

IGSN - xây dựng và mở rộng hệ thống PID do cộng đồng dẫn dắt

Tháng 2/2023

Dịch sang tiếng Việt: Lê Trung Nghĩa

Dịch xong: 15/06/2023

Bản gốc tiếng Anh: <https://doi.org/10.5281/zenodo.7330498>

IGSN – building and expanding a community-driven PID system

February 2023

Trường hợp điển hình này là một phần của loạt trường hợp điển hình đã được sản xuất trong nghiên cứu về “Rủi ro và lòng tin khi theo đuổi hạ tầng PID hoạt động tốt cho nghiên cứu” được Trao đổi Kiến thức - KE (Knowledge Exchange) ủy quyền vào tháng 7/2021. Kết quả chính của nghiên cứu này là một báo cáo xem xét tổng quan PID hiện hành với sự nhấn mạnh vào các vấn đề rủi ro và lòng tin có liên quan của nó.

Loạt các trường hợp điển hình bổ sung này nhằm cung cấp sự thấu hiểu sâu hơn trong các lĩnh vực hoạt động cụ thể, các tiến trình và các bên liên quan trong tổng quan PID rộng lớn hơn này.

Tiêu đề: IGSN - xây dựng và mở rộng hệ thống PID do cộng đồng dẫn dắt

Tác giả: Pablo de Castro, Ulrich Herb, Laura Rothfritz và Joachim Schöpfel (tư vấn khoa học scidecode).

Thư điện tử: office@knowledge-exchange.info

DOI: 10.5281/zenodo.7330498



Tất cả các nội dung được xuất bản được chia sẻ theo giấy phép Creative Commons Ghi công (CC BY 4.0) creativecommons.org/licenses/by/4.0



All content published is shared under a Creative Commons Attribution licence (CC BY 4.0) creativecommons.org/licenses/by/4.0

Mục lục

1. Cơ sở lý luận	5
2 Xây dựng PID do cộng đồng dẫn dắt	7
2.1 Nhu cầu của cộng đồng như một động lực	7
2.2 Kết hợp chuyên môn và thực hành tốt nhất như những người tạo thuận lợi	7
2.3 Quản lý tăng trưởng	9
2.4 Các câu trả lời được tùy chỉnh cho các nhu cầu cộng đồng	10
2.5 Các yếu tố thành công	11
3. Các vấn đề xung quanh các rủi ro và lòng tin về sự nổi lên của các PID mới	16
3.1 Tính bền vững	16
3.2 Các yêu cầu của cộng đồng	16
3.3 Các hiệu ứng thương hiệu	16
3.4 Độ tin cậy và tính cập nhật kỹ thuật	18
3.5 Các chỉ số về lòng tin	19
3.6 Các vấn đề về rủi ro	19
4. Quyền tác giả	20
5. Tài liệu tham khảo	21

1. Cơ sở lý luận

Số Mẫu Chung Quốc tế - IGSN (International Generic Sample Number) ban đầu từng là mã nhận diện thường trực cho các đối tượng (mẫu) vật lý.

Nó được một loạt các tổ chức sử dụng, ví dụ, Hệ thống Đăng ký Mẫu Trái đất - SESAR (System for Earth Sample Registration), Khoa học địa lý Úc, Tổ chức Nghiên cứu Khoa học và Công nghiệp Khối thịnh vượng chung - CSIRO (Commonwealth Scientific and Industrial Research Organisation), Điều chung của Dữ liệu Nghiên cứu Úc - ARDC (Australian Research Data Commons), Đại học MARUM Bremen, Trung tâm Nghiên cứu Khoa học địa lý Đức - GFZ (German Research Center for Geosciences), Viện Nghiên cứu Khai thác Biển của Pháp - IFREMER (Institut Français de Recherche pour l'Exploitation de la Mer) và Viện Khoa học địa lý & Tài nguyên Khoáng sản Hàn Quốc - KIGAM (Korea Institute of Geoscience & Mineral Resources). Ngay cả dù nó đã được phát triển để chỉ định các PID chỉ cho các mẫu vật lý từ Khoa học địa lý, nó có thể được sử dụng ngày nay để chỉ định các PID cho bất kỳ đối tượng vật lý hay bộ sưu tập các đối tượng nào. Nhờ có tính mở kỹ thuật và tổ chức của nó cho các trường hợp điển hình khác, IGSN đã trở thành một dịch vụ thú vị cho các cộng đồng khác - nó cũng được phản ánh bởi quyết định của tổ chức triển khai IGSN (IGSN e.V.)¹ thay đổi tên của mã nhận diện từ Số Mẫu Địa lý Quốc tế thành Số Mẫu Chung Quốc tế và đổi tên tổ chức một cách tương ứng. Cuối năm 2021 IGSN e.V. và DataCite đã nhất trí rằng DataCite sẽ vận hành các dịch vụ đăng ký ID và hỗ trợ công nghệ của IGSN. Vào ngày 21/09/2022, DataCite đã công bố rằng việc đăng ký các IGSN bây giờ đã có thể về mặt kỹ thuật².

IGSN được coi là trường hợp điển hình có giá trị trước hết vì các ID của nó trở tới các đối tượng vật lý thay vì sở hữu trí tuệ hoặc các kết quả đầu ra (như các DOI hầu hết làm) hoặc tới các nhà tạo lập của chúng. Ngoài ra, bản thân dịch vụ đó và khung tổ chức của nó đã được phát triển từ dưới lên trong một nỗ lực khổng lồ dựa vào cộng đồng. Nỗ lực này đã thành công theo cách làm cho đáng để nghiên cứu các điều kiện tiên quyết của thành công này về các khía cạnh phát triển tổ chức/kỹ thuật (và các điều này được quản lý và mở rộng phạm vi như thế nào). Ngay từ đầu, dịch vụ này tìm kiếm sự tương đồng về kỹ thuật - tổ chức cho thực hành tốt nhất - một chiến lược đã

1 <https://www.igsn.org/>

2 <https://blog.datacite.org/start-registering-igsn-ids-with-datacite-now/>

làm dễ dàng để bỏ hệ thống xử lý riêng của mình để hợp tác với nhà cung cấp PID đã được thiết lập là DataCite.

