

Mesa 4: Infraestructuras tecnológicas para gestión de datos de investigación

LUIS ELIÉCER CADENAS MARÍN

Director Ejecutivo de Red CLARA (Venezuela)

MANUELA KLANOVICZ FERREIRA

Analista de Sistemas del Centro de Procesamiento de Datos, Universidade Federal do Rio Grande do Sul (Brasil)

LAUTARO MATAS

Secretario Ejecutivo y Técnico LA Referencia (Argentina)

Moderadora: **MARISA R. DE GIUSTI**

Universidad Nacional de La Plata (Argentina)

RESUMEN EXTENDIDO

¿Qué infraestructuras de datos y a qué niveles considera importante desarrollar en un país: regional, nacional y/o institucional? ¿Por qué? ¿A qué nivel se deben manejar los grandes volúmenes de datos generados?

LUIS ELIECER CADENAS (Venezuela): Es una pregunta compleja, diría yo, que tiene más de una respuesta, y depende eventualmente de muchos factores y también del nivel, tanto geográfico como de alcance que se piensa, como las disciplinas sobre las cuales estamos tratando de manejar o utilizar estos datos. Desde Red CLARA he tenido la oportunidad de interactuar con múltiples iniciativas globales que están trabajando el tema de los datos científicos, en el marco de lo que se ha denominado «nubes de ciencia abierta» y que hay ya varias experiencias y esfuerzos a nivel global. Eso incluye, por ejemplo, la experiencia europea, la European Open Science Cloud, incluye el proyecto africano de Open Science Cloud, y múltiples iniciativas que buscan generar

consensos en términos de cómo manejar los datos de una manera que puedan hacerse interoperables y podamos sacar el mayor provecho en términos de inversión que hacemos en la región. El elemento importante a considerar es lo que representan esos datos en términos de inversión por parte de gobiernos en toda nuestra región y la relevancia que tiene les saquemos el mayor provecho posible. Yo pienso que en término de los niveles, debe haber infraestructura a todos los niveles, debe haber infraestructura local, dependiendo y justamente por las características de desarrollo tecnológico o de digitalización que tiene Latinoamérica y la gran variedad de niveles, conocimientos y capacidades que tienen las propias universidades, tenemos que ayudar a que esto ocurra de una manera más o menos generalizada y eso supone crear mecanismos de apoyo y soporte a esas instituciones para puedan fomentar y desarrollar su infraestructura de datos. Las grandes universidades las tienen, obviamente, y tienen resuelto este problema, pero queda una enorme cantidad de instituciones que eventualmente no conocen, no saben o no tienen idea de cómo mantener una infraestructura de esta naturaleza. Creo que hay que pensar en lo local, en donde lo más apropiado entiendo que es una integración de los datos ascendente, es decir, de lo más bajo hacia lo más alto en términos de integración; buscar mecanismos de interoperabilidad y coordinación para poder construir una visión de esos datos en una manera distribuida a nivel regional y yo pensaría, a lo menos, en una instancia que permitiera integrar esos datos a nivel latinoamericano. Construir los mecanismos para ayudar a esas instituciones pequeñas o que no conocen de este tema para que puedan fácilmente comenzar a almacenar y manejar sus datos de forma inteligente. En las instituciones grandes, dependiendo de las disciplinas hay mucha diferencia, o sea, no es lo mismo los datos que manejan en un repositorios los físicos de altas energías, como el CERN, que los datos que puede necesitar alguna de las ciencias sociales, ni la manera en que eso se almacena y se utiliza. Por eso decía que la pregunta es muy compleja, porque es un espacio muy fragmentado en todos los sentidos: en el sentido de las estrategias de las organizaciones están buscando solventar esta problemática o apoyar en mejorar el manejo de los datos, y después hay un componente que tiene que ver con estas

