 2020.

© European Union, 2020



Chính sách sử dụng lại các tài liệu của Ủy ban châu Âu được triển khai dựa vào Quyết định của Ủy ban 2011/833/EU ngày 12/12/2011 về sử dụng lại các tài liệu của Ủy ban (OJ L 330, 14.12.2011, tr. 39). Ngoại trừ được lưu ý khác, sử dụng lại tài liệu này là được phép theo giấy phép Creative Commons Ghi công 4.0 Quốc tế (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>). Điều này có nghĩa là sử dụng lại là được phép miễn là thừa nhận ghi công đúng cách được đưa ra và bất kỳ thay đổi nào cũng sẽ được chỉ ra.

Sử dụng hoặc tái tạo lại bất kỳ hình ảnh hoặc các tư liệu khác nào không phải là bản quyền của Ủy ban châu Âu, thì sự cho phép phải được tìm trực tiếp từ những người nắm giữ bản quyền.



The reuse policy of European Commission documents is implemented based on Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of photos or other material that is not under the copyright of the European Union, permission must be sought directly from the copyright holders.

Thừa nhận ghi công hình ảnh:

Trang bìa: © Lonely #46246900, ag visuell #16440826, Sean Gladwell #6018533, LwRedStorm #3348265, 2011; kras99 #43746830, 2012. Nguồn: Fotolia.com.

Chính sách mã nhận diện thường trực (PID) cho Đám mây Khoa học Mở châu Âu

Báo cáo từ các Nhóm Làm việc về Kiến trúc và FAIR của Đám mây Khoa học Mở châu Âu

Biên tập: Ban Điều hành EOSC

Tháng 10/2020

Ban Tổng giám đốc về Nghiên cứu và Đổi mới

Ban Điều hành EOSC

Mục lục

TÓM TẮT	7
1. BỐI CẢNH	8
2. CÁC NGUYÊN TẮC	8
3. ĐỊNH NGHĨA CHUNG CỦA PID	10
4. CÁC VAI TRÒ VÀ TRÁCH NHIỆM	13
5. CÁC ỨNG DỤNG PID	14
6. CÁC DẠNG PID	16
7. CÁC DỊCH VỤ PID VÀ CÁC NHÀ CUNG CẤP DỊCH VỤ PID	16
8. ĐIỀU HÀNH VÀ TÍNH BỀN VỮNG	17
PHỤ LỤC 1: DANH SÁCH CÁC NGUỒN LỰC ĐƯỢC TƯ VẤN	18
PHỤ LỤC 2: BẢNG CHÚ GIẢI	20
PHỤ LỤC 3: CÁC TÁC GIẢ CỦA CHÍNH SÁCH	22
PHỤ LỤC 4: CÁC PHIÊN BẢN VÀ NHỮNG THAY ĐỔI ĐƯỢC THEO DÕI	23

BAN ĐIỀU HÀNH EOSC, THÁNG 10 NĂM 2020

Karel LUYBEN: Executive Board Co-Chair, CESAER Representative

Cathrin STÖVER: Executive Board Co-Chair, GÉANT Representative

Jean-François ABRAMATIC: Architecture WG chair, Individual Expert

Juan BICARREGUI: Rules of Participation (RoP) WG Chair, Research Data Alliance (RDA) Representative

Jan HRUŠÁK: Landscape WG Chair, Individual Expert

Sarah JONES: FAIR WG Chair, Individual Expert

Rupert LÜCK: Sustainability WG Co-Chair, EBML Representative

Lidia BORRELL-DAMIAN: Sustainability WG Co-Chair, Science Europe Representative

Natalia MANOLA: Skills and Training WG Chair, OpenAIRE Representative

John WOMERSLEY: European Spallation Source ERIC Representative

Ron DEKKER: CESSDA Representative

Jeannette FREY: Observer, LIBER Representative

TÓM TẮT

Chính sách Mã nhận diện Thường trực - PID (Persistent Identifier) này được viết cho các nhà hoạch định chính sách trong các nhà cung cấp hạ tầng và dịch vụ EOSC tiềm năng, và sẽ là mối quan tâm của tất cả các bên liên quan của Đám mây Khoa học Mở châu Âu - EOSC (European Open Science Cloud). Nó xác định một tập hợp các kỳ vọng về các mã nhận diện thường trực nào sẽ được sử dụng để hỗ trợ cho việc vận hành môi trường nghiên cứu Tìm thấy được, Truy cập được, Tương hợp được, Sử dụng lại được - FAIR (Findable, Accessible, Interoperable, Reusable). Các yêu cầu của các nhà cung cấp và các dịch vụ cơ bản họ cung cấp cũng được nêu. Chính sách này được điều hành EOSC phê chuẩn, bên sẽ giám sát việc triển khai của nó. Việc triển khai sẽ được hướng dẫn thông qua các khuyến nghị về Kiến trúc Kỹ thuật của PID mà Nhóm Làm việc về Kiến trúc của EOSC sẽ cung cấp.

Tác giả của chính sách này là các đại diện của Nhóm Làm việc về FAIR của EOSC và các nhóm Làm việc về Kiến trúc EOSC (Maggie Hellström, André Heughebaert, Rachael Kotarski, Paolo Manghi, Brian Matthews, Raphael Ritz, Anders Sparre Conrad, Mario Valle, Tobias Weigel và Peter Wittenburg). Các bản thảo ban đầu đã được phát hành vào tháng 12/2019 và tháng 5/2020, và được làm cho sẵn sàng cho bình luận và tư vấn công khai.

Chính sách hoàn thành này đã được cung cấp cho Ban Điều hành và Lãnh đạo EOSC vào tháng 10/2020.