Biết rằng điều này cũng có nghĩa là các mã nhận diện IGSN có thể trở thành các DOI của DataCite, và bất kỳ thành viên nào của DataCite cũng sẽ có khả năng đăng ký các mã nhận diện cho các mẫu như các IGSN thông qua DataCite, trường hợp điển hình IGSN cũng có thể giúp minh họa cách bản thân hệ thống PID tiến hóa bằng việc mang các PID mới qua các dịch vụ cộng đồng - đặc biệt là khi các nhu cầu ban đầu được coi là quá nhỏ để được đáp ứng bởi các nhà cung cấp dịch vụ lâu đời không phục vụ cộng đồng. IGSN sử dụng có hệ thống các hiệu ứng thương hiệu (hoặc đã sản xuất chúng qua hợp tác) và sử dụng có hệ thống vốn cấp để bảo vệ sự tăng trưởng của nó và đã xác định rõ ràng nhu cầu chuyên nghiệp hóa. Nó cũng đã hưởng lợi từ kinh nghiệm (tổ chức và kỹ thuật) giành được bởi một trong những tác nhân chính trong triển khai và vận hành dịch vụ đó, điều đã có tác động sâu rộng đến công nghệ và hình thức tổ chức đã được chọn.

2 Xây dựng PID do cộng đồng dẫn dắt

2.1 Nhu cầu của cộng đồng như một động lực

Nhu cầu xây dựng PID chuyên dụng cho các mẫu hoặc các đối tượng vật lý phát sinh từ sự thiếu PID có thể được sử dụng xuyên suốt khắp các tổ chức: “*Các tổ chức như các viện bảo tàng, các khảo sát địa chất, và các chương trình nghiên cứu có kết nối mạng như Chương trình Khám phá Đại dương Quốc tế - IODP (International Ocean Discovery Program) có các hệ thống tại chỗ cho việc nhận diện duy nhất các mẫu của họ. Tuy nhiên, các hệ thống đó bị hạn chế về mức độ phạm vi của tổ chức, chúng không mở rộng vượt ra khỏi biên giới của cơ sở*”.

(Klump et al. 2021)

Vào thời điểm cần có PID cho các mẫu vật lý được khớp nối, đã không có hiểu biết sâu sắc về tổ chức để vận hành một dịch vụ như vậy: “*Khái niệm về IGSN đã được Kerstin Lehnert phát triển tại Đài quan sát Trái đất Lamont-Doherty vào năm 2004 trong một dự án tiền thân là Hệ thống Đăng ký Mẫu Trái đất – SESAR (System for Earth Sample Registration) (Lehnert et al, 2004). Khi đó DataCite còn chưa được thành lập và số ít các DOI đầu tiên cho các tập hợp dữ liệu đã được đăng ký thông qua Thư viện Khoa học và Công nghệ Quốc gia Đức (TIB) (Brase, 2004). TIB đã thấy nhu cầu về một hệ thống mã nhận diện thường trực cho các mẫu nhưng đã coi trường hợp điển hình này là nằm ngoài phạm vi các hoạt động DOI của họ*”.

(Phỏng vấn với Jens Klump).

Ngoài ra, tư cách thành viên của “*Tổ chức DOI Quốc tế – IDF (International DOI Foundation) để có thể gắn DOI độc lập với TIB đã bị loại trừ do phí thành viên IDF cao*” (Klump et al. 2021). Hệ quả là, đã quyết định chạy IGSN trong một “*triển khai chung hệ thống Handle.net, nó đã hoạt động vào năm 2008*”.

(Phỏng vấn với Jens Klump).

2.2 Kết hợp chuyên môn và thực hành tốt nhất như những người tạo thuận lợi

Một tay chơi chính trong phát triển IGSN (cả về kỹ thuật và tổ chức) đã và đang là Jens Klump, người đã được các nhà tư vấn tiến hành phỏng vấn. IGSN đã hưởng lợi từ kinh nghiệm của ông trong triển khai các DOI ở Đức bởi TIB Hannover đã giải quyết các nhu

cầu ngành từng không được các dịch vụ hiện hành đáp ứng: “*Tôi biết Kerstin Lehnert ở Đài quan sát Trái đất Lamont-Doherty, và đã học được về ý tưởng của bà về phát triển một mã nhận diện cho các mẫu địa chất. Đối với tôi, điều rõ ràng là phải sử dụng các mã nhận diện thường trực như DOI để nhận diện các mẫu địa chất, không phải điều gì đó tự làm ở nhà. (...) Vì thế chúng tôi đã tiếp cận TIB Hannover để hỏi họ liệu chúng tôi có thể sử dụng DOI cho các viên đá và Irina Sens đã nói bà yêu khái niệm đó, nhưng điều đó nằm ngoài phạm vi nhiệm vụ của TIB. Nên tôi đặt xuống 50 USD để mua không gian tên Handle. ‘OK, rồi chúng tôi làm thứ của riêng chúng tôi’. Chúng tôi đã vừa nhân bản STD DOI như nó được gọi vào thời điểm đó. Và vì các phần của hạ tầng đã được nhóm của tôi xây dựng, nên điều đó từng là dễ dàng”.*

(Phỏng vấn với Jens Klump).

Sau các bước đầu tiên đó, IGSN đã tập trung vào phát triển tổ chức của nó và phát triển hơn nữa, một lần nữa bám theo thực hành tốt nhất quen thuộc với Jens Klump: “*Và sau đó DataCite đã bắt đầu nổi lên và đã được kết hợp vào năm 2009 và sau đó chúng tôi có thể thấy khái niệm này của IGSN phát triển. Và chúng tôi thấy rằng nó sẽ không thực sự phát triển nếu chúng tôi giữ nó như một dự án NSF, như một dự án thuần túy dựa vào Mỹ. Vì vậy, tôi đề nghị trước tiên chúng ta hãy nghĩ xem nó sẽ trông như thế nào và xem xét ví dụ về DOI và DataCite, đồng thời có thể sao chép mô hình tổ chức. Vì vậy, chúng tôi cũng đã làm điều đó, chúng tôi đã sao chép các đạo luật và thành lập ‘Eingetragener Verein’ (e.V.), đến lãnh sự quán Đức ở San Francisco để đăng ký, bởi vì cuộc họp của Hiệp hội Địa vật lý Mỹ ở San Francisco là nơi duy nhất mà chúng tôi có thể tập hợp tất cả các thành viên sáng lập vào một phòng. Và đó là cách nó bắt đầu”.*

(Phỏng vấn với Jens Klump).

