|  |
| --- |
| **Ficha de necesidad preliminar** |
| **RETO: Enodatalake. Reutilización y valorización de los datos generados por la cadena de valor del sector vitivinícola** |
| **ENTIDAD COORDINADORA: Dirección General de Universidad y Política Científica** |

|  |
| --- |
| 1. Descripción - situación de partida:

El proyecto Enodatalake tiene por objetivo la reutilización y la valorización de los datos generados por la cadena de valor del sector vitivinícola con el fin de generar productos y servicios que incrementen su productividad y competitividad.A modo de resumen, Enodatalake pretende:* Reutilizar y revalorizar los datos generados por la cadena de valor del sector vitivinícola.
* Extender estándares europeos de compartición de datos entre el sector empresarial vitivinícola y agentes relacionados.
* Potenciar proyectos de generación, transformación y almacenamiento de dato en los agentes del sector e impulsar la computación en nube, *edge* u *on premise* en las empresas.
* Imponerse como marco de referencia para la compartición de datos del sector vitivinícola en La Rioja, siguiendo las estrictas exigencias funcionales y técnicas descritas en GAIA-X, en campos tales como:
	+ Gestión y verificación de Identidad
	+ Gestión de la Soberanía del dato: Protección del dato, control y gestión en tiempo real de políticas de uso y acceso
	+ Acceso a los datos de los agentes simplificado, homogéneo e integrado en formatos y estándares ampliamente utilizados en la industria en el Marketplace.
	+ Catalogación y categorización de los datos de acuerdo a las exigencias de Autodescripción.
	+ Certificación en GAIA-X del espacio de datos
* Atraer a *startups* y empresas innovadoras basadas en reutilización y revalorización de dato para que construyan servicios sobre los mismos. Se impulsará que los datos generados y enriquecidos por los nuevos servicios sean a su vez almacenados y publicados a través de los estándares de Enodatalake.
* Proponer y asesorar a los consumidores y proveedores de datos las mejores opciones tanto a nivel infraestructura como software para tomar el máximo partido de los datos y garantizar la integración.

2. Alcance del proyectoCon este objetivo, Enodatalake implementará un repositorio de datos y servicios al que agentes del sector y otras empresas tendrán acceso con el fin de reutilizar y revalorizar los datos, para finalmente generar nuevos servicios que den soluciones a desafíos en el ámbito de la vitivinicultura. Es un requisito básico del sistema la interoperabilidad y, entendiendo la apuesta clara de España y la Unión Europea por la economía del dato y su homogenización, seguirá las más estrictas recomendaciones funcionales, en diseño y técnicas para garantizarla. Para ello, el equipo del proyecto se apoyará continuamente en las recomendaciones de la Oficina del Dato y el *Hub* Nacional GAIA-X, ambos supervisados por la Secretaría de Estado de Digitalización e Inteligencia Artificial. Sabiendo que ambos entes comenzarán a trabajar en repositorios de datos sectoriales en otros ámbitos, como es el caso del *hub* con el repositorio de datos del sector turístico, el equipo se beneficiará de las buenas prácticas y metodologías utilizadas para llevar a cabo todo el proceso.El proyecto comenzará su despliegue con el desarrollo de una plataforma para el almacenamiento, gestión y tratamiento de datos que podrán ser utilizados para el desarrollo de trabajos de investigación a la que se añadirán todos los datos que ya están accesibles. La Comunidad Autónoma de La Rioja dispone de conjuntos de datos separados relativos al campo de la agricultura y la viticultura. Uno de los objetivos es incorporar dichos datos al espacio sectorial de datos para que puedan ser alojados y consultados de forma uniforme. Una explicación más detallada de dichas fuentes de datos viene detallada en el Anexo I.  La investigación actual genera y procesa una ingente cantidad de datos de muy diversas tipologías y procedencias, especialmente debido a la irrupción, entre otras cuestiones, de las nuevas tecnologías "ómicas" (genómica, trasncriptómica, proteómica, metabolómica). Las líneas de investigación en campos tales como, entre otros, la genética vegetal, la microbiología, la biotecnología, la agricultura de precisión, o la nutrición y la salud humana, requieren de infraestructuras tecnológicas que den respaldo a las necesidades de cálculo, generación de información, análisis, almacenamiento, y procesado de datos masivos mediante técnicas Big Data orientadas a la búsqueda, procesado y carga de datos, debido a su dimensión y complejidad.  Por este motivo, se abordará la inversión en un espacio de datos sectorial que facilite a todos los investigadores y técnicos, infraestructura, servicios y productos orientados a alta capacidad de cálculo y herramientas de análisis de datos masivos, para aplicaciones en diversos ámbitos de investigación y para la inclusión de modelos de Inteligencia Artificial en sus estudios.  A su vez, todos los datos de partida y generados por la investigación serán volcados, almacenados y compartidos permitiendo la creación de un repositorio abierto de datos científicos de la vid y el vino, que permita trabajar en red a los investigadores y técnicos y, al mismo tiempo, compartir sus resultados con terceros tales como universidades, centros de investigación, centros tecnológicos, *startups* y todos los agentes de la cadena de valor del sector vinícola.  Con el fin de garantizar la mayor interoperabilidad, y de acuerdo con la apuesta clara de España y la Unión Europea por la economía de dato y su homogeneización, el repositorio seguirá las recomendaciones funcionales técnicas y funcionales del marco de referencia GAIA-X, apoyados continuamente por la Oficina del Dato y el *Hub* Nacional GAIA-X, ambos impulsados por la Secretaría de Estado de Digitalización e Inteligencia Artificial. De esta forma, el repositorio extenderá los estándares de GAIA-X a lo largo de los investigadores y agentes involucrados en la cadena de valor del sector vinícola, consiguiendo uniformidad en lo referente a Gobernanza del Dato y maximizará la integración con otros espacios de datos sectoriales nacionales y europeos, creando posibles sinergias con otros generadores de datos y creando una futura red de intercambio permanente de información estructura e interoperable. Es importante resaltar que este espacio sectorial de datos será fundamental para, entre otras cosas, poner en marcha servicios de Bioinformática y Bioestadística, ya que permitirá el tratamiento y almacenamiento de datos genéticos, de modelado molecular y epidemiológicos, relacionados con diferentes organismos de interés en viticultura y enología, como el material vegetal (genética de variedades), los microorganismos (fitopatógenos, levaduras de vinificación), y humamos (percepción sensorial, estudios epidemiológicos de nutrición y salud) Los resultados esperados son:* Facilitar herramientas útiles de análisis de datos en los procesos de generación de conocimientos
* Generar y estructurar la información necesaria en la toma de decisiones por parte de los distintos agentes del sector vitivinícola
* Incorporación de bases de datos y series históricas necesarias en la modelización y predicción de futuros escenarios
* Intercambio de datos y metodologías de análisis de los mismos con otros centros de otras Comunidades Autónomas con similares características, así como del ámbito internacional.
* Reutilización y revalorización de los datos generados por la actividad investigadora en el ámbito de la vid y el vino
* Extensión y homogeneización del modelo de gobernanza y compartición de datos a través de los agentes del sector
* Atracción de *startups* y empresas basadas en innovación de dato que construyan nuevos productos y servicios sobre los datos de investigación del sector.