1. BỐI CẢNH

1.1 Chính sách mã nhận diện thường trực (PID) này được viết cho các nhà hoạch định chính sách cấp cao trong các nhà cung cấp hạ tầng và dịch vụ EOSC tiềm năng, và sẽ là mối quan tâm của tất cả các bên liên quan của EOSC. Nó xác định một tập hợp các kỳ vọng về các mã nhận diện thường trực nào sẽ được sử dụng để hỗ trợ cho việc vận hành môi trường nghiên cứu FAIR. Các yêu cầu của nhà cung cấp và các dịch vụ cơ bản họ chào cũng được nêu. Chính sách được điều hành EOSC phê chuẩn, bên sẽ giám sát triển khai của nó. Sự triển khai sẽ được hướng dẫn thông qua các khuyến nghị về Kiến trúc Kỹ thuật PID mà Nhóm Làm việc về Kiến trúc EOSC sẽ cung cấp.

1.2 Nhiều dạng mã nhận diện thường trực đang được sử dụng tích cực rồi cho một loạt các ứng dụng. Ngoài ra, vô số các công nghệ, cơ sở dữ liệu và các hệ thống đám mây sử dụng các mã nhận diện cục bộ địa phương. Chính sách PID này sẽ không cung cấp các tuyên bố chi tiết về các hệ thống đang tồn tại, mà cung cấp các định nghĩa, hướng dẫn và các yêu cầu sử dụng.

1.3 Quy định GDPR đang hướng dẫn chúng ta trong sử dụng các PID lưu trữ và cung cấp dữ liệu cá nhân. Các nhà cung cấp dịch vụ PID cần tuân thủ với GDPR. Dữ liệu các trợ cấp GDPR tuân thủ các quyền nhất định, bao gồm quyền yêu cầu các tổ chức xóa bỏ dữ liệu cá nhân của họ (quyền xóa). Vì các PID được các thực thể khác nhau (các kho, các tổ chức thực thi nghiên cứu, các hạ tầng nghiên cứu, EOSC, .v.v.) sử dụng, các chủ thể dữ liệu sẽ phải thực thi các quyền theo GDPR (như quyền xóa) trên cơ sở từng trường hợp một. Trên thực tế, các chủ thể dữ liệu có thể khó theo dõi nơi dữ liệu cá nhân của họ được xử lý.

2. CÁC NGUYÊN TẮC

2.1 Chính sách PID tập trung vào các nguyên tắc, các kết quả và điều hành được kỳ vọng được thiết kế để thiết lập một hạ tầng PID cường tráng, tin cậy phù hợp cho tính bền vững lâu dài của EOSC.

2.2 Có một dải các trường hợp sử dụng bao trùm sự đa dạng rộng lớn sử dụng trực tiếp của con người và việc xử lý và khả năng hành động của máy được tự động hóa. Chính

sách PID này sẽ dàn xếp một dải rộng lớn các trường hợp sử dụng và không đặt ra các rào cản để sử dụng các PID hiệu quả đáp ứng nhu cầu của cộng đồng nghiên cứu.

2.3 Chính sách sẽ có kết quả trong tương lai nơi các PID có thể được sử dụng như một phương pháp được ưu tiên để cung cấp cho thực thể được nó gắn PID, ở những nơi phù hợp, cùng với phương tiện con người đọc được, ví dụ, tên chung. Nhiều PID có thể nhận diện bất kỳ thực thể nhất định nào và những người sử dụng sẽ có khả năng sử dụng bất kỳ thứ gì họ thấy thuận tiện nhất (but cf.3.6).

2.4 Chính sách PID sẽ xúc tác cho môi trường thực hành nghiên cứu, và các dịch vụ thỏa mãn các nguyên tắc FAIR một cách thích hợp cho các lĩnh vực sử dụng đặc biệt. Trọng tâm đối với việc hiện thực hóa FAIR là các Đối tượng Số FAIR và các PID là cốt lõi cho ý tưởng của các Đối tượng Số FAIR, như được nêu bật trong báo cáo Biến FAIR Thành Hiện thực (Nhóm Chuyên gia FAIR, 2018).

2.5 Để tạo ra một hệ sinh thái PID hoạt động được và để hiện thực hóa các Nguyên tắc FAIR có liên quan, các Dịch vụ PID sẽ phải tương hợp với nhau và giữa các nhà cung cấp dịch vụ PID, cũng như xuyên suốt hạ tầng nghiên cứu.

2.6 Độc lập công nghệ của các PID được yêu cầu để cho phép thay đổi công nghệ. Các dịch vụ PID sẽ biến động về độ chín qua thời gian và chính sách PID sẽ nhận diện mức độ chín của các dịch vụ phù hợp cho việc áp dụng EOSC.

2.7 Chính sách tìm cách dàn xếp độ chín và thực hành, lược đồ, công nghệ và nhà cung cấp PID được thiết lập mà có sự hiện diện toàn cầu. Chính sách cũng cần phải cân bằng và không ưu tiên bất kỳ một cách tiếp cận hay công nghệ nào.

2.8 Để gắn với Chính sách này, các nhà cung cấp và người sử dụng PID được kỳ vọng cam kết duy trì bền vững các hạ tầng PID và các đối tượng số về lâu dài.

2.9 Chính sách nên khuyến khích các dịch vụ và công cụ mới và đổi mới sáng tạo, được sử dụng và xây dựng trên hạ tầng PID.

2.10 Chính sách nên được một Thực thể Hợp pháp của EOSC duy trì và được nhóm các chuyên gia và các bên liên quan chuyên tâm rà soát lại trên cơ sở thường xuyên và ít nhất trong thời gian 3 năm. Tất cả những ai thực hiện các Vai trò được xác định trong Phần 4 sẽ được tư vấn và tham gia trong quy trình rà soát lại đó.

