El futuro digital es de todos

MinTIC

Diseño de piloto de Data Commons

Hechos QUE CONECTAN

Tabla de contenido

1. Introducción 5

1.1. Descripción general 5

1.2. Objetivos 7

1.2.1. Objetivo general 7

1.2.2. Objetivos específicos 7

1.3. Problema, necesidad, o caso de uso para el intercambio de datos 8

2. Ecosistema: actores y flujos de datos 10

2.1. Actores involucrados 10

2.2. Flujo de datos 12

2.2.1. Estado actual del flujo de datos 12

2.3. Aprovechamiento de datos 18

2.4. Interoperabilidad actualmente 19

3. Solución de Data Commons 20

3.1. Sobre el modelo Data Commons 20

3.2. Data Commons para el SIGRA 21

3.3. Casos similares 24

3.4. Gobernanza del Data Commons 25

3.4.1. Roles y responsabilidades en el Data Commons 28

3.5. Requerimientos técnicos 30

3.5.1. Tecnología 30

3.5.2. Arquitectura técnica 32

3.5.3. Seguridad 35

3.5.4. Equipo: recursos humanos 36

4. Análisis de costos 39

5. Producto Mínimo Viable 40

6. Hoja de Ruta 41

6.1. Enfoque 41

6.2. Cronograma 42
7. Bibliografía 46

Tabla de gráficos

Figura 1. Pasos de la metodología de priorización del caso de uso 6
Figura 2. Actores involucrados según su rol en el ecosistema de datos 11
Figura 3. Mapeo inicial del ecosistema de datos 12
Figura 4. ODI - The Data Spectrum 14
Figura 5. Conjuntos de datos identificados 14
Figura 6. Diagrama arquitectura de red X-Road 20
Figura 7. X-Road y Data Commons 23
Figura 8. Proceso de definición de dinámicas de gobernanza 26
Figura 9. Principios para un Data Commons 27
Figura 10. Roles y responsabilidades sugeridas 29
Figura 11. Arquitectura técnica de referencia 32
Figura 12. Estándares y necesidades de usuario 35
Figura 13. Roles dentro de un equipo para diseñar e implementar el Data Commons 37
Figura 14. Desglose de roles y días requeridos 39
Figura 15. Cronograma 42

Ministerio de Tecnologías de la Información y las Comunicaciones
Carmen Ligia Valderrama Rojas - Ministra de Tecnologías de la Información y las Comunicaciones
Iván Mauricio Durán Pabón - Viceministro de Transformación Digital
Ingrid Tatiana Montealegre Arboleda – Directora de Gobierno Digital
Luisa Fernanda Medina Martínez - Subdirectora de Fortalecimiento de Capacidades Públicas Digitales
Carlos Julio León Caicedo - Líder del Equipo Datos Abiertos y Software Libre
Manuela Serrano Romero – Líder de Uso de Datos Abiertos

Departamento Nacional de Planeación
Alejandra Botero Barco - Directora General del Departamento Nacional de Planeación
Daniel Gómez Gaviria - Subdirector General Sectorial
Viviana Rocío Vanegas Barrero - Directora de Desarrollo Digital
Diana Ramírez Roa – Asesora - Dirección de Desarrollo Digital
Francisco Perea De Zubiría – Asesor - Dirección de Desarrollo
Digital

Banco de Desarrollo de America Latina (CAF)
María Isabel Mejía Jaramillo - Ejecutiva Senior / Dirección de Transformación Digital

Oxford Insights
Pablo Fuentes Nettel - Director de proyecto
Alejandra Finotto - Investigadora principal
Kate Iida - Investigadora
Richard Stirling - Asesor experto
Peter Wells - Asesor experto

Versión	Observaciones
Versión 1 Agosto 2022	Diseño de Piloto de Data Commons

Comentarios, sugerencias o correcciones pueden ser enviadas al correo electrónico: gobiernodigital@mintic.gov.co.

Este documento de la Dirección de Gobierno Digital se encuentra bajo una Licencia Creative Commons Atribución 4.0 Internacional.