capacidades técnicas y humanas en cada una de las instituciones. Hay un componente que tiene que ver con las características de la ciencia sobre la cual o de la cual se están almacenando los datos. Es decir, es muy muy complejo y yo creo que hay que ser flexible en términos del modelo que se implemente y tener una especie de caja de herramientas con muchas alternativas para poder atender cada necesidad en el caso, en el contexto, con las condiciones y con las características que haga falta. Tratar de tener esta caja de herramientas lo más completa posible; seguramente Lautaro comentará luego alguna de las cosas que estamos haciendo conjuntamente la Red CLARA, LA Referencia, el Sistema de Cómputo Avanzado para América Latina y el Caribe (SCALAC) para crear un ecosistema que efectivamente apoye este manejo de los datos de una manera eficiente. Para concluir mi respuesta, quisiera comentar solamente la importancia de que miremos también las tecnologías emergentes de transformación digital y el impacto que puede tener el uso de tecnologías como la inteligencia artificial o las tecnologías de block-chain en el proceso de investigación y en el mejor aprovechamiento de los datos científicos. Hay una revolución alrededor de la capacidad para producir ciencia, gracias a la capacidad para gestionar y procesar los datos, y creo que esos componentes debemos tomarlos en consideración porque representan un cambio o pueden impactar de manera muy importante la competitividad o la capacidad productiva de nuestros investigadores en la región. Estamos ante un gran reto. Yo creo que están confluyendo muchos intereses, muchos actores, hay mucha cooperación internacional buscando fortalecer los sistemas científicos; hemos vivido en el marco de una pandemia que ha empujado una concepción y un entendimiento mucho mayor de lo que significa la ciencia y yo creo que es momento de ponernos todos de acuerdo y empezar a construir colectivamente, con flexibilidad, sabiendo que no vamos a tener una solución que resuelva todo y que vamos a tener que tener alternativas que puedan ayudar en los distintos tipos de problemáticas. Entonces creo que tiene que ser local, a nivel de la institución, nacional para integrar lo que tiene que ver con las instituciones, y tiene que ser latinoamericano y probablemente global, si queremos estar inmersos en lo que es la ciencia global y sus potencialidades y sus logros.

MANUELA KLANOVICZ FERREIRA (Brasil): Conuerdo con algunas de cosas que dijo Luis. Aquí, en Brasil, todas las infraestructuras, tanto locales como regionales como institucionales, son las tres importantes. Los investigadores tienen varias formas de trabajo, dependiendo del área de investigación, y muchas veces dentro de la misma área proyectos con recursos distintos trabajan de forma diferente. Entonces los tres niveles son importantes. Por ejemplo, en la infraestructura local o institucional, hemos tenido oportunidad de promover la preservación durante la investigación; es decir que los datos antes de ser publicados ya pueden ser preservados por los investigadores y esos datos ya tienen una curaduría, una procedencia documentada, de forma que cuando llegue el fin de la investigación y se proceda a publicar un trabajo y disponibilizar un conjunto de datos referente a ese trabajo, de modo que puedan ser reutilizados, ya tienen una infraestructura de preservación de esos datos, de modo que no sea tan trabajoso a la hora de compartirlos. Existen muchos proyectos, y hay ya muchos conjuntos de datos, que no tienen un gran tamaño, pero sí hay una gran cantidad de ellos; generalmente pertenecen a proyectos menores, a grupos de investigación menores, que no tienen tantos recursos. Para esos investigadores, lo ideal sería preservar los datos de inmediato. Pero también hay proyectos grandes, como los del CERN, que hoy citó Luis, u otro tipo de proyectos con datos médicos, por ejemplo, es importante que estén en repositorios internacionales y grandes, pues de ese modo se pueden preservar y optimizar mejor, además de que pueden ser entregados de acuerdo a la necesidad del usuario. Todo esto es diferente dependiendo del formato de los datos y del área de investigación a que correspondan. Esto muestra también la importancia de los repositorios internacionales dedicados a un área específica, porque uno de los mayores problemas que se enfrentan es la diversidad de datos y formatos. Las infraestructuras deben servir a la mayoría de los investigadores pero muchas veces no se consigue «customizar» lo suficiente para ello, para que la preservación se torne más automatizada. Entonces, concluyendo, los tres niveles son importantes: el internacional, el nacional, porque por ejemplo aquí en Brasil existen algunas cuestiones respecto de colocar los datos en la nube,

porque anteriormente la legislación sólo permitía colocar datos en nubes locales, no internacionales, y asimismo las iniciativas nacionales, sobre todo en términos de recursos, ya que las universidades aquí son federales y muchas veces el gobierno no ve la importancia de destinar recursos a estas iniciativas.