3. ĐỊNH NGHĨA CHUNG CỦA PID

3.1 Vì các mục đích của Chính sách này, một mã nhận diện thường trực hỗ trợ và xúc tác cho nghiên cứu tuân thủ FAIR là mã duy nhất toàn cầu, thường trực và phân giải được. Từng trong số 3 đặc tính đó có thể được cung cấp thông qua các phương tiện và công nghệ khác nhau, nhưng áp dụng các nguyên tắc cơ bản y hệt nhau.

3.2 Duy nhất toàn cầu

3.2.1 Để xúc tác cho tính duy nhất toàn cầu, một PID cần tuân thủ với cú pháp được kiểm soát để tránh xung đột, ví dụ thế, bằng việc có không gian tên do các cơ quan quản lý được xác định rõ ràng điều hành quản lý.

3.3 Thường trực

3.3.1 Thường trực liên quan tới 3 khía cạnh. Hai khía cạnh đầu liên quan trực tiếp tới bản thân PID.

3.3.1.1 PID nên được quản lý và điều hành theo một cách thức cộng đồng có thể tin tưởng nó để giữ được là duy nhất và phân giải được về lâu dài. Điều này có thể vượt ra khỏi cuộc đời của thực thể nó nhận diện, của những người tạo lập ra thực thể đó hoặc thậm chí cả bản thân nhà cung cấp dịch vụ PID đó. Điểm sau phụ thuộc vào điều 8.4.

3.3.1.2 Cú pháp - bao gồm giao diện lập trình ứng dụng - API (Application Programming Interface) - của hệ thống PID cũng sẽ là thường trực.

3.3.1.3 Cuối cùng, thường trực cũng liên quan tới đối tượng theo đó PID phân giải (đối tượng được tham chiếu). Đối tượng được tham chiếu cũng sẽ là ổn định, dù nó là đối tượng số bẩm sinh hay là một đại diện số của một thực thể vật lý hoặc trừu tượng. Trong một vài trường hợp việc chốt tuyệt đối đối tượng được tham chiếu sẽ được yêu cầu, ví dụ, cho một tập hợp dữ liệu để hỗ trợ cho nghiên cứu sẽ khoogn thay đổi để đảm bảo khả năng tái tạo lại; và trong các trường hợp khác khái niệm thực thể như nó hiện được hiểu và được duy trì, ví dụ, một cơ sở đã thay đổi qua thời gian. Sự thường trực này cũng nên được cân nhắc kết hợp với quản lý phiên bản (xem phần 5.5 và 5.6).

3.4 Phân giải được

3.4.1 Có thể có 3 ý định phân giải PID. Một PID là phân giải được khi nó cho phép cả con người và những người sử dụng máy truy cập:

3.4.1.1 Một đối tượng số, một đại diện số, hoặc thông tin về cách để đối tượng đó có thể truy cập được.

3.4.1.2 Thông tin nhân (Kernel): Một hệ thống phân giải toàn cầu sẽ hỗ trợ quyền truy cập tới Thông tin Nhân (Kernel Information) từ PID của nó. Thông tin nhân là một bản ghi có cấu trúc với các thuộc tính mà ngữ nghĩa của nó là truy xuất được ở dạng máy giải nghĩa được (Weigel, T., et al. 2018). Nói chung, Thông tin Nhân nên ít nhất có chứa các thuộc tính trở tới nơi sự tuân tự theo bit của đối tượng được tham chiếu có thể được tìm thấy và một con trỏ trở tới một định nghĩa kiểu dạng. Tối ưu, nó có thể chứa các con trỏ trở tới các đối tượng theo ngữ cảnh hơn nữa, bao gồm cả siêu dữ liệu. Các Hồ sơ Thông tin Nhân PID phải được đăng ký trong các sổ đăng ký mở.

3.4.1.3 Khi một đối tượng hoặc đại diện của nó không còn sẵn sàng nữa, phân giải tới Thông tin Nhân vẫn phải là có thể. Trong các trường hợp như vậy, một thực hành tiêu chuẩn là đưa vào thông tin ‘bia mộ’ (tombstone) như một phần của thông tin nhân (cf. 5.6).

3.4.1.4 Sẽ có những trường hợp nơi mà, vì các lý do hợp pháp, thông tin cá nhân về đối tượng được tham chiếu phải bị xóa. Ví dụ, bám theo một yêu cầu quyền xóa từ một chủ thể dữ liệu theo GDPR. Trong các trường hợp như vậy, khả năng phân giải vẫn nên được duy trì cho thông tin bia mộ để cung cấp nhiều nhất có thể thông tin nhân trong khi vẫn tuân thủ với GDPR và các yêu cầu pháp lý áp dụng được.

3.4.2 Để làm cho nó phân giải được toàn cầu, PID đó cần phải là một phần của không gian tên được xác định bằng cú pháp được Cơ quan quản lý kiểm soát (phần 4.4).

3.5 Ngoài các tính năng đó của bản thân PID, việc sử dụng (hoặc ứng dụng) và quản lý PID đó, và các dịch vụ và điều hành liên quan tới nó, nên kết hợp để xúc tác cho nghiên cứu là FAIR.

3.6 Lưu ý là để tránh nghi ngờ, chính sách này có mục đích trung lập đối với thực tiễn có nhiều hơn một PID tham chiếu tới cùng một thực thể (nhưng cũng xem 2.3). Điều này xảy ra rồi nhưng Thông tin Nhân nên hỗ trợ sự hiểu biết của những chông chéo được biết thông qua sử dụng các thuộc tính tùy chọn như SameAs (tương tự như), AlsoKnownAs (còn được biết tới như) hoặc KnownAs (được biết tới như).

4. CÁC VAI TRÒ VÀ TRÁCH NHIỆM

4.1 Một hạ tầng PID trong EOSC có vài vai trò xác định mà các tác nhân có thể đảm nhận. Từng vai trò có trách nhiệm cho một thành phần cụ thể trong hạ tầng PID, với các bình luận cụ thể cho việc duy trì tính toàn vẹn của hạ tầng PID đó. Trong một kịch bản cụ thể, một tác nhân có thể đóng nhiều hơn một vai trò, hoặc đảm nhận vài hành động có giới hạn có liên quan tới nhiều vai trò, nhưng là hữu ích để chia tách chúng ra để thêm rõ ràng.