1. Introducción

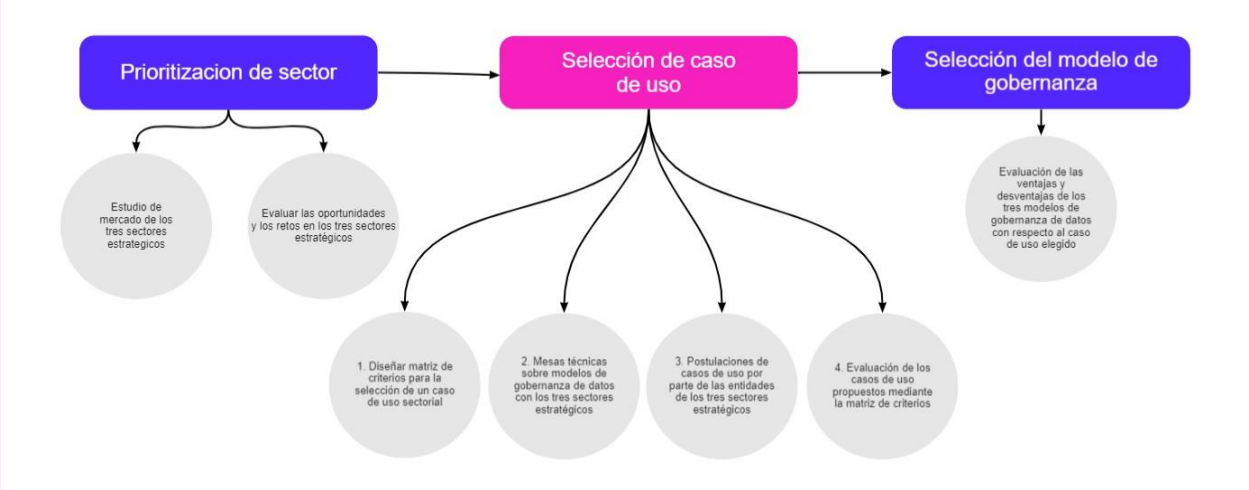
1.1. Descripción general

El presente documento parte de un proyecto consultoría llevado a cabo por Oxford Insights y comisionado por el Departamento Nacional de Planeación (DNP) y el Ministerio de Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC), que busca **contribuir al fortalecimiento del marco de infraestructura de datos en Colombia**. Tiene el propósito de contribuir al aprovechamiento de los datos en el país, así como cumplir con los objetivos del CONPES 4023 de 2021. Esto mediante la creación de un **documento de diseño para la posterior implementación de un mecanismo de intercambio de datos**.

El documento de diseño de piloto resulta de un proceso de investigación y análisis de oportunidades potenciales para el intercambio de datos en sectores considerados estratégicos para Colombia: transporte, agropecuario, y minas y energía. El **proyecto partió de lo general a lo particular, explorando diferentes sectores y modelos de gobernanza hasta llegar al caso de uso que se aborda en el presente documento**. Para ello se siguió una metodología compuesta por tres etapas: primero una priorización de sector[1], luego una selección de caso de uso[2], y finalmente la selección del modelo de gobernanza que mejor se ajustaba al caso de uso propuesto (ver Figura 1). Se seleccionó el caso de uso "**SIGRA**" (Sistema de Información para la Gestión de Riesgos Agropecuarios), **propuesto por el Ministerio de Agricultura y la UPRA[3]**, y el modelo del **Data Commons** como el mecanismo de gobernanza más apropiado para este.

[1] Referencia al Entregable 1
[2] Referencia al Entregable 2
[3] Referencia al Entregable 3

Figura 1. Pasos de la metodología de priorización del caso de uso



Fuente: Elaboración propia

Por medio de la implementación del Data Commons, se espera que los distintos actores del sector agropecuario, forestal, pesquero y de la acuicultura, puedan intercambiar datos cuyo análisis y uso **fortalezca las capacidades en los diferentes eslabones de las cadenas productivas, así como la gestión de riesgos asociados a la productividad.**

Por otro lado, de acuerdo a la naturaleza de la información que se intercambiará en el piloto, se considera que pueda haber un **impacto positivo a nivel social y económico**. Esto debido a que, mediante el intercambio de datos, se podrán generar capacidades y mitigar riesgos, resultando en un beneficio económico y productivo para las comunidades enfocadas en la producción agropecuaria.

Adicionalmente, se espera que la mitigación de riesgos contribuya a que los intermediarios financieros y aseguradoras puedan estructurar mejor sus productos y servicios, acorde a las necesidades de los productores. Esto permitiría optimizar la mitigación de los riesgos, resultando en **un campo colombiano más competitivo**, con mejores capacidades que le permitan abrirse a nuevos mercados.

El presente documento comienza ofreciendo información de contexto y estableciendo los objetivos que se quieren lograr a través de la implementación del Data Commons para el caso de uso de SIGRA. Posteriormente, se abordan temas relacionados con el ecosistema de datos actual, indicando los temas donde se deberá buscar mayor detalle en fases posteriores del proyecto.

Más adelante, se ofrece información sobre la solución de Data Commons, incluyendo:

dinámicas de gobernanza sugeridas para el piloto; requerimientos técnicos, con especial énfasis en tecnología y seguridad; y recursos humanos. En la siguiente sección, se presenta información de referencia para la estimación de costos. El documento concluye con la definición de una hoja de ruta, en la cual se describe el enfoque que se deberá aplicar al proyecto, así como las distintas fases que lo conforman.

