

Identificadores persistentes: construyendo el avión mientras volamos en él

Persistent identifiers: building the plane as we fly it

Pablo De-Castro

De-Castro, Pablo (2022). "Identificadores persistentes: construyendo el avión mientras volamos en él". *Anuario ThinkEPI*, v. 17, e17a24.

<https://doi.org/10.3145/thinkepi.2023.e17a24>

Publicado en *IweTel* el 19 de julio de 2023

Pablo De-Castro

<https://orcid.org/0000-0001-6300-1033>

<https://www.directorioexit.info/ficha2429>

University of Strathclyde & euroCRIS

pcastrmartin@gmail.com



Resumen: Los identificadores persistentes o PIDs se definen como referencias permanentes y globalmente únicas a cualquier clase de entidad digital o física y constituyen un área en rápido desarrollo en la actualidad. Además de los PIDs ya consolidados como los DOIs o los ORCIDs, existe toda una panoplia de identificadores emergentes –desde los *grant IDs* o identificadores persistentes de financiaciones hasta los identificadores de equipos e instalaciones de investigación pasando por los OrgIDs o identificadores de organizaciones y los IGSNs o identificadores de geomuestras. Se ofrecen aquí algunas reflexiones sobre la gradual evolución en esta materia y la conveniencia de contar con estrategias para la implantación y el uso de esta tecnología.

Palabras clave: Identificadores persistentes; Análisis del panorama; Estudios de caso; Nivel de madurez; Estrategias PID nacionales.

Abstract: Persistent identifiers (PIDs) may be defined as permanent and globally unique references to any kind of digital or physical entity. The area of PIDs is currently experiencing a remarkable development. There is a whole range of emerging PIDs on top of the already consolidated ones such as digital object identifiers (DOIs) and ORCIDs for authors. Among the emerging PIDs, it is worth mentioning grant identifiers issued by research funders, identifiers for research equipment and facilities, organizational identifiers (Org IDs), and identifiers for physical geosamples (IGSN). Some reflections are shared here on the gradual evolution in the discipline and the advisability of having strategies in place for the adoption and use of this technology.

Keywords: Persistent identifiers (PIDs); Landscape analysis; Case studies; Level of maturity; National PID strategies.

1. PIDs: una infraestructura emergente

Son numerosos los países que han definido o están definiendo una estrategia nacional en torno a los identificadores persistentes (**Walker**, 2023). Bajo la coordinación de un grupo de trabajo de la *Research Data Alliance (RDA)*, estos países mantienen reuniones y conversaciones periódicas para aprender los unos de los otros y avanzar juntos en la definición e implantación de una hoja de ruta conjunta. En Europa, países como Holanda, Finlandia o el Reino Unido han sido pioneros en definir una clara estrategia a seguir en este ámbito y están actualmente trabajando en su implantación. Estas iniciativas se unen a las que vienen llevando a cabo otros países fuera de Europa, en particular Australia, que tradicionalmente ha constituido un ejemplo a seguir.