4.2 Một hạ tầng PID cũng có thể được hình thành từ vài thành phần, hoặc vài dịch vụ, quy tắc hoặc tiêu chuẩn, chúng xúc tác cho hạ tầng đó để vận hành tuân theo các chính sách và kỳ vọng của các cộng đồng mục tiêu của nó.

4.3 Biểu đồ PID (Thành phần). Một tập hợp các quy tắc và tiêu chuẩn xác định bản chất tự nhiên của một PID. Điều này nên bao gồm một tập hợp các quy tắc định dạng từ vựng cho các PID trong không gian tên. Nó cũng có thể xác định, ví dụ: dạng PID có liên quan; định nghĩa siêu dữ liệu có liên quan; các điều kiện đảm bảo chất lượng; các quyền, các điều khoản và điều kiện sử dụng, và các phương pháp thuật toán để sinh các tên PID và thực thi các thuộc tính PID.

4.4 Cơ quan PID (Vai trò). Một tổ chức kiểm soát có trách nhiệm duy trì các quy tắc nhằm xác định tính toàn vẹn của các PID trong Biểu đồ PID. Các quy tắc đó có thể bao gồm việc thiết lập các tiêu chuẩn cho các định dạng từ vựng, các thuật toán và giao thức để đảm bảo tính duy nhất toàn cầu, cùng với việc thiết lập chất lượng các điều kiện dịch vụ để thực thi tuân thủ với các quy tắc đó. Cơ quan PID có thể là các tổ chức (ví dụ, DOI.org), nó thực thi kiểm soát đối với hạ tầng PID. Nhưng cũng có thể có các cơ quan không có sự kiểm soát tập trung (ví dụ, các mã nhận diện thường trực của Di sản Phần Mềm¹ - Software Heritage và mã nhận diện Phi tập trung của W3C²), mà cung cấp một cơ chế tiêu chuẩn hóa cộng đồng chỉ định sự tuân thủ của các PID đối với một Biểu đồ PID.

4.5 Nhà cung cấp Dịch vụ PID (Vai trò). Một tổ chức cung cấp các dịch vụ PID tuân thủ với Biểu đồ PID, tuân theo Cơ quan PID của nó. Nhà cung cấp Dịch vụ PID có trách

1 <https://docs.softwareheritage.org/devel/swh-model/persistent-identifiers.html#>

2 <https://www.w3.org/TR/did-core/>

nhiệm cung cấp, tính toàn vẹn, độ tin cậy và khả năng mở rộng các Dịch vụ PID, đặc biệt việc phát hành và phân giải các PID, mà còn cả tra cứu và tìm kiếm các dịch vụ, và tính tương hợp với hệ thống phân giải chung.

4.6 Dịch vụ PID (Thành phần). Các dịch vụ cơ bản là các dịch vụ tạo lập, quản lý và phân giải các PID và thông tin nhân có liên quan của chúng mà tuân thủ với Biểu đồ PID. Các dịch vụ cao cấp, giá trị gia tăng cũng có thể được cung cấp, ví dụ, tìm kiếm hoặc đo lường thuộc tính.

4.7 Nhà quản lý PID (Vai trò). Nhà quản lý PID có các trách nhiệm duy trì tính toàn vẹn mối quan hệ giữa các thực thể và các PID của họ, tuân thủ với Biểu đồ PID được Cơ quan PID xác định. Một Nhà quản lý PID thường sẽ đăng ký với các dịch vụ PID để chào chức năng cho các Chủ sở hữu PID trong các dịch vụ của Nhà quản lý PID. Một ví dụ là Nhà cung cấp Dịch vụ sử dụng các Dịch vụ PID như một phần của việc cung cấp dịch vụ của riêng mình. Ví dụ, Nhà quản lý PID có thể bao gồm nhà cung cấp kho dữ liệu, danh mục dữ liệu, hoặc hệ thống tiến trình công việc nghiên cứu.

4.8 Chủ sở hữu PID (Vai trò). Một tác nhân (một tổ chức hoặc cá nhân) người có quyền tạo lập một PID, gắn PID cho một thực thể, cung cấp và duy trì Thông tin Nhân chính xác, cho PID đó. Chủ sở hữu của một PID mới phải được nhận diện và các trách nhiệm này được chuyển giao, nếu Chủ sở hữu PID hiện hành không còn khả năng triển khai chúng nữa.

4.9 Người sử dụng đầu cuối (Vai trò). Người sử dụng đầu cuối các Dịch vụ PID, ví dụ, các nhà nghiên cứu, hoặc phần mềm, hoặc các dịch vụ được sản xuất để hỗ trợ cho các nhà nghiên cứu.

5. CÁC ỨNG DỤNG PID

5.1 Một hệ sinh thái các hạ tầng PID là cần thiết để hỗ trợ cho sự đa dạng rộng lớn các ứng dụng khoa học và chào sự linh hoạt đủ (các nhà nghiên cứu dịch vụ, biểu đồ, tập hợp các thuộc tính) và khả năng. Nên có một Giao diện Lập trình Ứng dụng - API (Application Programming Interface³) chung để tương tác với các PID (tạo lập, phân

3 The RDA Working Groups on PID Information Types and PID Kernel Information have produced preliminary recommendations and prototypes for such an interface and principles for Kernel

giải và sửa đổi PID và Thông tin Nhân PID) sẽ được tất cả các nhà cung cấp dịch vụ PID cung cấp. Các chi tiết của điều này sẽ được xác định tiếp như một phần của công việc triển khai (Nhóm Làm việc về Kiến trúc EOSC).