1.2. Objetivos

1.2.1. Objetivo general

El objetivo de este documento es habilitar la implementación de un piloto de intercambio de datos en un sector estratégico en Colombia, que en este caso será un Data Commons para el sector agropecuario. Dado el caso de uso seleccionado, se busca que, mediante el intercambio de datos, se facilite la generación de indicadores de riesgos de mercado, financieros, climáticos y sanitarios en el sector agro colombiano (acorde a los objetivos y la naturaleza del SIGRA).

1.2.2. Objetivos específicos

1. Esclarecer el ecosistema de datos actual e identificar oportunidades para el intercambio de datos.
2. Generar bases sólidas sobre el modelo de gobernanza que mejor se adapta a atender la problemática: Data Commons.
    a. Identificar y explicar el valor añadido de un Data Commons para el caso de uso de SIGRA en comparación con otros sistemas de intercambio de información.
    b. Identificar y explicar los diferentes roles y responsabilidades de los actores involucrados en el Data Commons.
    c. Presentar un mapeo inicial sobre los requerimientos tecnológicos para la posterior implementación de un piloto de Data Commons.
    d. Brindar información de referencia sobre el equipo humano técnico requerido para el pilotaje del proyecto.

3. Ofrecer información de referencia sobre costos en torno a la implementación del Data Commons.

4. Establecer una hoja de ruta para la implementación del piloto que permita identificar las distintas fases del proceso.

1.3. Problema, necesidad, o caso de uso para el intercambio de datos

El caso de uso seleccionado para este piloto de intercambio de datos parte de la necesidad de implementar el **Sistema de Información para la Gestión de Riesgos Agropecuarios (SIGRA)**.

Dada la naturaleza y objetivos del SIGRA, **se requiere que las entidades participantes puedan acceder a información integrada que les permita generar indicadores de riesgos de mercado, financieros, agro-climáticos y fitosanitarios**.

Debido a lo anterior, resulta clave impulsar la creación de un mecanismo que facilite el intercambio de datos entre distintas entidades del sector agropecuario. En ese sentido, se busca diseñar y posteriormente implementar un modelo de Data Commons para el caso de uso específico del SIGRA. Las necesidades que ha identificado el caso de uso mencionado son las siguientes:

1. Producir y difundir estadísticas de utilidad para el sector.
2. Integrar información y contar con datos interoperables.
3. Fomentar el conocimiento, la generación, el análisis y el uso de la información sobre los distintos riesgos que afectan la actividad agropecuaria, forestal, pesquera y de la acuicultura.
4. Facilitar un intercambio de información y acceso a datos formal entre las entidades del sector, y actores no gubernamentales o entidades no adscritas al Ministerio de Agricultura y Desarrollo Rural, tales como las organizaciones gremiales.
5. Habilitar la generación de indicadores de riesgos **de mercado, financieros, climáticos y fitosanitarios**.

Finalmente, al habilitar la gestión integral de los riesgos agropecuarios y el intercambio de datos, se busca contribuir a aumentar la competitividad del sector agro colombiano. Asimismo, se quiere brindar información valiosa para implementación y seguimiento de

políticas, estrategias e instrumentos del Ministerio de Agricultura y Desarrollo Rural, de sus entidades adscritas y vinculadas y demás entidades interesadas en esta materia.

2. Ecosistema: actores y flujos de datos

2.1. Actores involucrados

En el ecosistema de datos en el sector agro colombiano interactúan una serie de instituciones. A continuación, una lista de las entidades que se han identificado como clave en el intercambio de datos:

- **MADR** – Ministerio de Agricultura y Desarrollo Rural. Dirección de Financiamiento y Riesgos Agropecuarios y Oficina de las Tecnologías de la Información y Comunicación.
- **UPRA** - Unidad de Planificación Rural Agropecuaria. Dirección de Uso Eficiente del Suelo y Oficina de las Tecnologías de la Información y Comunicación.
- **FINAGRO** – Fondo para el Financiamiento del Sector Agropecuario. Dirección de Información Sectorial y Riesgos Agropecuarios.
- **IDEAM** – Instituto de Hidrología, Meteorología y Estudios Ambientales. Subdirección de Meteorología.
- **ICA** – Instituto Colombiano Agropecuario. Dependencia Sanidad Animal/Vegetal y Oficina de las Tecnologías de la Información y Comunicación.
- **BANCO AGRARIO**.
- **BOLSA MERCANTIL** – Gerencia Corporativa de analítica y estudios económicos.
- **DANE** – Departamento Administrativo Nacional de Estadística.
- **FASECOLDA** – Federación de Aseguradores Colombianos.
- **MinTIC** – Ministerio de Tecnologías de la Información y las Comunicaciones.
- **DNP** – Departamento Nacional de Planeación.
- **DAPRE** – Departamento Administrativo De La Presidencia De La República.
- **C4RI** – Centro para la Cuarta Revolución Industrial.
- **SIC** – Superintendencia de Industria y Comercio.
- **IGAC** – Instituto Geográfico Agustín Codazzi.
- **Organizaciones gremiales del sector agropecuario**.