Tuy nhiên, không giống như DataCite, IGSN đã tập trung vào biện hộ cộng đồng hơn là các cơ sở hạ tầng quốc gia trong pha biệt hóa này. Mặc dù IGSN một lần nữa đã bám theo ví dụ của DataCite và đã kết hợp điều hành và vận hành các dịch vụ IGSN trung tâm vào Tổ chức Triển khai Số Mẫu Địa lý Quốc gia - IGSN e.V. (International Geo Sample Number Implementation Organization).

Vì IGSN rõ ràng đáp ứng được nhu cầu bức thiết trong cộng đồng và không có các lựa chọn thay thế đáng kể nào tồn tại, IGSN đã chiếm ưu thế như là mã nhận diện tiêu

chuẩn cho các mẫu: “*Suốt các năm qua, IGSN đã phát triển vượt bậc từ một giải pháp ngách cho thạch học để trở thành một hệ thống nhận diện toàn cầu cho các mẫu với gần 10 triệu đối tượng được đăng ký. Sự hấp thu IGSN bởi các tổ chức khảo sát địa chất quốc gia và các bộ sưu tập chính, cũng như sự tích hợp IGSN vào hồ sơ khoa học thông qua các đường liên kết trong tư liệu khoa học, biến IGSN thành một giải pháp ứng viên mạnh cho mã nhận diện duy nhất ở mức toàn cầu cho các mẫu vật lý*”.

(Phòng vấn với Jens Klump).

Rõ ràng, các PID khác cho các đối tượng vật lý đã được thảo luận, ví dụ, Mã nhận diện Khoa học Đời sống - LSID (Life Science Identifier), tuy nhiên, chúng chưa giành được sự chấp nhận ở mức đó mà IGSN có: “*Cùng lúc, các hệ thống mã nhận diện mẫu đặc thù cộng đồng đã được giới thiệu, được/bị thuyết phục tích cực nhất trong khoa học đời sống và khoa học địa lý. Ví dụ, các cộng đồng tin sinh và đa dạng sinh học đã tạo ra hệ thống mã nhận diện (Mã nhận diện Khoa học Đời sống, LSID) để nhận diện các mẫu và phân loại sinh học. Vì các lý do xã hội - kỹ thuật khác nhau, LSID đã không được áp dụng, và cộng đồng thực dụng đã quyết định ngừng LSID*”.

(Phòng vấn với Jens Klump, xem thêm Klump & Huber 2017).

LSID từng dựa vào không gian tên urn:nbn không được đăng ký với Cơ quan Số được Chỉ định của Internet - IANA (Internet Assigned Numbers Authority), điều giải thích vì sao các ID đó không phải là các URN hoặc URI khắt khe. Tuy nhiên, việc triển khai và ứng dụng của LSID đã gây tranh cãi về việc liệu nên tuân theo cách tiếp cận LSID URN hay cách tiếp cận HTTP URI cho các hồ sơ³.

2.3 Quản lý tăng trưởng

Đổi lại, sự thành công của IGSN là một thách thức lớn đến mức được khuyến nên điều chỉnh lại tổ chức và kỹ thuật cho phù hợp. “*Sự điều hành và các giải pháp kỹ thuật đều trong thời điểm chuyển đổi, vì một trong những điều chúng tôi đã nhận ra là chúng tôi đã và đang vật lộn để giữ cho mọi điều làm việc được. Và tôi đã yêu cầu điều mà lúc đó tôi gọi là chuyên nghiệp hóa tổ chức, không điều hành nó trên cơ sở tình nguyện với sự đóng góp của các thành viên, mà điều hành nó với một người phụ trách có nhiệm vụ duy trì sự nghiệp của họ. Con đường tương tự DataCite đã đi. Nhưng không nhất thiết*

3 See for instance <https://zoobank.org/References/33521a45-4306-4a49-89fc-ac47a015551d>

sao chép điều đó. Vì vậy, chúng tôi đã nộp đơn xin Quỹ Sloan tài trợ cho dự án. (...) Và chúng tôi đã nhận ra rằng trong dự án này chúng tôi cần thay đổi mô hình kinh doanh của chúng tôi, chúng tôi cần tìm ra một đối tác (...)... nơi chúng tôi có thể đối tác trong việc chia sẻ một vài khía cạnh cơ bản của việc vận hành một tổ chức. Và rồi hóa ra là, (...) DataCite thực sự đã rất quan tâm đối tác với IGSN. Vì thế vào tháng 5 của năm ngoái, chúng tôi đã bắt đầu thương thảo về các điều khoản và đến cuối tháng 9, chúng tôi đã ký kết biên bản thỏa thuận giữa 2 tổ chức để đối tác trong việc vận hành hạ tầng kỹ thuật của IGSN. Vì vậy, IGSN, hiện đang chuyển sang khía cạnh kỹ thuật, đang chuyển đổi từ Handle của riêng nó thành DOI được phát hành thông qua DataCite.”

(Phòng vấn với Jens Klump).

Vào cuối năm 2021 IGSN e.V. và DataCite đã nhất trí là DataCite sẽ vận hành các dịch vụ đăng ký và hỗ trợ công nghệ cho IGSN ID. Điều này cũng đồng nghĩa là các mã nhận diện IGSN sẽ trở thành các DataCite DOI, và bất kỳ thành viên nào của DataCite cũng sẽ có khả năng đăng ký các mã nhận diện cho các mẫu như các IGSN thông qua DataCite⁴. (Klump, 2021).