5.2 Các dịch vụ PID cho EOSC cần đề cập sự đa dạng rộng lớn các ứng dụng bao gồm các ứng dụng đòi hỏi các cơ chế an toàn được xây dựng trong hạ tầng PID đó. Một ví dụ là trường hợp dữ liệu bí mật nơi kiểm soát quyền truy cập và/hoặc mã hóa Thông tin Nhân có thể là phù hợp.

5.3 Phải rõ ràng mọi lúc ai là Chủ sở hữu PID và vì thế ai được phép thực hiện các thay đổi các thuộc tính. Chủ sở hữu PID có trách nhiệm về quản lý đúng cách các PID và giữ cho tập hợp các thuộc tính được cập nhật.

5.4 Mức chi tiết của các PID phụ thuộc rất nhiều vào các cộng đồng và nó sẽ thay đổi qua thời gian. Nhiều mức chi tiết nên được hệ sinh thái PID hỗ trợ và các liên kết giữa các mức chi tiết nên được khuyến khích. Mức độ chi tiết sâu hơn bao giờ hết và sự gia tăng sau đó về số lượng các PID sẽ được tạo ra, quản lý và phân giải, điều này có thể làm tăng nỗ lực và chi phí quản lý Dịch vụ PID cũng như chức năng và tính năng của nó. Và vì thế các mức chi tiết phù hợp cho thực hành tốt nhất và các trường hợp sử dụng của cộng đồng nên được cung cấp, trong khi cho phép linh hoạt để đáp lại việc làm thế nào các nhu cầu và các thực hành sẽ tiến hóa.

5.5 Các dịch vụ PID nên hỗ trợ quản lý phiên bản. Các nhà cung cấp dịch vụ PID và các kho phải có các chính sách và hướng dẫn rõ ràng về cách làm thế nào để quản lý phiên bản trong trường hợp Đối tượng Số FAIR hoặc thực thể thay đổi.

5.6 Các PID không nên bị/được gắn lại hoặc xóa. Trong trường hợp một thực thể đang được nhận diện bị xóa hoặc ngừng tồn tại, thông tin bia mộ cần phải có trong một tập hợp các thuộc tính PID. Ngoài ra, mỗi PID nên cung cấp thông tin về dạng cơ bản và chính sách quản lý đối tượng được tham chiếu theo cách máy hành động được, dù thực thể được tham chiếu có là sẵn sàng dài hạn hay không và dù nó có thể được kỳ vọng thay đổi hay không.

Information that can form the basis for further work.

6. CÁC DẠNG PID

6.1 Các PID có thể nhận diện nhiều thực thể khác nhau. Chúng có thể là bản sinh số (ví dụ, các tài liệu, dữ liệu, phần mềm, dịch vụ - được biết như là các đối tượng số - và các bộ sưu tập được tạo ra từ chúng), vật lý (ví dụ, con người, các công cụ, các chế tác, các mẫu), hoặc khái niệm (ví dụ, các tổ chức, dự án, từ vựng).

6.2 Các thực thể vật lý và khái niệm phải được đại diện thông qua một đại diện số (ví dụ, trang đích (landing page), siêu dữ liệu, tập hợp các thuộc tính, chỉ mục cơ sở dữ liệu) để có sự hiện diện trong bối cảnh số. Tất cả các đại diện số phải là các Đối tượng Số FAIR.

6.3 Các lớp đối tượng số có thể cần các tập hợp thuộc tính khác nhau mà một PID phân giải tới. Là trách nhiệm của cộng đồng thực hành để xác định và ghi thành tài liệu các tập hợp thuộc tính đó (các Hồ sơ Thông tin Nhân PID - PID Kernel Information Profiles).

6.4 Bản ghi thông tin nhân PID là một nguồn siêu dữ liệu không có thẩm quyền nhằm tạo thuận lợi cho việc tự động hóa các quy trình. Nếu thông tin về một thuộc tính đúp bản siêu dữ liệu được duy trì ở đâu đó khác, nguồn bên ngoài đó là có thẩm quyền.

7. CÁC DỊCH VỤ PID VÀ CÁC NHÀ CUNG CẤP DỊCH VỤ PID

7.1 Một nhà cung cấp dịch vụ PID sẽ chào các dịch vụ tích hợp tốt với các hạ tầng nghiên cứu châu Âu, nhưng không loại trừ cộng đồng nghiên cứu rộng lớn hơn. Các dịch vụ cơ bản về đăng ký và phân giải PID sẽ không mất chi phí đối với những người sử dụng đầu cuối.

7.2 Như với các dịch vụ EOSC khác, các dịch vụ PID cơ bản sẽ ở Mức Sẵn sàng Công nghệ - TRL (Technology Readiness Level) 8 (hệ thống hoàn chỉnh và đủ điều kiện) hoặc 9 (hệ thống thực sự được chứng minh trong môi trường vận hành). Các dịch vụ PID giá trị gia tăng có thể được cung cấp ở các TRL thấp hơn.

7.3 Các nhà cung cấp dịch vụ PID cần đảm bảo gần như sẵn sàng “24/7” đối với các dịch vụ PID cơ bản. Các trách nhiệm về duy trì và tính sẵn sàng dịch vụ cần phải được ghi thành tài liệu rõ ràng.

7.4 Các nhà cung cấp dịch vụ PID cần phải có kế hoạch về tính sẵn sàng và kế thừa rõ ràng với một chiến lược rút lui đảm bảo tính liên tục của sự phân giải (như mục 3.4.1) đối với các PID của nó.

7.5 Một tập hợp các Cơ quan đăng ký PID tin cậy và các nhà cung cấp dịch vụ PID là cần thiết thường xuyên được chứng thực dựa vào các tập hợp quy tắc đồng thuận. Chứng thực sẽ bao trùm các khả năng phân giải các PID tới thông tin từ các nhà cung cấp dịch vụ PID và các quy trình quản lý của họ đối với việc duy trì các PID. Nên rõ ràng ai có trách nhiệm duy trì Thông tin Nhân được cập nhật, nếu cần, bằng việc xúc tác cho các bên thứ 3 để sửa đổi nó.