La existencia de casos de uso permitirá la incorporación de otras entidades de la cadena de valor al proyecto, sobre todo del ámbito industrial. Estos agentes serán al mismo tiempo proveedores de datos y clientes de los diferentes casos de uso que se generen.A modo de resumen, los principales hitos a cumplir son:* Repositorio de datos de investigación.
* Repositorio de datos públicos de interés para la toma de decisiones por los diferentes agentes del sistema, o para el diseño de nuevos casos de uso.
* Incorporación de 15 agentes referentes de la cadena de valor para incorporar datos de sus procesos desde el campo hasta la copa.
* Se dispondrá de un catálogo de casos de uso y servicios que se irá completando a medida que avancemos en la construcción del Enodatalake. Será necesaria la incorporación de datos de manera permanente, su mantenimiento, la incorporación de agentes y la venta de servicios, por lo que se dotará asimismo de personal suficiente para llevarlo a cabo.
* Determinación del dimensionamiento y elección de las infraestructuras necesarias para albergar el Enodatalake y planes de escalado
1. **Componentes innovadores**
* Técnicas de procesamiento y análisis masivo de información
* Lagos de datos interoperables bajo marco de trabajo GAIA-X
* Innovación en Gobernanza y gestión del dato
* Innovación en seguridad de la información
* Implantación de sistema para Garantía del Soberanía del Dato
* Dimensionamiento, diseño y gestión de Infraestructura Hardware. Plan de escalado.
1. **Fecha prevista de cierre de convocatoria**: 14 de septiembre de 2022.
2. **Presentación de solicitudes**:

La presentación de solicitudes se podrá realizar a través de una instancia general que podrá presentarse de dos maneras:* De forma electrónica en:

<https://www.larioja.org/oficina-electronica/es?web=000&proc=12327>* De forma presencial a través de:
* La Oficina de Registro del órgano competente
* Oficinas de registro

Si desea información sobre otros lugares de presentación de solicitudes, póngase en contacto con el Servicio de Atención al CiudadanoTéngase en cuenta que el artículo 14.2 de la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas, establece la obligatoriedad de la gestión electrónica para determinados sujetos como, por ejemplo, las personas jurídicas1. **Documentación anexa**:
* Anexo I. Fuentes de Datos
 |