LAUTARO MATAS (Argentina): Con respecto a la pregunta, tomo todo lo que dijeron Luis y Manuela, los tres niveles son importantes, pero voy a patear un poco el tablero con algunas cosas. Sí, esa es la solución ideal, pero no es la solución que vamos a tener, al menos no es la solución que vamos a tener en la mayoría de los países de nuestra región. Si vemos lo que ocurre con los repositorios de literatura, vemos que aún en los repositorios de literatura, mayormente con instalaciones de DSpace, tenemos instalaciones muy antiguas en gran cantidad de instituciones, instituciones que no tienen recursos para mantenerlas actualizadas y que ya es un desafío para muchas áreas, especialmente la de las bibliotecas con sus recursos, mantener eso. Entonces, para ser realista, sería ideal que todas las instituciones puedan tener un repositorio de datos a disposición de sus investigadores para el depósito. Pero eso no va a ocurrir o al menos no va a ocurrir en el mediano plazo. Entonces, los niveles superiores, el nacional y el regional, y quiero hacer un matiz respecto de internacional y regional: digo regional porque por supuesto que depositar hoy en Zenodo es una alternativa a falta de otra cosa, pero es importante que pensemos que al final del día vamos a querer tener muchos esos datos en una infraestructura regional, que todavía no existe y que en un momento voy a comentar un poco qué estamos pensando junto con Red CLARA para intentar construirla. Luego está el nivel nacional y en el nivel nacional, a veces se vuelve a repetir el mismo problema en el nivel institucional. Mantener un repositorio, crear un repositorio de datos para grandes volúmenes de información, como bien dijeron mis compañeros de mesa, es un desafío para muchas disciplinas, y a veces es un desafío que puede exceder las infraestructuras nacionales, incluso de muchos países con los recursos que tienen para investigación, pensemos en los países pequeños y también por los volúmenes de producción que tienen. Y esto con las universidades también pasa: hay instituciones pequeñas que, por el volumen de producción que tienen, hay que ver si tiene

sentido en todos los lugares tener esos dos niveles. Es una discusión que no es mía solamente, sino que la estamos teniendo como países, cuando ellos se ponen a desarrollar en el marco de LA Referencia y nosotros participamos en las discusiones a nivel nacional de algunos países, de qué van a hacer con su infraestructura de datos, es algo que estamos discutiendo: ¿vale la pena empezar con un repositorio nacional o juntar esfuerzos de universidades desde un repositorio que compartan y entonces puedan compartir recursos y tener una infraestructura más confiable? Porque no se olviden que aquí el tema es tener una infraestructura confiable: hay algo peor que no tener esto, que es tener algo malo y que los investigadores no confíen eso o que pierdan sus datos o lo que sea. Eso sería peor que no tenerlo porque nuevamente lo más importante aquí es el cambio cultural en los investigadores para que ellos depositen y reutilicen los datos de su investigación, entonces tenemos que crear infraestructuras confiables, tenemos que ordenar también, a nivel nacional e institucional, el mensaje que le damos a los investigadores y no estar cambiando de estrategias entre medio de uno u otro. La organización al fin tiene que ser nacional, tiene que haber una buena relación entre eso y las decisiones institucionales y tiene que haber políticas nacionales. Nosotros, a nivel regional con LA Referencia y con Red CLARA, estamos intentando organizar toda esta discusión, por un lado para articular las distintas políticas en los distintos países, y aprovechar las experiencias de los que, como Brasil, están avanzando más rápido que otros en este problema, entonces acercar y tratar de hacer redes de colaboración para que haya transferencia de conocimiento de los países que están más avanzados hacia los que están empezando. Y a nivel regional, esta idea que estamos trabajando con Luis y con el equipo de Red CLARA, que es la de ir hacia un ecosistema de ciencia abierta en la región y una parte fundamental de ese ecosistema de ciencia abierta es tener una infraestructura de depósito de datos. Esa estrategia no es para reemplazar los otros niveles, sino que es para hacer una infraestructura regional que esté a disposición de los países y de los investigadores cuando no tengan uno en su institución. La estrategia será trabajar juntos como países para consolidar una infraestructura regional, buscar financiamiento para eso,