7.6 Có nhu cầu về một hệ thống phân giải PID chung, toàn cầu xuyên khắp tất cả các hệ thống PID và các nhà cung cấp dịch vụ. Để xúc tác cho điều này, các nhà cung cấp dịch vụ PID cần đảm bảo hệ thống của họ hỗ trợ các API cần thiết.

8. ĐIỀU HÀNH VÀ TÍNH BỀN VỮNG

8.1 Các nhà cung cấp dịch vụ PID sẽ áp dụng điều hành cộng đồng đúng cách để đảm bảo rằng các dịch vụ và hệ thống PID của họ gắn với các chính sách đó, và là lan lẹ và có trách nhiệm đối với các nhu cầu nghiên cứu, Khoa học Mở và EOSC. Như vậy, cộng đồng nghiên cứu tích cực của Liên minh châu Âu cần phải có đại diện trong cấu trúc điều hành để có khả năng gây ảnh hưởng tới các hoạt động và các mô hình kinh doanh.

8.2 Các dịch vụ PID sẽ được cung cấp với chi phí hợp lý cho các dịch vụ của EOSC và các nguồn lực sử dụng các dịch vụ đó như các chủ sở hữu PID và các nhà quản lý PID.

8.3 Cấu trúc điều hành sẽ được nhúng trong điều hành toàn cầu.

8.4 Các nhà cung cấp dịch vụ PID sẽ có kế hoạch rút lui thăm định được độc lập đảm bảo tính liên tục của các PID và các dịch vụ PID nếu họ ngừng hoạt động. Kế hoạch này sẽ đảm bảo các nhà cung cấp khác có thể nhân bản các PID và các dịch vụ mà nhà cung cấp gốc ban đầu ngừng tồn tại.

8.5 EOSC sẽ khuyến khích các sử dụng, các dịch vụ và công cụ mới và đổi mới sáng tạo, mà chúng sử dụng và xây dựng dựa vào hạ tầng PID đó.

PHỤ LỤC 1: DANH SÁCH CÁC NGUỒN LỰC ĐƯỢC TƯ VẤN

Đây là danh sách các xuất bản phẩm, chính sách, khuyến nghị và các văn bản khác hiện đang có được tư vấn về quyền tác giả của tuyên bố chính sách này.

Berg-Cross G, Ritz R, Wittenburg P (2018). RDA DFT Core Terms and Model. B2SHARE. <http://hdl.handle.net/11304/5d760a3e-991d-11e5-9bb4-2b0aad496318>

European Commission Expert Group on FAIR Data (2018). Turning FAIR into Reality. Publications Office of the European Union. <https://doi.org/10.2777/1524>

Fenner M, Demeranville T, Kotarski R et al (2016). THOR: Conceptual Model Of Persistent Identifier Linking. Zenodo. <https://doi.org/10.5281/ZENODO.48705>

FORCE11 (2015). The FAIR Principles (2015) <https://www.force11.org/group/fairgroup/fairprinciples>

Lannom L, Broeder D, Manepalli G (2015): Research Data Alliance: Data Type Registries working group output. <https://doi.org/10.15497/A5BCD108-ECC4-41BE-91A7-20112FF77458>

McMurry JA, Juty N, Blomberg N et al. (2017) Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. PloS Biol 15(6): e2001414. <https://doi.org/10.1371/journal.pbio.2001414>

Meadows A, Krznarich L, Brown J, et al (2019). Recommendations for Supporting ORCID in Repositories. figshare. <https://doi.org/10.23640/07243.7777274.v3>

The ODIN Consortium (2013): D4.1 Conceptual model of interoperability. Figshare. <https://doi.org/10.6084/m9.figshare.824314.v1>

Parland-von Essen J (2019). The Use of Persistent Identifiers for Research Datasets Recommendation by the Finnish Scientific Community for Open Research. Edunni wiki. <https://wiki.eduuni.fi/display/csctuha/The+Use+of+Persistent+Identifiers+for+Research+Datasets>

- Rauber A, Asmi A, van Uytvanck D, Proell S (2015): Research Data Alliance Working Group output: Data Citation of Evolving Data: Recommendations of the Working Group on Data Citation (WGDC). <https://doi.org/10.15497/RDA00016>
- RDA GEDE Digital Object Topic Group (2019). FAIR Digital Object Framework <https://github.com/GEDE-RDA-Europe/GEDE/tree/master/FAIR%20Digital%20Objects/FDOF>
- Reed D, Sporny M, Sabadello M, et al (2019). Decentralized Identifiers (DIDs) v1.0. Core Data Model and Syntaxes. W3C Working Draft 27 November 2019. w3.org. <https://www.w3.org/TR/2019/WD-did-core-20191127/>
- TechLib libraries (DTU Copenhagen; ETH Zurich; TIB Hannover; TU Delft) (2019). Pitching for PIDs: European support for a sustainable PID infrastructure - Avoiding a PIDfall. Hannover : Institutionelles Repositorium der Leibniz Universität Hannover. <https://doi.org/10.15488/8825>
- Weigel T, DiLauro T, Zastrow T (2015). Research Data Alliance: PID Information Types WG final deliverable. Zenodo. <http://doi.org/10.15497/FDAA09D5-5ED0-403D-B97A-2675E1EBE786>
- Weigel T, Plale B, Parsons M et al (2018). RDA Recommendation on PID Kernel Information (Version 1). Research Data Alliance. <https://doi.org/10.15497/RDA00031>
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016) <https://doi.org/10.1038/sdata.2016.18>
- Wittenburg P, Hellström M, Zwölf C-M et al. (2017). Persistent identifiers: Consolidated assertions. Status of November, 2017. Zenodo. <http://doi.org/10.5281/zenodo.1116189>

PHỤ LỤC 2: BẢNG CHÚ GIẢI

Các khái niệm được cung cấp để giúp hiểu chính sách, và bổ sung cho các định nghĩa các vai trò và trách nhiệm được xác định trong Phần 4. Tất cả các khái niệm cần được coi là có liên quan tới các PID.