Los actores anteriormente mencionados pueden involucrarse en diversas etapas del ciclo de vida de los datos. Pueden encontrarse en una o más de las siguientes categorías según la Metodología del ODI (ODI, 2019)[4]:

[4] Para más información referirse a MinTIC. (2022). Pasos para pilotear proyectos de intercambio de datos. Asimismo, se puede consultar en la fuente primaria: http://theodi.org/wp-content/uploads/2019/06/ODI-Data-Ecosystem-Mapping-%E2%80%93-print-at-home-guide-A4-%E2%80%93-2019-06-26.pdf

Figura 2. Actores involucrados según su rol en el ecosistema de datos

Rol	Entidades
Guardianes de datos	MADR, UPRA, FINAGRO.
Beneficiarios	MADR, UPRA, FINAGRO, IDEAM, ICA, BANCO AGRARIO, BOLSA MERCANTIL, DANE, FASECOLDA.
Contribuidores	MADR, UPRA, FINAGRO, IDEAM, ICA, BANCO AGRARIO, BOLSA MERCANTIL, DANE, FASECOLDA.
Intermediarios	MADR, UPRA, FINAGRO.
Creadores o usuarios de los datos	MADR y UPRA, FINAGRO, IDEAM, ICA, BANCO AGRARIO, BOLSA MERCANTIL, DANE, FASECOLDA.
Reguladores	MADR, UPRA y SIC.
Policymakers (hacedores de políticas)	MinTIC, MADR, UPRA, DAPRE, DNP, C4RI.

Fuente: elaboración propia basada en la metodología ODI, 2019.