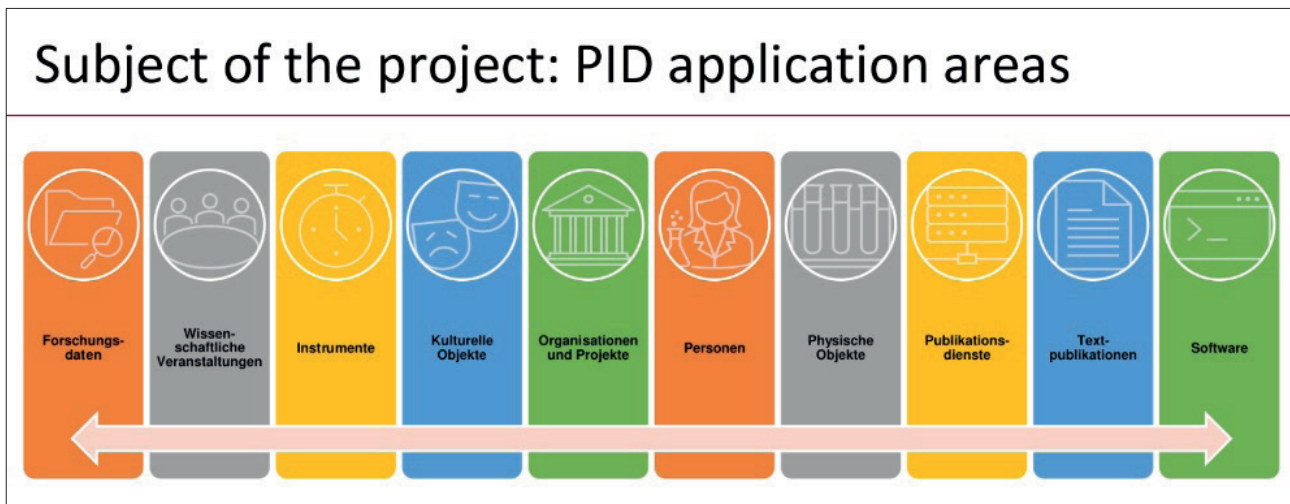


Figura 1. Áreas de actividad del proyecto "PID Network Germany" (2023-2026) financiado por la *German Research Foundation (DFG)*, <https://www.pid-network.de/en>

Las diez áreas son: datasets, eventos, instrumentos, objetos culturales, organizaciones y proyectos, personas, objetos físicos, servicios de publicaciones, publicaciones textuales y software

El desarrollo de una red de identificadores persistentes es también de suma relevancia para la consolidación de la *European Open Science Cloud (EOSC)*, porque es en el área de la gestión de datos de investigación –incluyendo los desarrollos de software– donde se perciben las mayores oportunidades para esta infraestructura. De hecho, el desarrollo e implantación de identificadores persistentes son elementos clave en proyectos en curso en la órbita de la *EOSC* tales como *FAIRCORE4EOSC* coordinado desde *CSC* in Finlandia) y *Fair Impact* (desde *DANS* en los Países Bajos). Uno de los objetivos principales del primero de ellos es por ejemplo la puesta en servicio de los *RAiDs* (*Research Activity IDs*) como mecanismos basados en la tecnología *handle* para la identificación persistente de proyectos en universidades y centros de investigación de la Unión Europea (*FAIRCORE4EOSC*, 2023). Hasta el momento únicamente Australia ha avanzado en la emisión de *RAiDs*, pero existen iniciativas independientes –más o menos coordinadas– para establecer agencias *RAiD* en la UE, en EUA y en el Reino Unido.

Desde las instituciones, y en particular desde las bibliotecas de investigación, viene promoviéndose desde hace tiempo el uso de los identificadores persistentes más habituales –*DOIs* para publicaciones y datasets, *Orcids* para autores y crecientemente *RORs* para organizaciones– en el marco de las tareas de apoyo a la investigación relacionadas con la gestión de resultados científicos. Organizaciones como *Crossref*, *DataCite* y *Orcid* constituyen actores clave para el progreso en la adopción de estos *PIDs* más consolidados y lo serán también para los *PIDs* emergentes. La idea fundamental es que, en la descripción de metadatos de los objetos digitales oportunos, los campos de texto libre puedan gradualmente ir reemplazándose (donde corresponda) por *PIDs* legibles por máquina que permitan explotar los grafos de investigación subyacentes. Estos *PID graphs* enlazan entidades tales como autores, instituciones, resultados de investigación, agencias de financiación, proyectos etc., facilitando el análisis y la generación de informes desde diversas perspectivas, dependiendo del punto de acceso a la red.

A partir de un *PID graph* plenamente desarrollado resulta por ejemplo muy sencillo para una agencia de financiación identificar el conjunto de resultados de investigación (publicaciones, datasets, patentes, tesis doctorales...) asociados a un proyecto concreto financiado por la agencia. La consultora *MoreBrains*, que viene proporcionando el soporte teórico para la expansión de la adopción de *PIDs* en el Reino Unido, ha elaborado una serie de diagramas mostrando ejemplos donde el uso de los *PIDs* mejora significativamente la eficiencia de numerosos flujos de trabajo relacionados con la gestión de la información científica (*MoreBrains*, 2022).