Attribute - Thuộc tính	Giá trị mô tả tính năng của một đối tượng hoặc đại diện của nó, như một phần của thông tin nhân PID hoặc siêu dữ liệu khác.
Digital Object - Đối tượng số	Một đối tượng số có tuần tự bit có thể được lưu trữ trong nhiều kho và có liên quan tới một mã nhận diện thường trực (PID) và siêu dữ liệu.
Digital Entity - Thực thể số	Một thực thể số ngụ ý bất kỳ dạng tuần tự bit nào đang được lưu trữ hoặc truyền mà không có đăng ký để xúc tác cho chia sẻ.
European Open Science Cloud - Đám mây Khoa học Mở châu Âu (EOSC)	Một hạ tầng tích hợp để tạo lập một web dữ liệu FAIR. Sự phát triển của EOSC là một sáng kiến lớn và đang diễn ra với nhiều bên liên quan.
FAIR Digital Object - Đối tượng số FAIR	Một mô hình được báo cáo <i>Biến FAIR thành Hiện thực</i> đề xuất, ngụ ý các yếu tố nào là cần thiết cho một Đối tượng Số để trở thành FAIR và vì thế máy hành động được.
FAIR Principles - Các nguyên tắc FAIR	FAIR là Tìm thấy được, Truy cập được, Tương hợp được, Sử dụng lại được (Findable, Accessible, Interoperable, Reusable). Nó tham chiếu tới các nguyên tắc dữ liệu FAIR được cộng đồng FORRCE 11 phát triển, khuyến nghị dữ liệu nên chia sẻ theo 4 khái niệm đó.
Granularity - Mức độ chi tiết	Các mức phân tầng hoặc các phần cấu thành khác nhau có thể tạo nên dữ liệu hoặc các kết quả đầu ra nghiên cứu khác. Ví dụ, các mức chi tiết khác nhau của một xuất bản phẩm nghiên cứu, đi từ toàn bộ ấn bản tạp chí, tới các bài báo cấu thành của nó, tới các phần hoặc hình ảnh cấu thành của một bài báo, các mức trong một bộ sưu tập khoa học phức tạp hoặc mức chi tiết trong một cơ sở dữ liệu khoa học lớn.
Kernel Information - Thông tin Nhân	Một PID cần phải được phân giải tới một bản ghi có cấu trúc gồm các thuộc tính được xác định tốt để cho phép máy hành động được. Đây là khái niệm về Thông tin Nhân PID như được mô tả trong Khuyến nghị khớp nối của RDA (Weigel, T., et al. 2018).
Landing Page - Trang đích	Một trang người đọc được, hiển thị trên trình duyệt, cung cấp cho người sử dụng là con người thông tin về làm thế nào để truy cập và/hoặc giải nghĩa đối tượng số hoặc đại diện của nó được một PID nhận diện.

Machine Actionable - Máy hành động được	Máy hành động được ngụ ý một tuyên bố chính thức được xác định về cú pháp và ngữ nghĩa xúc tác cho các hệ thống tính toán triển khai việc xử lý tự động.
Metadata - Siêu dữ liệu	Với siêu dữ liệu trong ngữ cảnh của chính sách PID này chúng tôi ngụ ý tất cả các dạng khẳng định về các thuộc tính của tuần tự bit của đối tượng số như có tính mô tả, khoa học sâu, ngữ cảnh, nguồn gốc xuất xứ, các quyền truy cập, giao dịch .v.v. Dạng siêu dữ liệu này không được lưu trữ trong bản ghi PID như là các thuộc tính của nhân, tuy nhiên bản ghi PID nói chung nên trở tới siêu dữ liệu đó.
Namespace - Không gian tên	Không gian tên đảm bảo là tất cả các mã nhận diện trong một mức nhất định có các tên duy nhất sao cho có thể dễ dàng được nhận diện.
Persistent Identifier - mã nhận diện thường trực (PID)	Một mã nhận diện thường trực, duy nhất và phân giải được toàn cầu dựa vào Biểu đồ PID được chỉ định mở.
PID Ecosystem - Hệ sinh thái PID	Tập hợp các thành phần và dịch vụ PID có thể lập thành liên đoàn hoặc tương hợp được để hỗ trợ cho EOSC và nghiên cứu mà chúng là FAIR theo một cách thức hiệu quả và bền vững.
PID Kernel Information Profile - Hồ sơ thông tin nhân PID	Hồ sơ Thông tin Nhân PID mô tả một tập hợp các thuộc tính nhân đang được một kho lựa chọn để mô tả đối tượng số (DO). Hồ sơ đó phụ thuộc vào dạng đối tượng và các nhu cầu của cộng đồng. Chỉ các thuộc tính được công bố mở với các dạng ngữ nghĩa được đồng thuận rộng rãi sẽ cung cấp khả năng hành động của máy.
Prefix - Tiền tố	Thành phần cú pháp PID biểu thị quyền hoặc phân chia không gian tên. Tiền tố cần phải là duy nhất toàn cầu có liên quan tới các cơ quan quản lý địa phương nhất định mà là tự do không mất tiền về cách họ sinh ra các mã nhận diện (ID) duy nhất cục bộ.
Semantics - Ngữ nghĩa	Ý nghĩa hoặc giải nghĩa ý nghĩa được gắn vào một chuỗi văn bản nhất định. Được khuyến nghị không đưa ngữ nghĩa vào một chuỗi của mã nhận diện nơi ý nghĩa có thể thay đổi qua thời gian hoặc có thể không hiểu được qua sự khác biệt về ngôn ngữ và văn hóa.
Suffix - Hậu tố	“Hậu tố” là một phần của một mã nhận diện thường trực duy nhất được cơ quan quản lý địa phương tạo ra cũng chỉ định các quy tắc ngữ nghĩa cho hậu tố đó.