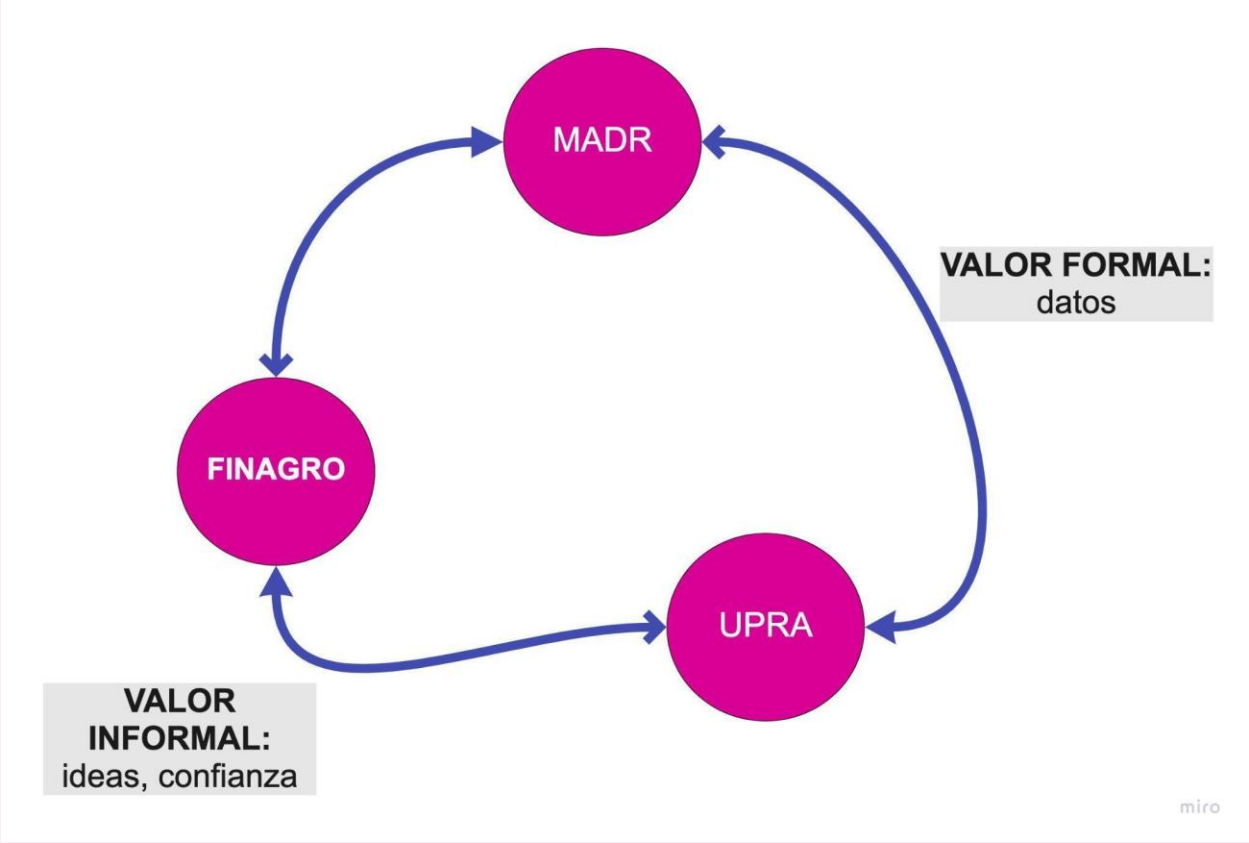
2.2. Flujo de datos

2.2.1. Estado actual del flujo de datos

Previo al diseño e implementación de un mecanismo de intercambio de datos, es crucial contar con un diagnóstico sobre el estado actual de flujo de datos en el caso de uso explorado. En ese sentido, **será fundamental que durante la siguiente fase del piloto se lleve a cabo un mapeo detallado y basado en investigación de usuario de estos flujos**.

Con la finalidad de ofrecer información de contexto, a continuación se presenta un mapeo inicial del ecosistema de datos. Según la metodología del ODI para mapear los ecosistemas de datos, existen intercambios de información formales e informales en los ecosistemas de datos (ODI, 2019).

Figura 3. Mapeo inicial del ecosistema de datos



Fuente: elaboración propia basada en la metodología ODI, 2019.

Para el **intercambio formal**, los contribuyentes actualmente intercambian información y datos principalmente por medio de correo electrónico, y suele ser en un formato de Excel. **La plataforma X-Road,** implementada por MinAgricultura y UPRA, **también es habilitadora para el intercambio de información actualmente**.

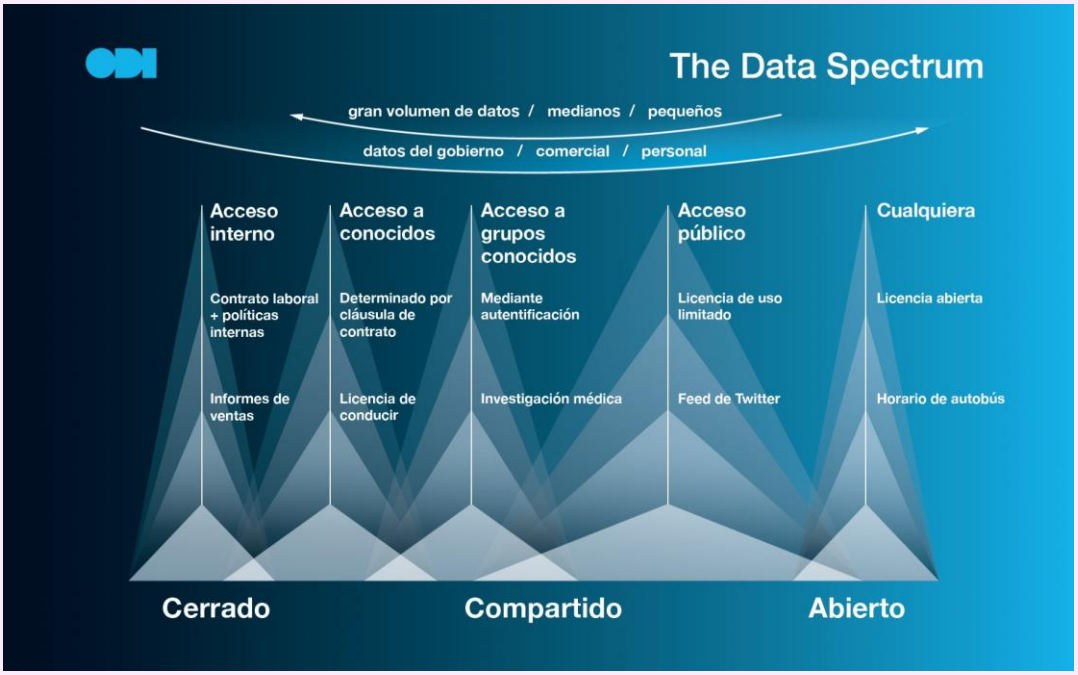
El **intercambio informal** de datos se refiere a la forma en que se intercambian las ideas y los conocimientos relacionados dentro del ecosistema de datos. Uno de los elementos que habilitan el intercambio informal de los datos es lo que el ODI ha categorizado como aseguranza de datos (*Data Assurance*). Este concepto se refiere a los procesos que fomentan la confianza en el intercambio de datos. La aseguranza de datos da certeza a las instituciones de que los datos se utilizarán para abordar necesidades específicas. Asimismo, asegura que las entidades que coleccionan, acceden y usan los datos lo hacen de forma responsable.