En los diez ejemplos del "ciclo de investigación optimizado a través del uso de *PIDs*" que expone *MoreBrains* no se detallan aún los posibles beneficios derivados de contar con identificadores persistentes para equipos e instalaciones de investigación. La disponibilidad de estos *PIDINSTs* –a

"Estamos volando el avión mientras lo construimos, es decir, estamos tratando de explotar los beneficios de una red de identificación persistente que está aún muy lejos de estar completa"

cuya definición y gradual implantación dedica sus esfuerzos otro grupo de trabajo internacional de la *Research Data Alliance (RDA, 2022)*– ofrecería potencialmente a instituciones y sobre todo a las agencias de financiación (quienes como norma financian estos a menudo carísimos equipos) la oportunidad de obtener informes de uso de dichos equipos e instalaciones: qué investigadores, qué organizaciones (tales como universidades y centros de investigación pero también empresas externas) los utilizan, qué proyectos y con qué resultados de investigación (publicaciones, datasets, etc.). Si este flujo de trabajo concreto no se muestra en el diagrama de *MoreBrains* es porque aún no contamos con PIDs para este tipo de equipos de investigación, por mucho que actores como el grupo de trabajo de la *RDA, Data-Cite* y algunas agencias nacionales de gestión de la información científica en determinados países estén dando pasos muy significativos para el establecimiento de estos mecanismos. Es indudable también que la incorporación de una red de PIDs para equipos e instalaciones de investigación supondría también un valioso activo adicional para la *EOSC*.

Es en este sentido que estamos volando el avión mientras lo construimos, es decir, estamos tratando de explotar los beneficios de una red de identificación persistente que está aún muy lejos de estar completa. “Construyendo el avión mientras volamos en él: la promesa de los identificadores persistentes” es de hecho el título de un informe sobre el desarrollo e implantación actuales de los PIDs (**De-Castro et al., 2023**) encargado por la organización *Knowledge Exchange* a un equipo de consultores del cual forma parte el autor de estas líneas. Este informe, publicado en febrero de este año junto a siete estudios de caso (**De-Castro, 2023**) que examinan el estado del arte en diversas áreas de la identificación persistente, plantea como un objetivo crítico el análisis de los riesgos y de la necesaria fiabilidad en la construcción y operación de la infraestructura de identificadores persistentes tal como se está llevando a cabo en la actualidad.

2. Build it and they'll come?

El informe y sus estudios de caso asociados se basan en buena medida en una serie de entrevistas con expertos en el área de los PIDs en diversos países (con atención particular a los seis que componen la red de colaboración *Knowledge Exchange*, esto es, Alemania, Dinamarca, Finlandia, Francia, Países Bajos y Reino Unido). Los expertos se seleccionaron de manera que representaran diversas clases de organizaciones, tales como agencias de financiación, agencias nacionales de gestión de la información científica, universidades, bibliotecas nacionales, editores comerciales y proveedores de PIDs entre otras. Las impresiones de estos expertos sobre diferentes aspectos se destacan entrecomilladas en el informe y en los estudios de caso y proporcionan una panorámica sobre los retos y las oportunidades en la materia.

El informe contiene una serie de recomendaciones para los diversos grupos de organizaciones que se perciben como relevantes para el desarrollo de los PIDs, incluyendo universidades, proveedores de



Figura 2. Planes para la implantación de identificadores persistentes para equipos e instalaciones de investigación en la hoja de ruta 2023/24 del proyecto *equipment.data* de *Jisc* para una base de datos nacional de equipos e instalaciones (ya existente) en el Reino Unido. <https://equipment.data.ac.uk>

infraestructura PID, *funders*, agencias nacionales, editoriales, investigadores entre otros. De hecho, uno de los riesgos que se describen en el trabajo es la considerable complejidad de la red de actores involucrados en el desarrollo y adopción de esta tecnología y la dificultad de coordinar iniciativas estrechamente relacionadas que vienen desarrollándose en paralelo sin contacto entre las organizaciones que las llevan a cabo.

Se observa una fuerte tendencia en este sentido a tratar de simplificar los flujos de trabajo asociados al desarrollo y la implantación de PIDs. Esta actitud se percibe en particular entre proveedores de PIDs tales como *Crossref*, *DataCite* y *Orcid*, que a menudo consideran que por lo que respecta a los PIDs emergentes estamos ante un nuevo caso de “*build it and they’ll come*”. Esto quiere decir que la estrategia adoptada en exitosos desarrollos previos tales como los DOIs y los Orcids se considera reutilizable para los PIDs emergentes. Esta estrategia “*build it and they’ll come*” consiste en desarrollar en primer lugar una serie de soluciones técnicas –incluyendo entre otras cosas los esquemas de metadatos asociados a los nuevos PIDs– para después invitar a una serie de usuarios (investigadores, universidades, editoriales, repositorios y sistemas CRIS, etc.) a que hagan uso de la infraestructura creada para ellos en la confianza de que satisfará sus necesidades.