que no es algo fácil, porque nosotros no tenemos algo como tiene OpenAIRE o Zenodo, lo mismo con el CERN, que tienen financiamiento europeo y fuentes de financiamiento regional. Nosotros nos tenemos que basar o en financiamientos nacionales en colaboración o como contraparte de financiamiento internacional de otros fondos europeos o del BID, o de otras iniciativas, entonces eso es un trabajo que tenemos que hacer. Paralelamente, por supuesto, discutir y consolidar a nivel nacional las estrategias y ser claro con los investigadores, y en las instituciones trabajar en tratar de ver qué tiene sentido, qué se puede sustentar y qué se puede crear con el tiempo. Siempre tiene que haber una red de contención para el siguiente nivel, es decir, si el nivel institucional no existe, tiene que estar el nivel nacional, y en los países en que ese nivel no existe, tenemos que nosotros trabajar juntos para construirlo a nivel regional y tratar de mantener los datos dentro de nuestra región.

¿Qué alternativas de cooperación se están planteando en su país y para América Latina en su conjunto (tecnológicas, políticas, compras consorciadas de identificadores para datos y metadatos, acuerdos)?

L. E. CADENAS: Sí, hay varios temas que se vinculan con el comentario que hizo previamente Lautaro de las dificultades que tenemos en la región, porque las fuentes de financiamiento no son tan abundantes como lo son en las otras regiones de la Tierra. Sin embargo, Red CLARA ha logrado, por un apoyo sostenido de la Comisión Europea desde su fundación, construir una infraestructura muy relevante en términos de magnitud, con conectividad entre los países, que constituye una base importante para poder generar esta integración regional. LA Referencia, por su parte, tiene la representación institucional de los gobiernos que participan en la red y esos son espacios que efectivamente creo que podemos utilizar para promover iniciativas que sean colectivas y regionales. Creo que es lo más difícil de hacer: a nivel nacional, por supuesto, dependiendo de los gobiernos y de las orientaciones y de la percepción que se tiene de la ciencia y de la ciencia abierta puede que cambien las políticas y sean muy diversas. Creo que a nivel regional, las organizaciones que están trabajando en esto están bastante claras y tenemos la oportunidad,

porque tenemos infraestructura construida que permite efectivamente tener repositorios de datos que pueden ser usados por muchos países a nivel regional y que puedan conectar estos datos con datos en otras partes del mundo. Recientemente, ustedes podrán haber visto que inauguramos la conexión directa que ahora tenemos con Europa con el cable submarino, y eso no va a dar una posibilidad de integración muy importante también con las iniciativas europeas en temas de ciencia abierta. Creo que es una labor que va tomando tiempo, pero quizás podamos conseguir hacer confluir los intereses y las voluntades no solo de los gobiernos, sino de las agencias internacionales de cooperación, que están siendo muy importantes en términos de impulsar esta visión, esta necesidad de contar con datos abiertos y de contar con una visión de ciencia abierta a nivel global. Eso ustedes lo conocen mucho mejor que yo, pero todo lo que tiene que ver con la resolución de la UNESCO alrededor del tema de ciencia abierta, que esperamos que esté refrendada (que no sé ni siquiera si ya está refrendada o no, creo que no, pero debe estar por estarlo), eso genera un marco institucional y un marco político muy relevante a nivel global, con una cantidad enorme de países respaldando la necesidad de tener este enfoque de ciencia abierta. En los cuatro pilares planteados por el documento de la UNESCO, está el tema de las infraestructuras, que es un tema que tenemos que reforzar enormemente. Lautaro mencionaba la necesidad de que los datos estén en la región y yo coincido plenamente. Recientemente, la UDAT publicó su reporte sobre economía de este año, un reporte muy interesante que refleja, principalmente, todo lo que tiene que ver con los datos y de qué manera los datos están siendo dominados o manejados por básicamente cuatro o cinco empresas a nivel global, que tienen toda la capacidad de almacenamiento, de gestión, de procesamiento y yo creo que hay que lograr un equilibrio, ningún desequilibrio es sano; yo creo que hay espacio para los servicios de nube privados o comerciales, pero también es necesario tener políticas de datos en los países, que garanticen que los datos propios son manejados de una forma adecuada. En ese mismo reporte se hablaba de que el 95 % de la innovación de las *startups* en inteligencia artificial están en Estados Unidos y en China. El otro 5 % está en el resto del mundo incluyendo muchos