Un factor que puede resultar en mayor aseguranza de datos es el establecimiento de convenios. Actualmente, existen convenios en el ecosistema de datos del sector agro, que pueden ser sumamente benéficos. Tal es el caso del **convenio entre el MinAgricultura, UPRA y FINAGRO en el marco de SIGRA**. Por otra parte, existen **convenios entre MinAgricultura y el Banco Agrario de Colombia y MinAgricultura y Bolsa Mercantil de Colombia**.

Por otro lado, clasificar los datos disponibles de acuerdo a su nivel de accesibilidad es una herramienta crucial para mapear el ecosistema de datos. En ese sentido, el Data Spectrum propuesto por el ODI es una referencia útil. Dicho marco clasifican los datos en tres niveles distintos según su grado de accesibilidad:

- Cerrado - suelen ser datos de acceso interno o acceso a conocidos.
- Compartido - suelen ser datos de acceso a conocidos o acceso público.
- Abierto - suelen ser datos de acceso totalmente público y abierto.

Figura 4. ODI - The Data Spectrum



Fuente: ODI

A continuación presentamos una tabla que enlista los datos que las entidades buscan compartir y su nivel de accesibilidad actual así como la que se desea. Entre estos datos hay series históricas, de monitoreo, predicciones e información actualizada sobre precios y rendimientos de cultivo.

Figura 5. Conjuntos de datos identificados

Conjunto de datos	Titularidad	Nivel de accesibilidad actual	Nivel de accesibilidad que se desea	Tipo de datos
Resultados del Indicador sensibilidad - utilidad de arroz secano y riego, a nivel nacional y con frecuencia semestral	MADR	Cerrado	Compartido	Nacional
Resultados del Indicador sensibilidad - utilidad de maíz amarillo y blanco, tradicional y tecnificado a nivel nacional y con frecuencia semestral	MADR	Cerrado	Compartido	Nacional
Resultados del Indicador sensibilidad - utilidad de papa capira y suprema, a nivel nacional y departamental (Cundinamarca, Boyacá y Nariño) y con frecuencia semestral	MADR	Cerrado	Compartido	Cundinamarca, Boyacá y Nariño

Resultados del Indicador sensibilidad - utilidad de leche en polvo y cruda, a nivel nacional y con frecuencia semestral	MADR	Cerrado	Compartido	Nacional
Resultados del Indicador de evolución del porcentaje de siniestralidad general del FAG a nivel nacional.	Finagro	Cerrado	Compartido	Nacional
Resultados del Indicador de evolución del porcentaje de siniestralidad, a nivel nacional, por categorías (Arroz) del FAG.	Finagro	Cerrado	Compartido	A nivel nacional por categoría de productos (Arroz, maíz, papa, leche y ganado)
Resultados del Indicador de evolución del porcentaje de siniestralidad, a nivel nacional, por categorías (maíz) del FAG.	Finagro	Cerrado	Compartido	A nivel nacional por categoría de productos (Arroz, maíz, papa, leche y ganado)
Resultados del Indicador de evolución del porcentaje de siniestralidad, a nivel nacional, por categorías (papa) del FAG.	Finagro	Cerrado	Compartido	A nivel nacional por categoría de productos (Arroz, maíz, papa, leche y ganado)
Resultados del Indicador de evolución del porcentaje de siniestralidad, a nivel nacional, por categorías (leche) del FAG.	Finagro	Cerrado	Compartido	A nivel nacional por categoría de productos (Arroz, maíz, papa, leche y ganado)
Resultados del Indicador de evolución del porcentaje de siniestralidad, a nivel nacional, por categorías (ganado) del FAG.	Finagro	Cerrado	Compartido	A nivel nacional por categoría de productos (Arroz, maíz, papa, leche y ganado)
Resultados del Indicador de Máximos puntos adicionales a la IBR nominal, para el cultivo de arroz, para los sistemas secano y riego.	Finagro	Cerrado	Compartido	Nacional y Zona productoras más representativas (Llanos orientales, Bajo cauca, Costa Norte, Santanderes y Centro)
Resultados del Indicador de Máximos puntos adicionales a la IBR nomina para el cultivo de maíz.	Finagro	Cerrado	Compartido	Nacional y departamentos más representativos (Córdoba, Cundinamarca, Tolima, Meta y Valle)
Resultados del Indicador de Máximos puntos adicionales a la IBR nomina para el cultivo de papa.	Finagro	Cerrado	Compartido	Nacional y departamentos más representativos (Cundinamarca, Boyacá y Nariño).
Resultados del Indicador de Utilidad Neta Simulada para el cultivo de arroz, para los sistemas secano y riego.	Finagro	Cerrado	Compartido	Nacional y Zona productoras más representativas (Llanos orientales, Bajo cauca, Costa Norte, Santanderes y Centro)