Esta estrategia –que incluye como norma consultas con la futura comunidad de usuarios– ha funcionado muy bien en áreas tales como autores (Orcid) y publicaciones (DOI), en las que existía una clarísima necesidad de contar con mecanismos de identificación persistente. En esos casos, las organizaciones usuarias de la infraestructura PID mencionadas más arriba han percibido las evidentes ventajas de contar con estos identificadores y se han apresurado a transmitirlos a los investigadores y a dar sus propios pasos para su adopción. Existe no obstante un cierto riesgo de falta de información, fragmentación y en consecuencia de falta de adopción en ámbitos mucho más complejos como la identificación persistente de organizaciones de investigación o de equipos e instalaciones.

En el caso de los OrgIDs, los casos de uso permanecen insuficientemente definidos, con una fuerte tendencia a simplificar excesivamente su diseño de modo que se empleen únicamente para definir entidades de nivel superior o *top-level entities*. Esta problemática, que viene siendo objeto de una intensa discusión en los foros públicos, puede dificultar la aplicación de los OrgIDs a situaciones en evidente necesidad de estandarización como es la definición de afiliaciones. Como ejemplo, el ROR para el CS/C es uno de los más sofisticados disponibles en la base de datos de este registro de organizaciones de investigación, con 122 “entidades hijas” que incluyen todos los centros e institutos de investigación del Consejo, así como algunas entidades relacionadas.

<https://ror.org/02gfc7t72>

Sin embargo, afiliaciones más detalladas como –por ejemplo– el *Departamento de Ecología Evolutiva de la Estación Biológica de Doñana* que los investigadores utilizan con regularidad al firmar sus trabajos, carece de entrada persistente ROR.

<https://doi.org/10.1016/j.envexpbot.2020.104336>

Para aquellos que comenzamos nuestra carrera profesional como catalogadores es fascinante atisbar el inesperado regreso al primer plano de los polvorientos catálogos de autoridades y la constatación de que, a diferencia de los DOIs y los Orcids, para gestionar identificadores de organizaciones que cambian con relativa frecuencia se precisa una entidad (históricamente, la biblioteca) que asuma la propiedad y el mantenimiento de estos OrgIDs.

Por contra, si el principal caso de uso de los OrgIDs es la adecuada identificación de la organización responsable del pago de costes por procesamiento de artículos (APCs), la limitación de la identificación a entradas de nivel superior tiene todo el sentido dado que estos pagos suelen estar administrativamente centralizados, en particular una vez consolidados los flujos de trabajo relativos a los así llamados acuerdos transformativos. Se aprecia en todo caso un cierto riesgo de opacidad derivado de dejar el diseño y desarrollo de la infraestructura PID en manos de organizaciones en las que las entidades usuarias de la infraestructura pueden no estar suficientemente representadas. Son los riesgos tradicionalmente asociados al hecho de ser usuarios de tecnología en lugar de creadores de la misma. Contar con un consorcio Orcid y un nodo *DataCite* en el país puede por tanto no ser suficiente para evitar estos riesgos si no se está representado en ninguno de los grupos de trabajo dedicados a tratar de consensuar el diseño de la infraestructura emergente.

“A diferencia de los DOIs y los Orcids, para gestionar identificadores de organizaciones que cambian con relativa frecuencia se precisa una entidad (históricamente, la biblioteca) que asuma la propiedad y el mantenimiento de estos OrgIDs”

3. Soberanía en el ámbito de la identificación persistente

Relacionado con lo anterior, uno de los riesgos que alguno de los expertos entrevistados para el trabajo de *Knowledge Exchange* mencionado más arriba ha planteado es el de la sostenibilidad de un modelo de financiación basado en la membresía de entidades sin ánimo de lucro como *Crossref*, *Orcid* o *DataCite*. La expresión concreta utilizada por el entrevistado para plantear ese riesgo fue

“existe un límite para el número de organizaciones de las que universidades y centros de investigación pueden hacerse miembro”.