otros países que son desarrollados, pero que no tienen el mismo nivel de desarrollo en términos de esta tecnología. Eso va a generar un gran desbalance y vamos a necesitar un enfoque global y un enfoque regional alrededor del tema. Red CLARA participa en muchas instancias de cooperación, obviamente las instancias que tienen que ver con cooperación con Europa, participamos en instancias que tienen que ver con la cooperación en redes de investigación y educación a nivel global, estamos hablando de más de 140 redes; participamos en organizaciones donde podemos llevar este mensaje, esta necesidad y esta idea de poder construir colectivamente un espacio de datos para los investigadores, no sólo a nivel de Latinoamérica, sino a nivel global. Yo creo que hay que tomar todas esas instancias, que están allí, que están además trabajando colectivamente para lograr manipular cada vez más datos enormes. Por ejemplo, en el ámbito del cambio climático: es un ámbito que está permitiendo organizar mucho, cómo estructurar y manejar datos en una determinada área, por la relevancia que esto tiene para temas de monitoreo la Tierra o para tomar decisiones. Todos los esfuerzos que hay al nivel de Naciones Unidas en términos de los datos, de las estadísticas y de las informaciones, etc. Creo que hay que juntar todo esos esfuerzos y buscar los espacios necesarios. Ahí está la disposición de las organizaciones que tenemos en la región: LA Referencia es un mecanismo enorme y muy importante para coordinar las acciones políticas o por lo menos plantear ideas en el marco de acciones políticas regionales dentro de los espacios, los ministerios de ciencia y tecnología en la región y tenemos que aprovecharlos todos, porque realmente el reto es extremadamente grande, los recursos son pocos, y por eso es indispensable que tengamos una estrategia colaborativa, una estrategia donde nos sumemos unos con otros y podamos complementar lo que no tenemos a través de financiamiento estructurado a nivel regional. Hay que buscar todos los espacios abiertos para esta conversación, los distintos grupos que existen; por ejemplo, participé hace poco en un evento de la Alianza del Pacífico: ese es un ámbito en el cual hay que plantear el tema y decir hace falta infraestructura, hace falta financiamiento, hace falta una visión común en el marco de todas las instancias, tanto políticas como técnicas como de todos

los niveles que tenemos en la región, hay que llevar este mensaje y defender la idea de la ciencia abierta y la importancia que tiene para nuestra región y la relevancia que tiene que tengamos dominio sobre nuestros datos, que decidamos qué podemos poner donde queramos y qué queremos tener nosotros, porque es indispensable para nuestro desarrollo.