2.4 Các câu trả lời được tùy chỉnh cho các nhu cầu cộng đồng

Như được nêu ở đầu phần 2.1, các hệ thống mã nhận diện cho các đối tượng vật lý đã tồn tại, nhưng chúng đã không làm việc xuyên khắp các tổ chức và vì thế đã không đảm bảo, ví dụ, sự định hướng rõ ràng của một đối tượng: “*Việc giải quyết các dạng mù mờ đó từng là động lực chính cho sự phát triển mã nhận diện duy nhất toàn cầu, sau đó được gọi là Số Mẫu Địa lý Quốc tế (IGSN)*” (Phòng vấn với Jens Klump), vì vậy có một nhu cầu mạnh mẽ của cộng đồng về một PID đáp ứng nhu cầu của họ.

Một yêu cầu khác được IGSN giải quyết là nhu cầu nhận diện không chỉ các mẫu đặc biệt, mà còn các bộ sưu tập vật lý: “*IGSN bây giờ vừa là sự điều hành vừa là một hệ thống kỹ thuật để chỉ định và duy trì các mã nhận diện thường trực duy nhất toàn cầu cho các mẫu vật lý và tổng hợp các mẫu*”.

(Phòng vấn với Jens Klump)

4 As mentioned, on the 21st of September 2022, DataCite announced that the registration of IGSNs was now technically possible.

Một đặc tính khác là các IGSN ID không chỉ được gán cho các đối tượng hay các bộ sưu tập các đối tượng, mà còn cho các đối tượng/thực thể cần thiết để tạo mẫu: “*IGSN có thể được sử dụng để nhận diện các thực thể có kết nối chặt chẽ với các mẫu vật lý. Các ví dụ là các lỗ khoan, mỏ, mỏm đá, hoặc các địa điểm khác. Tất cả chúng đều có chung một điểm là bản thân chúng đều không phải là các mẫu, mà (...) là các đặc tính mẫu*”.

(Phòng vấn với Jens Klump)

2.5 Các yếu tố thành công

Đáp ứng các nhu cầu đặc thù và biết các nhu cầu đó của cộng đồng

Ngoài ra, IGSN ID xúc tác cho quản lý các lược đồ siêu dữ liệu khác nhau, cho phép các ứng dụng tiểu ngành và làm gia tăng việc áp dụng IGSN:

“*Việc tách biệt siêu dữ liệu đăng ký của mã nhận diện khỏi mô tả đối tượng trao cho hệ thống IGSN sự linh hoạt để dần xếp được sự đa dạng lớn hơn của các ứng dụng, điều có thể đòi hỏi các hồ sơ siêu dữ liệu khác nhau mô tả các mẫu của chúng, ví dụ, cho các ngành hoặc các trường hợp điển hình khác nhau*”.

(Phòng vấn với Jens Klump)

“*Việc cho phép các hồ sơ siêu dữ liệu đặc thù ứng dụng như vậy trao cho các đại lý IGSN sự linh hoạt để mô tả các mẫu cho các ứng dụng khác nhau, ví dụ cho phép thu thập các dạng mẫu nhất định với siêu dữ liệu mô tả đặc thù lĩnh vực của chúng được yêu cầu cho các danh mục & các ứng dụng đặc thù lĩnh vực [xem Devaraju et al, 2017]*”.

(Phòng vấn với Jens Klump)

Sự tham gia tích cực của các cộng đồng trong phát triển các lược đồ đó có khả năng đã có ảnh hưởng tích cực lên khả năng áp dụng và tùy chỉnh của chúng:

“*Lược đồ Siêu dữ liệu Mô tả IGSN⁵ được các thành viên IGSN phát triển với các đầu vào từ cộng đồng thực hành trong các khoa học trái đất và môi trường. Nó được sử dụng để lập danh mục cho một tập hợp tối thiểu các thuộc tính mẫu và các bộ sưu tập mẫu có tính mô tả, như dạng mẫu, dạng tư liệu, người đóng góp, và hoạt động tạo mẫu,*

5 <http://schema.igsn.org/description/> and <http://igsn.github.io/metadata>, see also the crosswalk from IGSN to DataCite metadata, <https://support.datacite.org/docs/igsn-id-metadata-recommendations>

tổng hợp các danh mục các mẫu khắp các đại lý IGSN vào trong các cổng điện tử bao quát. Lược đồ này có chủ ý được giữ chung chung để cho phép biên soạn một danh mục toàn cầu về, ví dụ, các mẫu địa chất và sinh học và các bộ sưu tập mẫu”.

(Phỏng vấn với Jens Klump)

Sự linh hoạt này cũng đã cho phép chuyển đổi/phát triển IGSN từ một PID cho thạch học (petrology) thành một PID cho các đối tượng vật lý và các bộ sưu tập: “*Chúng tôi đã phải tìm kiếm vài ý nghĩa cho nó, nhưng trong truyền thông, nó sẽ chỉ là IGSN. Nhưng câu hỏi sau đó là điều này có ý nghĩa gì đối với cộng đồng? Hiện tại, nó thuần túy là cộng đồng địa lý và môi trường, chúng tôi có vài nhà quan sát từ khoa học vật liệu và khảo cổ học, nhưng khái niệm mà chúng tôi phát triển để mở rộng ứng dụng các mã nhận diện IGSN là bằng việc phát triển khái niệm này của ‘cộng đồng các cộng đồng’. Vì thế dạng các cộng đồng ngành bản thân họ vận hành, vì họ biết họ cần gì”.*

(Phỏng vấn với Jens Klump).