Resultados del Indicador de Utilidad Neta Simulada para el cultivo de maíz.	Finagro	Cerrado	Compartido	Nacional y departamentos más representativos (Córdoba, Cundinamarca, Tolima, Meta y Valle)
Resultados del Indicador de Utilidad Neta Simulada para el cultivo de papa.	Finagro	Cerrado	Compartido	Nacional y departamentos más representativos (Cundinamarca, Boyacá y Nariño).
Resultados del Indicador de Margen de rentabilidad neta y bruta para el cultivo de arroz, para los sistemas secano y riego.	Finagro	Cerrado	Compartido	Nacional y Zona productoras más representativas (Llanos orientales, Bajo cauca, Costa Norte, Santanderes y Centro)
Resultados del Indicador de Margen de rentabilidad neta y bruta para el cultivo de maíz.	Finagro	Cerrado	Compartido	Nacional y departamentos más representativos (Córdoba, Cundinamarca, Tolima, Meta y Valle)
Resultados del Indicador de Margen de rentabilidad neta y bruta para el cultivo de papa	Finagro	Cerrado	Compartido	Nacional y departamentos más representativos (Cundinamarca, Boyacá y Nariño).
Resultados del Indicador de Valor en riesgo (95% de confianza) y Número de veces que puede cubrir un crédito para el cultivo de arroz, para los sistemas secano y riego.	Finagro	Cerrado	Compartido	Nacional y Zona productoras más representativas (Llanos orientales, Bajo cauca, Costa Norte, Santanderes y Centro)
Resultados del Indicador de Valor en riesgo (95% de confianza) y Número de veces que puede cubrir un crédito para el cultivo de maíz.	Finagro	Cerrado	Compartido	Nacional y departamentos más representativos (Córdoba, Cundinamarca, Tolima, Meta y Valle)
Resultados del Indicador de Valor en riesgo (95% de confianza) y Número de veces que puede cubrir un crédito para el cultivo de papa.	Finagro	Cerrado	Compartido	Nacional y departamentos más representativos (Cundinamara, Boyacá y Nariño).
Resultados del Indicador de Probabilidad de incurrir en pérdidas para el cultivo de arroz, para los sistemas secano y riego.	Finagro	Cerrado	Compartido	Nacional y Zona productoras más representativas (Llanos orientales, Bajo cauca, Costa Norte, Santanderes y Centro)
Resultados del Indicador de Probabilidad de incurrir en pérdidas para el cultivo de maíz.	Finagro	Cerrado	Compartido	Nacional y departamentos más representativos (Córdoba, Cundinamarca, Tolima, Meta y Valle)
Resultados del Indicador de Probabilidad de incurrir en pérdidas para el cultivo de papa.	Finagro	Cerrado	Compartido	Nacional y departamentos más representativos (Cundinamarca, Boyacá y Nariño).
Inventarios arroz.	MADR	Cerrado	Compartido	Nacional

Precios pronosticados de arroz a nivel nacional y por los principales municipios productivos, mensuales y en rangos de valor mínimo, valor medio y valor máximo.	MADR	Cerrado	Compartido	Espinal, Ibagué, Montería, Neiva, Valle, Villavicencio y Yopal.
Precios pronosticados de papa Capira y Suprema a nivel nacional y por los principales departamentos productivos, mensuales y en rangos de valor mÍnimo, valor medio y valor máximo.	MADR	Cerrado	Compartido	Córdoba, Meta, Tolima y Valle. (Precio en zona de producción reportado por productores a Ingenieros regionales)
Precios pronosticados de maíz amarillo y blanco a nivel nacional y por los principales departamentos productivos, mensuales y en rangos de valor mÍnimo, valor medio y valor máximo.	MADR	Cerrado	Compartido	Antioquia, Boyacá, Cauca, Cundinamarca, Nariño, Santander y Valle.
Precios pronosticados de leche en polvo y cruda a nivel nacional, mensuales y en rangos de valor minimo, valor medio y valor máximo.	MADR	Cerrado	Compartido	Nacional
Rendimientos Arroz	UPRA	Cerrado	Compartido	
Rendimientos Papa	UPRA	Cerrado	Compartido	
Rendimientos Leche	UPRA	Cerrado	Compartido	
Rendimientos Maíz	UPRA	Cerrado	Compartido	

Fuente: Elaboración propia

Actualmente, la totalidad de los conjuntos de datos que se desean compartir se encuentran en el nivel “Cerrado”. Para que los conjuntos de datos se puedan aprovechar en el Data Commons, su nivel de acceso tendrá que pasar de “Cerrado” a “Compartido” (Acceso a grupos conocidos). Esto permitirá que las entidades involucradas en el mecanismo de intercambio de datos puedan acceder a estos conjuntos de datos directamente, mediante autenticación.