M. KLANOVICZ FERREIRA: Luis mencionó varios puntos relacionados con la Red CLARA, y creo que estas iniciativas son muy importantes porque hay instituciones que tal vez no tienen capacidades o infraestructura y hay que conseguir promover la preservación de los datos localmente, para tener cierta uniformización entre países distintos, así como promover la importancia de la ciencia abierta y promover incentivos en esa área. Aquí en Brasil, existen algunas iniciativas de cooperación, en las cuales estamos participando y procurando recursos pero el objetivo de uno de los consorcios es la infraestructura para lograr un ecosistema de investigación colectiva. Inicialmente, este consorcio sería utilizado para comprar DOI, a partir de Data Cite, porque de este modo existen algunos descuentos importantes. Este sería un primer paso para incentivar la preservación de los datos. Esto sería sólo el comienzo, la idea es que haya más asociaciones colaborativas que compartan sus recursos. Otra iniciativa que existe en Brasil ha hecho un relevamiento de la situación de los los repositorios de datos del país y de sus necesidades. A partir de eso este año fue lanzado un plan para las instituciones que no tienen infraestructura suficiente y durante un período de incubación de nueve meses se les brinda capacitación y transferencia de recursos. Estas son dos iniciativas nacionales que veo muy importantes y necesarias para que se dé la voz de largada de los repositorios de datos; todo esto es fundamental para la comunicación de la ciencia, aunque a veces se vea como una «obligación» más para los investigadores, ya que deben también depositar los datos. Es decir, que estas iniciativas son importantes para que haya una concientización de la relevancia de los datos, de cómo teniendo los datos disponibles de una investigación anterior, teniendo datos confiables y pudiendo reproducirla, así se evolucionará más rápido y podremos llegar, como sociedad, en Brasil y en el mundo, a nuevos descubrimientos útiles. Esa concientización entonces es muy

importante y estas iniciativas, en mi opinión, sirven no sólo para brindar infraestructura sino para esto, si todo el mundo lo está haciendo, por qué no vamos a hacerlo; es preciso convencer a los dirigentes de que esta es una aplicación adecuada de los escasos recursos con que se cuenta.

L. MATAS: Bueno, lo bueno y lo malo de hablar después de gente tan completa es que quedan pocas cosas que decir. Hablando en serio, más allá de los puntos específicos en los que estoy de acuerdo, aquí es una actuación conjunta la que estamos haciendo. Entonces muchas de las cosas que yo diría, son las que ha mencionado Luis. Lo que quiero remarcar es que, como decía antes, sí, estamos muy activos buscando esta integración regional y estamos complementando la experiencia de Red CLARA en los proyectos y también el capital simbólico que tiene Red CLARA frente a las iniciativas europeas y además con las grandes inversiones que han hecho en términos de conectividad en la región, con este rol que tenemos nosotros es de poder poner en una mesa los organismos de ciencia y tecnología en temas de ciencia abierta a discutir y a consensuar políticas y también estrategias y ojalá a poder consensuar también con otras iniciativas de la región en trabajar juntos en un proyecto que nos permita crear y financiar estas infraestructuras. Nosotros tenemos que crear artificialmente esto que otras regiones tienen como dado, que es el financiamiento conjunto para ser usado en este tipo de iniciativas. Esta articulación que estamos haciendo con Red CLARA desde lo cooperativo, que para nosotros es muy fácil porque estamos muy relacionados y confiamos mucho en esa experiencia, pero, por otra parte, la opinión de otros países y esa mesa de discusión esperamos que nos permita crear uno o varios proyectos lo suficientemente robustos para poder conseguir el financiamiento que necesitamos. Sobre todo el segundo desafío que viene, porque a veces hablamos de financiamiento para conseguir montarlo, pero el problema en sí no es montarlo, el problema es sostenerlo. Ahí es donde está el mayor desafío a largo plazo porque nosotros, con suerte, podríamos montar una infraestructura compartida o distribuida en varios nodos, de distintas cosas, que para mí sería lo ideal, tratar de hacer replicación y crear algo que sea realmente distribuido, para que la preservación sea confiable y, de nuevo, este tema de darles a los