Tính mở có tính khái niệm này làm cho IGSN trở thành đối tác hấp dẫn cho các triển khai PID cả từ các ngành và các ngữ cảnh khác: “*Chúng tôi đã được DiSSCo⁶, một Tập đoàn về các bộ sưu tập lịch sử quốc gia ở châu Âu tiếp cận, liệu chúng tôi có thể cung cấp các mã nhận diện cho họ hay không, và họ đã dự kiến cho 3 tỷ đối tượng. Và DataCite đã nói rồi, ‘Không, quá lớn’. Và tôi đã nghĩ, tốt thôi, rồi thì điều này sẽ không xảy ra vào ngày mai, điều này sẽ mất 10 năm, trong thời gian 10 năm chúng tôi có thể làm điều đó. Nên tôi đã nói có, tất nhiên. Nhưng chúng tôi sẽ phải nghĩ về cách thức làm điều này như thế nào”.*

(Phỏng vấn với Jens Klump)⁷.

⁶ <https://www.dissco.eu/>

⁷ DiSSCo, the Distributed System of Scientific Collections (DiSSCo) is a research infrastructure planned in Europe to commence implementation in 2024. In order to start out as a FAIRified system, different handle-based PIDs were evaluated by the project to determine their approach to persistent identification. IGSN was one of them, but DiSSCo, after careful evaluation decided to implement a “DOI-driven approach”, because of the “substantial achievements, operational experience and reputation of DOI/ IDF to date”. (Hardisty et al., 2021)

Các trường hợp điển hình linh hoạt và tính mở về khái niệm kỹ thuật, xã hội

IGSN do một tổ chức quốc tế, Tổ chức Triển khai IGSN (IGSN e.V., <http://www.igsn.org>) điều hành, (Klump, 2021). Trong năm 2021, IGSN và DataCite đã ký kết một Biên bản Thỏa thuận và đã đối tác với nhau. Điều này có nghĩa là các Handle ID IGSN bây giờ sẽ được đăng ký thành các DOI ID IGSN và các Handle có bí danh là các DOI để đảm bảo rằng chúng tiếp tục phân giải. DataCite sẽ cung cấp hạ tầng kỹ thuật và tất cả các thành viên của DataCite cũng sẽ có khả năng chỉ định các ID IGSN.

Tuy nhiên, lược đồ siêu dữ liệu cho các IGSN sẽ không bị thay đổi vì sự nhận diện thường trực các mẫu địa chất có các yêu cầu rất khác nhau so với lược đồ mà DataCite định hướng xuất bản có thể cung cấp. Cách tiếp cận của IGSN về chỉ định siêu dữ liệu dựa vào các quyết định từ các cộng đồng thực hành⁸, trong khi các yêu cầu cho một tập hợp cơ bản các siêu dữ liệu mô tả đã được đồng thuận⁹ là khá độc đáo. Điều này làm cho IGSN rất linh hoạt và hấp dẫn cho các cộng đồng và các trường hợp điển hình khác nhau. Ví dụ lĩnh vực đặc thù các cổng điện tử cộng đồng, như Hệ thống Đăng ký Mẫu Khoa học Trái đất - SESAR¹⁰ (System for Earth Science Sample Registration) vận hành như một đầu mối để làm cho các mẫu tìm kiếm được và truy cập được.

Sự tham gia của cộng đồng

Chắc chắn sẽ không sai nếu cũng đề cập đến sự tin tưởng vào năng lực của những người có trách nhiệm trong cộng đồng, những người đã chứng minh được mình có kiến thức về kỹ thuật - tổ chức (trong trường hợp này là từ ứng dụng DOI ở Đức) để phát triển ứng dụng PID cho cộng đồng.

Sự tham gia của cộng đồng là rất quan trọng cho các hệ thống PID đặc biệt phục vụ cho các nhu cầu của một ngành hoặc cộng đồng nghiên cứu, vì các tính năng của một dịch vụ được phát triển với các khía cạnh gần gũi hơn nhiều với các trường hợp điển hình lấy cộng đồng làm trung tâm hơn là các hệ thống PID chung chung (xem ở trên). Việc thiết lập vai trò mới người quản lý cộng đồng IGSN ở DataCite nhấn mạnh yếu tố quan trọng này.

8 <https://igsn.github.io/communities/>

9 <https://igsn.github.io/metadata/>

10 <https://www.geosamples.org/>

Hợp nhất với các nhà cung cấp PID đã thành lập

Vì một tác nhân trung tâm, Jens Klump, đã quen với các triển khai DOI ở TIB Hannover, thiết kế IGSN có định hướng mạnh tới những phát triển ở TIB và hệ quả là tới DataCite ngay từ đầu: Bắt đầu bằng việc nhân bản triển khai các DOI ở TIB Hannover của Đức (được nêu trong cuộc phỏng vấn), cho tới dạng tổ chức như một hiệp hội có đăng ký (Eingetragener Verein, e.V.). Việc phù hợp với DataCite là yếu tố quan trọng trong phát triển của IGSN ngay từ đầu: *“Điều này bao gồm sự lựa chọn Handle như là giao thức mã nhận diện thường trực nằm bên dưới, điều đã được chọn vào năm 2008 để giữ cho IGSN tương hợp được với DataCite càng nhiều càng tốt”*.

(Phỏng vấn với Jens Klump)

Vì cách tiếp cận này không chỉ dựa vào sự tương đồng với TIB/DataCite, mà còn dựa vào sự cân nhắc các lựa chọn thay thế, không chỉ đã làm việc tương tự như một yếu tố tin cậy, mà còn cả sự đánh giá của DataCite như một thực hành tốt nhất - đặc biệt khả năng của nó để xây dựng hoặc tích hợp các cộng đồng.

“Lý do vì sao chúng tôi đã quyết định đối tác với DataCite từng là vì mô hình kinh nghiệm tập hợp các thành viên mới đi cùng hiện đã chín muồi. Và ngày nay, họ đang hồ hởi để mở rộng các dịch vụ của họ tới các ứng dụng khác. Và họ đã bắt đầu với thứ gì đó rất giống thư mục. Vì siêu dữ liệu gốc bắt đầu từ STD DOI từng dựa vào siêu dữ liệu thư mục theo ISO 690¹¹, nó từng là hạt nhân và mô tả mục đích dự kiến nó đã phải đưa vào trong danh mục của TIB, vì đã không có các cổng điện tử khi đó, cổng điện tử từng là điều mới mẻ. Và sau đó mọi điều đã thay đổi qua năm tháng. Nhưng DataCite vẫn là một hệ thống nặng về thư mục, mà bây giờ các ứng dụng khác muốn sử dụng các mã nhận diện thường trực, như ROR và các công cụ, và các ứng dụng khác bây giờ đang tiếp cận DataCite. Và DataCite là mở cho điều đó. Nhưng với IGSN, họ cũng đã thấy cơ hội đưa một cộng đồng hoàn toàn chín muồi đi cùng, với một trường hợp điển hình khác, nơi hệ thống đó có rồi 10 triệu mã nhận diện được đăng ký, số lượng đó không nhỏ hơn mấy so với DataCite”.