En un momento en el que se percibe como prioritaria la expansión de la implantación de los identificadores persistentes más consolidados a regiones tradicionalmente menos propensas a su utilización, este riesgo se pone particularmente de manifiesto. Las cuotas para la membresía de estas organizaciones pueden no ser gravosas para entidades en el llamado “norte global”, pero no todas las regiones del planeta se encuentran en las mismas circunstancias. Es relevante mencionar en este aspecto la importante componente económica en la emergencia del concepto de soberanía en el desarrollo de infraestructura de identificación persistente.

Este movimiento hacia una “soberanía PID” se aprecia en particular en América Latina, donde pese a la significativamente buena ubicación de países como Perú, Colombia, Chile o Brasil en la gráfica de adopción y uso de *Orcid* (Porter, 2022), existe una fuerte conciencia de las dificultades de asumir modelos de financiación basados en pagos a organizaciones que se perciben a menudo como al servicio de los intereses de la investigación del norte global. Dada la diversidad de mecanismos técnicos disponibles para la identificación persistente, no es descabellado plantear entonces un modelo alternativo de identificación persistente basado por ejemplo en la tecnología ARK (Matas et al., 2023) cuyo diseño e implantación se adapten a las circunstancias y a las necesidades específicas de la región.

El informe para *Knowledge Exchange* no aborda los riesgos asociados a los aspectos económicos y sociopolíticos en la implantación de la tecnología PID dado que esto excede el área de cobertura de un trabajo centrado en los seis países europeos que forman el entramado KE. Sin embargo, cuando se pone de manifiesto que los intentos de obtener una licencia para el uso a nivel nacional de la tecnología *handle* en Cuba se han demostrado baldíos por consideraciones de esta naturaleza, no sorprende que se opte por soluciones alternativas (Cervantes Producciones Digital, 2022). De nuevo algunos de los entrevistados para el trabajo sobre PIDs para *Knowledge Exchange* estiman que la existencia de soluciones competidoras desarrollándose en paralelo –tales como RORs e identificadores *Ringgold* para la identificación persistente de organizaciones– no constituye un serio riesgo de fragmentación del panorama PID global sino al contrario, un incentivo para que los diferentes actores involucrados hagan las cosas mejor.

En este sentido, una de las principales recomendaciones del informe de *Knowledge Exchange* es el establecimiento de un “observatorio de PIDs” que permita un seguimiento de los progresos en las diferentes áreas por parte de los diversos actores. En la actualidad existen espacios como el “PID Forum” (Freya, 2019), pero un observatorio gestionado por una entidad independiente de los proveedores de infraestructura PID representaría un complemento muy necesario al panorama PID actual.

4. Conclusión

En vista de la rápida evolución del entorno de identificadores persistentes es importante que las diversas organizaciones involucradas en su adopción y su uso permanezcan al tanto de los desarrollos en la materia. Es razonable también considerar la participación en la conversación que a nivel internacional viene manteniéndose en el ámbito del establecimiento de estrategias nacionales para la implantación de PIDs. En este aspecto, y dada la relevancia de este campo para la *EOSC*, la celebración del *EOSC Symposium 2023* en Madrid el próximo mes de septiembre (EOSC, 2023) puede constituir una buena oportunidad para ponerse al día con respecto a los progresos de proyectos como *FAIRCORE4OESC* y del grupo de trabajo “PID policy and implementation” que forma parte del grupo de 13 “task forces” dedicado a definir áreas clave para la implantación de la *EOSC* (EOSC, 2022).

“El desarrollo de una red de identificadores persistentes es de suma relevancia para la consolidación de la European Open Science Cloud (EOSC)”

“Uno de los riesgos que se plantean es el de la sostenibilidad de un modelo de financiación basado en la membresía de entidades sin ánimo de lucro como Crossref, Orcid o DataCite”

5. Referencias

Cervantes Producciones Digital (2022). *Identificador digital cubano*.
<https://lcu-id.com>

EOSC (2022). *PID Policy and implementation Task Force (PID TF)*.
<https://www.eosc.eu/advisory-groups/pid-policy-implementation>

EOSC (2023). *EOSC Symposium 2023 (Madrid, 20-22 Sep 2023)*.
<https://symposium23.eoscfuture.eu>