PHỤ LỤC 3: CÁC TÁC GIẢ CỦA CHÍNH SÁCH

Các tác giả của chính sách này là một tập con của các thành viên của các Nhóm Làm việc về Kiến trúc EOSC và EOSC FAIR, cũng như các chuyên gia PID được mời đóng góp cho Đội Đặc nhiệm về PID Hạ tầng.

Tên	Liên kết EOSC	Liên kết
Maggie Hellström	Architecture PID TF	Lund University and ICOS RI
André Heughebaert	FAIR WG	Belgian Biodiversity Platform
Rachael Kotarski	FAIR WG	British Library
Paolo Manghi	Architecture WG	ISTI/CNR
Brian Matthews	Architecture WG	UKRI-STFC
Raphael Ritz	Architecture WG	MPCDF/MPG
Anders Sparre Conrad	FAIR WG	Danish e-Infrastructure Cooperation
Mario Valle	Architecture PID TF	Swiss National Supercomputing Centre (CSCS)
Tobias Weigel	Architecture WG	DKRZ
Peter Wittenburg	FAIR WG	Max Planck Institute

PHỤ LỤC 4: CÁC PHIÊN BẢN VÀ NHỮNG THAY ĐỔI ĐƯỢC THEO DÕI

Phiên bản	Ngày phát hành	Địa chỉ	Các thay đổi từ phiên bản trước
1.0	13/12/2019	https://doi.org/10.5281/zenodo.3574203	N/A
2.0	01/05/2020	https://doi.org/10.5281/zenodo.3780423	Nhiều thay đổi trả lời cho phản hồi từ: - PIDapalooza 2020 - Diễn đàn PID - thư điện tử & phản hồi cá nhân khác DOI mới cho phiên bản 2.
3.0 (phiên bản này)	30/09/2020	https://doi.org/10.2777/zenodo.926037	Các thay đổi trả lời cho phản hồi từ: - Diễn đàn PID - Thư điện tử & phản hồi cá nhân khác cho các tác giả - Nhóm Làm việc về Tính tương hợp Pháp lý về GDPR DOI mới cho phiên bản cuối cùng

Liên hệ với Liên minh châu Âu (EU)

CÁ NHÂN

Tất cả khắp Liên minh châu Âu có hàng trăm trung tâm thông tin Europe Direct.

Bạn có thể thấy địa chỉ của trung tâm gần bạn nhất tại: https://europa.eu/european-union/contact_en

BẰNG ĐIỆN THOẠI HOẶC THƯ ĐIỆN TỬ

Europe Direct là dịch vụ trả lời các câu hỏi của bạn về Liên minh châu Âu.

Bạn có thể liên hệ dịch vụ này:

- bằng điện thoại tự do: 00 800 6 7 8 9 10 11 (các nhà vận hành cụ thể có thể lấy tiền cho các cuộc gọi đó),
- tại số tiêu chuẩn sau đây: +32 22999696, hoặc
- bằng thư điện tử qua: https://europa.eu/european-union/contact_en

Tìm thông tin về Liên minh châu Âu

TRÊN TRỰC TUYẾN

Thông tin về Liên minh châu Âu ở tất cả các ngôn ngữ chính thức của Liên minh châu Âu là sẵn sàng trên website của châu Âu: https://europa.eu/european-union/index_en

CÁC XUẤT BẢN PHẨM CỦA LIÊN MINH CHÂU ÂU

Bạn có thể tải về hoặc đặt các xuất bản phẩm tự do không mất tiền và mất tiền của Liên minh châu Âu từ:

<https://op.europa.eu/en/publications>. Nhiều bản sao các xuất bản phẩm tự do không mất tiền có thể có được bằng việc liên hệ với Europe Direct hoặc trung tâm thông tin địa phương của bạn (xem https://europa.eu/european-union/contact_en)

LUẬT VÀ CÁC TÀI LIỆU LIÊN QUAN CỦA LIÊN MINH CHÂU ÂU

Để truy cập tới thông tin pháp lý từ Liên minh châu Âu, bao gồm tất cả các luật của Liên minh châu Âu từ 1952 trong tất cả các phiên bản ngôn ngữ chính thức, hãy đi tới EUR-Lex tại: <http://eur-lex.europa.eu>

DỮ LIỆU MỞ TỪ LIÊN MINH CHÂU ÂU

Cổng điện tử Dữ liệu Mở của Liên minh châu Âu (<http://data.europa.eu/euodp/en>) cung cấp quyền truy cập tới các tập hợp dữ liệu từ Liên minh châu Âu.

Dữ liệu có thể được tải về và sử dụng miễn phí, cho cả các mục đích thương mại và phi thương mại.

Chính sách này có tác giả là các đại diện của Nhóm Làm việc về FAIR của EOSC và Nhóm Làm việc về Kiến trúc của EOSC.

Chính sách mã nhận diện thường trực (PID) này được viết cho các nhà hoạch định chính sách cấp cao trong các nhà cung cấp dịch vụ và hạ tầng tiềm năng của EOSC và mối quan tâm của tất cả các bên liên quan của EOSC. Nó xác định một tập hợp các kỳ vọng về các mã nhận diện thường trực nào sẽ được sử dụng để hỗ trợ cho một môi trường vận hành nghiên cứu FAIR. Các yêu cầu của các nhà cung cấp và các dịch vụ cơ bản họ chào cũng được nêu.

Nghiên cứu và đổi mới sáng tạo