(Phỏng vấn với Jens Klump)

11 ISO. (2010). Guidelines for bibliographic references and citations to information resources (ISO 690:2010) (Standard No. ISO 690:2010). Geneva, Switzerland: International Organization for Standardization (ISO).

Lòng tin này vào khả năng của DataCite để quản lý các cộng đồng đã được sự cung cấp các tài nguyên của DataCite chống trụ, điều tất nhiên củng cố hơn nữa cho lòng tin đó:

“Vì những gì tôi thực sự không kỳ vọng từ DataCite từng là cam kết cho quan hệ đối tác này. Vì vậy, DataCite đã huy động vốn cấp cho việc này, nhưng họ đang cam kết hai người triển khai IGSN. Do đó họ vừa tuyển dụng một người quản lý cộng đồng. Và họ đang trong quá trình tuyển dụng một vai trò hỗ trợ kỹ thuật để giúp cho các thành viên của IGSN ở mức kỹ thuật để tiến hành chuyển đổi. Nhưng các mẫu Người quản lý Cộng đồng là ai đó sẽ vừa làm việc với cơ chế thành viên hiện hành của IGSN tất cả đều là các thành viên của DataCite, và với các thành viên của DataCite mà có thể có quan tâm trong việc đăng ký PID cho các mẫu vật lý để đưa họ cùng đi với quá trình này. (...) Vì lý do này, sẽ là dễ dàng hơn nhiều để sử dụng mạng lưới DataCite hiện có để tiếp cận những người sử dụng mới tiềm năng về khía cạnh này, cho những điều mà họ đã biết rồi. (...) Họ đang sử dụng rồi DataCite DOI. Họ có lẽ chỉ chưa sử dụng đối tượng vật lý dạng tài nguyên. Tôi thực sự muốn nói, IGSN có vị thế hoàn toàn vượt trội hiện nay khi có sự hỗ trợ tuyệt vời này từ DataCite, rằng họ cam kết hai người làm việc với cộng đồng này. Và thông qua thỏa thuận quan hệ đối tác có nhóm chỉ đạo quan hệ đối tác kết nối giữa hai tổ chức đó”.

(Phỏng vấn với Jens Klump)

3. Các vấn đề xung quanh các rủi ro và lòng tin về sự nổi lên của các PID mới

3.1 Tính bền vững

Trong phỏng vấn, đại diện của IGSN đã nhấn mạnh vai trò của các mô hình kinh doanh bền vững, hệt như ông và Robert Huber đã làm trong xuất bản phẩm của họ từ 2017: *“Có lẽ là viên thuốc đắng phải nuốt đối với vài thành viên trong cộng đồng dữ liệu nghiên cứu lo ngại về tất cả những điều thương mại, nhưng các mô hình kinh doanh là khía cạnh cơ bản của các hệ thống PID - các hệ thống PID bền vững không miễn phí”*.

(Klump & Huber 2017)

3.2 Các yêu cầu của cộng đồng

Tính bền vững này được bổ sung như là dấu hiệu của lòng tin đối với sự thừa nhận của cộng đồng rằng PID là có giá trị đối với nó, rằng nó hoạt động và phục vụ cho các yêu cầu của họ: *“Tôi nghĩ các yếu tố chính ở đây là có một mô hình kinh doanh bền vững. Vì cuối cùng, bạn phải có những người làm việc với nó. Đây là một hệ thống xã hội - kỹ thuật, nơi phần dễ dàng hơn đang chạy với công nghệ, phần khó khăn hơn là phần về xã hội, để có lòng tin của cộng đồng mà điều này thực sự là điều gì đó đáng giá, rằng nó làm việc được, nó phục vụ cho các nhu cầu của họ”*.

(Phỏng vấn với Jens Klump)

3.3 Các hiệu ứng thương hiệu

Ngoài ra, đại diện của IGSN nhấn mạnh dấu hiệu lòng tin mà một thương hiệu nổi tiếng trao và rằng IGSN muốn hưởng lợi từ việc đối tác với DataCite: *“Và vì thế đặt vai kề vai IGSN và DOI, DataCiteDOI và URN là sự so sánh thú vị, vì ngay cả dù IGSN đã có từ 2008, không có ngưng nghỉ, nó vẫn được một số người coi như là chưa chín muồi¹². Nên bạn sẽ thấy điều này trong các nghiên cứu và trong tư liệu rằng nó được coi như một mã nhận diện đang nổi lên. Và điều đó là không đúng về khía cạnh kỹ thuật, vì nó thực sự là tiên tiến. Nhưng về phía lòng tin của cộng đồng, nó là một hệ thống đang nổi lên. Do đó đó là khá thú vị để thấy khía cạnh đó. DOI có giá trị thương hiệu tuyệt vời, nó đã có giá trị đó ngay từ đầu, điều giải thích vì sao STD DOI đã chọn DOI chứ không thứ gì khác”*.