De Castro, Pablo (2023). "The outcomes of the Knowledge Exchange project on PIDs have now been published. The dissemination work is just starting". *Scidecode science consulting*, 3 February.
<https://scidecode.com/2023/02/03/the-outcomes-of-the-knowledge-exchange-project-on-pids-have-now-been-published-the-dissemination-work-is-just-starting>

De-Castro, Pablo; Herb, Ulrich; Rothfritz, Laura; Schöpfel, Joachim (2023). *Building the plane as we fly it: the promise of persistent identifiers*. Knowledge Exchange.
<https://doi.org/10.5281/zenodo.7258286>

FAIRCORE4EOSC (2023). "EOSC Core Components: Research Activity Identifier (RAiD)". FAIRCORE4EOSC.
<https://faircore4eosc.eu/eosc-core-components/research-activity-identifier-raid>

Freya Project (2019). The PID Forum.
<https://www.project-freya.eu/en/engagement/pid-forum>

Matas, Lautaro; Segundo, Washington; Nóbrega, Tiago; Filho, J. Edilson S.; Mena-Chalco, Jesús (2023). "dARK: A decentralized blockchain implementation of ARK Persistent Identifiers". Open Repositories 2023 (OR2023), Stellenbosch, South Africa, 12-15 June 2023.
<https://doi.org/10.5281/zenodo.8091668>

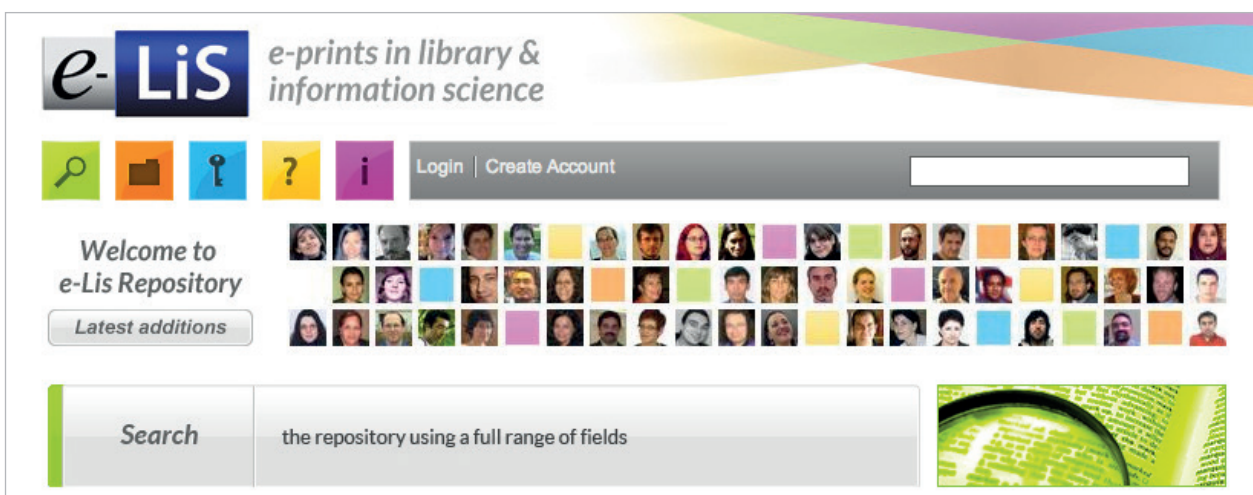
MoreBrains (2022). *The PID-optimised research cycle*.
<https://resources.morebrains.coop/pidcycle>

Porter, Simon J. (2022). Figure 2 "Estimated ORCID adoption and engagement by country". In "Measuring research information citizenship across Orcid practice". *Frontiers in research metrics and analytics*, n. 7, 779097.
https://www.frontiersin.org/files/Articles/779097/frma-07-779097-HTML/image_m/frma-07-779097-g002.jpg

RDA (2022). "Persistent identification of instruments WG". Research Data Alliance.
<https://www.rd-alliance.org/groups/persistent-identification-instruments-wg>

Walker, Bridget (2023). "RDA national PID strategies guide and checklist". Research Data Alliance.
<https://rd-alliance.org/group/national-pid-strategies-wg/outcomes/rda-national-pid-strategies-guide-and-checklist>

Give visibility to your work by depositing it in e-LIS, the largest international repository on library & information science, and communication



<http://eprints.rclis.org>