investigadores algo confiable donde depositar. Estuvimos discutiendo la posibilidad de hacer un Zenodo regional, que se pueda hacer algo en forma distribuida, y sea algo parecido a lo que es Zenodo; podríamos solucionar el problema de la soberanía de los datos que en algunos países, como decía Manuela, en términos de las legislaciones obligan a tener los datos, eso se puede así, pero replicado en otros lugares para seguridad. El problema es como sostenemos: hay que comprometer a los Estados en darle sustentabilidad a eso y yo creo que ahí el rol de LA Referencia, el rol de interlocución que podemos tener, tiene que ir en ese sentido, tiene que ir en el sentido de convencer, de convocar y de generar un ámbito, un fondo para luego mantener esto funcionando. Es un trabajo a muy largo plazo, un trabajo que ya estamos haciendo, pero que nos va a llevar tiempo y que es un trabajo siempre de mediano y largo plazo. En el ámbito de los datos, el tema de la infraestructura nos va a ocupar los próximos cinco o diez años de consolidar y llegar a un punto comparable al que otras regiones están en estos términos. Pero hay que hacerlo y lo estamos haciendo ya.

¿Cómo se deben pensar las plataformas de publicación para integrarlas con infraestructuras de datos? O sea, ¿cómo se adaptan los circuitos de publicación, qué protocolos de interoperabilidad se utilizan, qué servicios deben brindar los sistemas de publicación, etc.?

L. E. CADENAS: Yo tengo que hacer un *disclosure* aquí: de eso conozco muy poco realmente, con lo cual no me siento competente para responder apropiadamente. Sí puedo decir, en términos de los servicios, que mucho de lo que se está desarrollando en el marco de estas iniciativas que mencioné previamente, es lo que hace falta: tener garantías de que el investigador pueda manejar el ciclo completo de los datos y vincularlos con sus publicaciones. Creo que hay elementos importantes que considerar, por ejemplo la seguridad alrededor de los datos por una parte, y por la otra, la integridad y mantenimiento de esos datos, de manera que haya una garantía de que lo se publica y los datos en los cuales se basa la publicación sean preservados adecuadamente en el tiempo y no cambien y puedan ser reproducibles, a los efectos de esa

capacidad de reproducibilidad que queremos buscar a partir de este tipo de herramientas. También apoyar el manejo de las aplicaciones o los software para el desarrollo y para el procesamiento de esos datos es un elemento que es importante y la posibilidad para el investigador de definir los niveles de apertura y la forma en que desea manejar sus datos en el marco de las publicaciones que está proveyendo.

M. KLANOVICZ FERREIRA: Uno de los grandes problemas en relación con los datos, su reuso y la reproductibilidad de las investigaciones es que los datos estén documentados de forma adecuada. En la actualidad, existen los **principios FAIR**, que tienen como objetivo volver los datos encontrables, accesibles, interoperables y reutilizables. A partir de estos objetivos, los repositorios de datos deben implementar acciones para que sean alcanzados. Una de las formas para hacer eso, y que forma parte de mi investigación, por ejemplo, es la documentación automatizada de datos. Documentación relacionada a los datos o bien relacionada a las investigaciones en las que se obtuvieron esos datos. En la actualidad, gran parte de las investigaciones son llevadas a cabo mediante computadoras, pero no todo el mundo tiene la capacidad de describir las herramientas con las que se obtuvieron los datos o el modo en que un experimento fue ejecutado, y asimismo documentar la procedencia de los datos, qué días fueron tomados, en qué horario, qué entidades estuvieron involucradas en su recolección, ¿son confiables estos datos? Hoy día existen iniciativas en proyectos grandes, como en el CERN, que disponen de herramientas para que toda esta información sea guardada junto con los datos. Estas son herramientas que le facilitan la tarea al investigador, porque si un dato no fue bien documentado, si su procedencia no está documentada, otros investigadores no lo van a poder utilizar. Entonces es importante proveer de estas herramientas a los investigadores que hoy tienen dificultades para documentar correctamente los datos. Otra cosa importante es incentivar, y desde luego que todo depende de los recursos, la preservación de los datos dentro de la publicación. Muchas veces los investigadores necesitan publicar los datos porque se los exigen para poder enviar el trabajo a una publicación. ¿Por qué los editores quieren los datos? Porque así pueden