(Phỏng vấn với Jens Klump)

¹² See Kotarski et al. (2020)

Điểm này cũng có thể được chứng kiến trong quá trình ra quyết định của DiSSCo. Theo Park et al. (2001) họ lưu ý rằng: “*Việc cung cấp các dịch vụ PID nhất quán, chất lượng cao và xuất sắc được xây dựng đằng sau một nhãn hiệu tạo ra sự công nhận tức thời cho các PID của lược đồ đã được chọn dưới dạng các con trỏ để trỏ rõ ràng tới dữ liệu (kỹ thuật) số chính xác và xác thực cụ thể về một mẫu vật, bao gồm một liên kết không thể phá vỡ đến một mẫu vật vật lý tương ứng hành động để trao quyền. Đáng tin cậy nên làm theo. Chất lượng thông tin là yếu tố mạnh nhất gây ảnh hưởng đến những lợi ích của tổ chức thông qua tính hữu dụng và sự thỏa mãn của người sử dụng được thừa nhận*”. Họ kết luận rằng “*hoạt động dưới một tiền tố của hệ thống Handle khác với tiền tố được IDF và ePIC sử dụng là lựa chọn yếu nhất về cơ bản vì những khó khăn liên quan đến việc giới thiệu một mã nhận diện không được coi là DOI*”.

(Hardisty et al., 2021)

Trong những ngày đầu của nó, trước khi nó đối tác với DataCite, bản thân IGSN đã tìm cách thiết lập sự tin cậy thông qua sự hợp tác với một cơ sở có uy tín: “*Chúng tôi cố gắng phản ánh điều đó trong IGSN, nơi mặc dù mới bắt đầu làm việc với các nhà nghiên cứu riêng lẻ, nhưng ngay từ đầu chúng tôi đã nói rằng chúng tôi phải làm việc với cơ quan Khảo sát Địa chất Quốc gia. Bởi vì họ đáng tin cậy và họ có các bộ sưu tập lớn. Làm việc với các nhà nghiên cứu riêng lẻ sẽ là từng phần, nó sẽ rất khó. Nhưng nếu chúng tôi đưa được các tổ chức lớn đó đi theo cùng, chúng tôi có thể thể hiện lòng tin, và chúng tôi có thể phát triển rất nhanh*”.

(Phỏng vấn với Jens Klump)

Ngay cả DataCite cũng hưởng lợi ở mức độ nhất định từ giá trị thương hiệu của một cơ sở khác, TIB Hannover, nó vận hành các văn phòng DataCite: “*Vâng, tôi nghĩ những gì bạn nói trước đó về tổ chức có sự ủng hộ tốt, rồi thì nó có tính bền vững và phục hồi của riêng nó. Tôi nghĩ TIB đã đưa ra sự kết hợp rất may mắn ở đó họ đã thiết lập tốt, và cả việc hướng về phía trước nữa. Và rằng DOI không chỉ là cơ hội nơi tôi có thể thấy chúng hiện thực hóa tiềm năng của sự phát triển mới, nắm bắt nó. Vì thế với TIB, đó là sự kết hợp rất tốt. Nhưng để có nền tảng vững chắc, tôi nghĩ, thực sự là quan trọng. Và trong những ngày đầu, việc đưa dự án do DFG cấp vốn thành một dòng các hạng mục trong ngân sách của TIB thực sự là một điều quan trọng*”.

(Phỏng vấn với Jens Klump)

3.4 Độ tin cậy và tính cập nhật kỹ thuật

Là một thành phần tin cậy khác, chuyên gia tạo ra độ tin cậy kỹ thuật và tính cập nhật, điều này đã được đề cập trong giai đoạn đánh giá đối với URN và Handle: “*Và khi chúng tôi đã đánh giá và triển khai URN, chúng tôi đã gặp phải một vấn đề thú vị ở đó vì ở đây chúng tôi đã phải làm việc với dạng công nghệ chưa chín muồi. Đã có những phần ở nơi này nơi khác từng có hứa hẹn, nhưng chúng chưa bao giờ được hiện thực hóa. (...) Và trong ví dụ về URN, đây là một kinh nghiệm học tập thú vị về chất lượng dịch vụ. Vì vậy, dịch vụ này rất đáng tin cậy, bởi vì nó được Thư viện Quốc gia điều hành. Nhưng điều này từng rất cồng kềnh. Nhưng chúng tôi đang nói về năm 2004 ở đây, vì thế các dịch vụ web là điều gì đó mới. Khi Michael Diepenbroek và tôi đã yêu cầu TIB cung cấp cho chúng tôi một điểm cuối dịch vụ web, trong cuộc họp tiếp sau, chúng tôi đã được trình bày với một biểu mẫu web và chúng tôi đã nhìn anh ta đầy thắc mắc: ‘Cái gì thế?’ Đã có một chút khoảng trống văn hóa ở đó. Và Thư viện Quốc gia đã làm việc theo cách thức y hệt đối với chúng tôi để đăng ký các URN cho các tập hợp dữ liệu, họ đã muốn chúng tôi gửi các bảng tính Excel qua thư điện tử, và sau đó họ muốn trả về các mã nhận diện cho chúng tôi với việc tổng kiểm tra được thêm vào. Không, đó không phải là cách nó làm việc. Chúng tôi có hơn một triệu mã nhận diện rồi! Đây không phải là việc chúng tôi đang đẩy 40 cuốn sách mỗi ngày tới quầy.*”

(Phòng vấn với Jens Klump)

Nhận xét rằng các thư viện quốc gia, trong khi có lợi thế về sự tin cậy do đã tồn tại hàng thế kỷ, nhưng hơi kém linh hoạt về mặt kỹ thuật, có thể dẫn đến ý tưởng về một bức tranh lý tưởng về nhà cung cấp PID kết hợp cả hai đặc điểm, sự tồn tại lâu đời và có thiện chí đổi mới về mặt kỹ thuật, với nguồn cấp vốn vững chắc đặt lên hàng đầu (điều này có lẽ đã được thể hiện ở một mức độ nào đó trong TIB Hannover và DataCite): “*Vì vậy, bạn có thể thấy những dòng riêng biệt này ở những nơi, về mặt kỹ thuật thì không khó lắm, nhưng bạn cần phải có một nền tảng tổ chức vững mạnh để có thể chạy điều này trong nhiều thập kỷ. Phải có tiền vào. Và bạn phải tìm ra cách thức phục vụ cộng đồng, giải quyết các vấn đề và được coi là đáng tin cậy.*”

(Phòng vấn với Jens Klump)

3.5 Các chỉ số về lòng tin

Jens Klump cũng đã giới thiệu một chỉ số có thể cung cấp thông tin về mức độ ở đó cộng đồng (hoặc ít nhất các tổ chức như các đại diện của cộng đồng đó) tin cậy một dịch vụ, thiện chí của nó để cam kết và đầu tư vào dịch vụ đó¹³:

“Vì vậy, một thước đo quan trọng đối với tôi liệu họ có tin tưởng vào hệ thống hay không là liệu họ có đầu tư vào nó hay không. Liệu họ có đặt các nguồn lực cho nó hay không. Điều này là đặc biệt, với các cơ quan chính phủ, khi họ đưa ra một cam kết. Một khi họ đã biến điều này thành một dòng hạng mục, họ gắn với nó. Nhưng để đi tới được điểm đó là không dễ. Vì thế họ phải được/bị thuyết phục là đáng để đi qua tất cả điều này. Vì thế đó là từ quan điểm hạ tầng”.

(Phỏng vấn với Jens Klump)

Ngoài các tổ chức, với các cá nhân là đại diện của cộng đồng, các chỉ số về lòng tin khó xác định hơn, rất có thể là nhu cầu về dịch vụ:

“Và quan điểm của người sử dụng, là khó khăn hơn nhiều để chỉ ra liệu cộng đồng có tin vào hệ thống hay không. Đó là vì sự hấp thu là khó đo đếm. (...) Vì vậy, khi chúng tôi thấy mọi người yêu cầu cung cấp dịch vụ, thì ... điều này cũng là một chỉ số lòng tin. Với các hệ thống mã nhận diện được phát triển trong nội bộ khác, chúng tôi không thấy việc biện hộ của cộng đồng ở bất cứ đâu điều này nên được sử dụng và thấy sự hấp thu ở phía nhà cung cấp cũng không thật thường xuyên”.

(Phỏng vấn với Jens Klump)

3.6 Các vấn đề về rủi ro

Cho tới nay, không yếu tố rủi ro nào đã được nhận diện trong trường hợp điển hình này, ngay cả dù các dấu hiệu về quan hệ đối tác với DataCite là rất tích cực (tầm với, các tùy chỉnh có thể bởi các cộng đồng khác, sự đầu tư vào nguồn lực con người), có thể gây ra các rủi ro (quá tải của nhà cung cấp DataCite hoặc mất tính linh hoạt của IGSN thông qua hợp tác với DataCite) chỉ có thể được nhận diện qua thời gian.

13 Perhaps the caveat might be: provided the community has funds and can invest sovereignly to support the service.

4. Quyền tác giả

Trường hợp điển hình này chủ yếu được Ulrich Herb (Đại học Saarland, ORCID <https://orcid.org/0000-0002-3500-3119>) trong nhóm các nhà tư vấn bao gồm Pablo de Castro (Đại học Strathclyde và euroCRIS, ORCID <https://orcid.org/0000-0001-6300-1033>), Laura Rothfritz (Đại học Humboldt ở Berlin, ORCID <https://orcid.org/0000-0001-7525-0635>) và Joachim Schöpfel (Đại học Lille và euroCRIS, ORCID <https://orcid.org/0000-0002-4000-807X>) dưới cái ô của tư vấn khoa học scidecode (ROR <https://ror.org/02c0bjd31>). Công việc đã được Nhóm Nhiệm vụ & Hoàn thành của KE giám sát với thành phần của nó được liệt kê ở <https://www.knowledge-exchange.info/event/pids-risk-and-trust>.

5. Tài liệu tham khảo

- Brase, J (2004). Using Digital Library Techniques – Registration of Scientific Primary Data. In R. Heery & L. Lyon (Hrsg.), *Research and Advanced Technology for Digital Libraries* (vol. 3232, p. 488–494). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-30230-8_44
- Devaraju, A, Klump, J, Tey, V, Cox, SJD and Fraser, R. 2017. Towards a web-enabled geo-sample web: An open source registration and management system for connecting geo-samples to the web. In: *Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings*. Presented at the FOSS4G 2017. Boston, MA: OSGEO. p. Article 16. DOI: <https://doi.org/10.7275/R5P26W9H>
- Hardisty AR, Addink W, Glöckler F, Güntsch A, Islam S, Weiland C (2021) A choice of persistent identifier schemes for the Distributed System of Scientific Collections (DiSSCo). *Research Ideas and Outcomes* 7: e67379. <https://doi.org/10.3897/rio.7.e67379>
- Klump, J et al, (2021). Towards Globally Unique Identification of Physical Samples: Governance and Technical Implementation of the IGSN Global Sample Number. *Data Science Journal*, 20(1), p.33. DOI: <http://doi.org/10.5334/dsj-2021-033>
- Klump, J., & Huber, R. (2017). 20 Years of Persistent Identifiers – Which Systems are Here to Stay? *Data Science Journal*, 16, 09. <https://doi.org/10.5334/dsj-2017-009>
- Kotarski, R., Kirby, J., Madden, F., Mitchell, L., Padfield, J., Page, R., et al. (2020). PIDs as IRO Infrastructure – Early Findings. London, UK: British Library. <https://doi.org/10.23636/1214>
- Lehnert, KA, Goldstein, SL, Lenhardt, C, & Vinayagamoorthy, S (2004). SESAR: Addressing the need for unique sample identification in the Solid Earth Sciences. AGU Fall Meeting Abstracts, 2004, SF32A-06. <https://ui.adsabs.harvard.edu/abs/2004AGUFMSF32A..06L/abstract>
- Park, S., Zo, H., Ciganek, A. P., & Lim, G. G. (2011). Examining success factors in the adoption of digital object identifier systems. *Electronic Commerce Research and Applications*, 10(6), 626–636. <http://doi.org/10.1016/j.elerap.2011.05.004>

Knowledge Exchange

C/O Jisc

4 Portwall Lane,

Bristol, BS1 6NB

United Kingdom

T +44 203 697 5804

E office@knowledge-exchange.info