revisarlos antes de la publicación y esa revisión, siempre que sea posible, aunque es difícil, es muy importante, y teniéndolos bien documentados se facilita mucho más esa revisión. Esa revisión incentiva, a su vez, la calidad de los datos, siempre teniendo presentes los principios FAIR. No es sólo una cuestión de verificación de la investigación, sino que es una cuestión de incentivos a la investigación. Un tercer punto que me gustaría comentar, muy ligado a la pregunta anterior, es la evaluación de los datos, así los investigadores podrían ver cómo los evalúan los usuarios; además de eso, los accesos a los datos, accesos a los documentos, descargas, de forma que el usuario que vaya a usar esos datos consiga ver rápidamente si el dato ya fue usado, si tiene muchos *downloads*, cuántos accesos tiene, si fue calificado, etc. Conforme la utilización de los datos aumenta, se debe facilitar a los usuarios toda esta información contextual y así habrá también un aumento en la compartición de los datos, que puede ser exponencial. Este tipo de iniciativas son, entonces, muy interesantes y en mi opinión deben ser incentivadas por los repositorios.

L. MATAS: Creo que la respuesta de Manuela fue muy buena y interesante y abarca algunas cosas que no pensaba mencionar y que puedo entonces mencionar. En principio, lo importante es la calidad de los metadatos y en este sentido los nuevos estándares, ya no son tan nuevos, pero nosotros todavía estamos empujando para que se implementen, que son los estándares de metadatos OpenAIRE 4, que tienen contemplado poder vincular los datos con las publicaciones, por medio de los identificadores relacionados, especificando que eso es un dataset. Por eso es ideal que donde depositamos los datos, tengan identificadores únicos y así poder citarlos. La calidad, entonces, para que todos esos metadatos sean encontrables, reutilizables, la calidad con la que son depositados, los lugares donde son depositados y cuán bien curatos están esos datos y también el orden que tienen los investigadores para poder después en las publicaciones citarlos correctamente y vincularlos. En ese sentido, lo que nosotros tenemos en LA Referencia, es un grupo de trabajo, como decía antes, para intentar que todas esas implementaciones, la migración y la migración hacia atrás, que no es nada fácil, ocurra, para todo eso lo que

nosotros tenemos es la nueva versión del cosechador; también estamos trabajando en la posibilidad de hacer migraciones entre esquemas de metadatos de lo que tenemos de herencia en OpenAIRE 3 para OpenAIRE 4. Queremos poder propiciar eso y poder dar un servicio a los repositorios, darles una versión enriquecida de los datos en otro esquema que no tienen para que no hagan ese esfuerzo manualmente. Eso es un sueño que tenemos y la tecnología que tenemos y que hemos desarrollado en los últimos dos años nos lo permite hacer, hay que organizarlo y lo vamos a hacer. Por otro lado, me disparó una de las cuestiones que mencionó Manuela, la relación con un proyecto en el que estamos trabajando y mirando con mucha atención con Luis en el ámbito de Red CLARA también y con otros socios que es el proyecto de MyLab: que los investigadores y los grupos de investigación tengan un ambiente donde trabajar ordenadamente y que queden registrados los contextos donde la investigación es hecha. En las presentaciones de LA Referencia yo a veces digo que nosotros empezamos ocupándonos de la literatura, de los repositorios, pero que consideramos como producción científica no solamente a la literatura y los datos, sino a toda la información contextual que forma parte de los procesos de investigación y de generación de ese conocimiento. Entonces me pareció muy bueno lo que dijo Manuela, porque la reproductibilidad, que es la condición necesaria para ser ciencia, no es solamente tener el artículo y tener los datos, sino tener el contexto en que ese experimento o esos datos fueron generados. Lo de MyLab nos lo acerca, en el sentido de que ellos están trabajando en una herramienta para que los grupos de investigación puedan preservar hasta las discusiones que ocurren en el proceso de investigación. Lo tomo hoy, porque me parece que es algo interesante para profundizar y para poner muy de relevancia, en nuestro trabajo del próximo tiempo. Eso es lo que estamos haciendo, son muchas cosas y son muchos frentes, pero estamos trabajando un poquito en cada uno y tratando de avanzar como comunidad, en esto lo importante es avanzar sin prisa y sin pausa.