2.3. Aprovechamiento de datos

En el ecosistema actual de datos en el sector agro, se da por entendido que la fuente original garantiza la calidad de la información suministrada para la construcción de los

indicadores. Por otra parte, se cuentan con instrumentos para identificar y reportar los resultados de la verificación de calidad. El MADR (por medio de la dirección de financiamiento y riesgos) y la UPRA han expresado la voluntad de realizar dicha tarea conjuntamente en caso de que la implementación del piloto lo requiera. En línea con lo anterior, también deben tenerse en cuenta los lineamientos de protección e integridad de la información de Colombia, y por lo tanto se debe contemplar la generación de espacios donde se pueda procesar la información proveniente de las entidades.

En cuanto a estándares de calidad de los datos, ya se han establecido algunos por parte del MinTIC y parece que los conjuntos de datos que se quieren compartir ya los cumplen (por ejemplo, los estándares de Lenguaje Común de Intercambio de Datos y los estándares de Interoperabilidad).

Algunos obstáculos que se han identificado en el intercambio actual de datos son:

- El proceso de solicitud y recepción de datos puede ser poco eficiente. Se realiza principalmente por medio de correo electrónico y los tiempos de respuesta pueden ser un poco lentos.

- Dado que los datos no se comparten de forma inmediata, no se cuenta con información temporalmente de calidad ya que la frecuencia de actualización no es la más vigente. Esto afecta a la calidad de los datos.

- Los datos se comparten principalmente en formato de hojas de cálculo de Excel.
- Los datos no se encuentran en forma agregada.
- La fuente de información suministra los datos completos a nivel nacional. Sin embargo para desagregaciones geográficas y sectoriales se requiere realizar procesos de imputación que es necesario validar con la fuente de la información
- Existe cierta reticencia a compartir determinados datos por temas de seguridad.

Es importante recalcar que los conjuntos de datos que se quieren compartir actualmente no cuentan con información de tipo personal, por lo que no se requieren implementar requisitos para cumplir con estándares de privacidad en este momento. No obstante, **si en el futuro se quisieran incorporar datos personales, entonces sí habría que asegurarse que se siguen metodologías con privacidad**

por defecto a la hora de compartir los datos. De igual manera, es necesario que se cuente con mecanismos de seguridad que aseguren el uso íntegro de los datos[5].

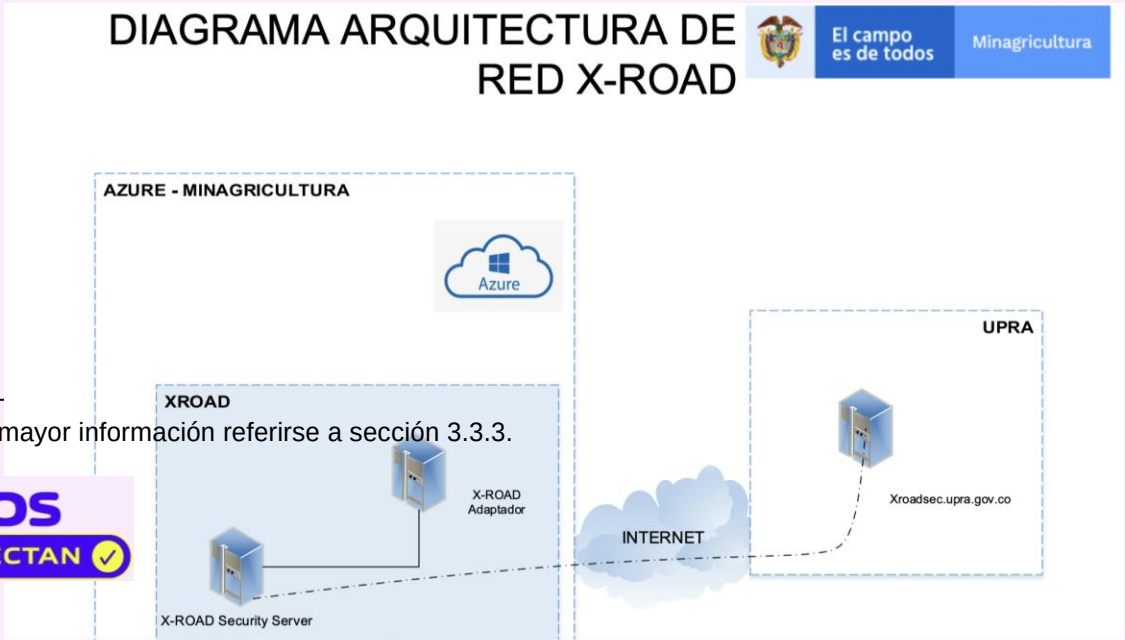
2.4. Interoperabilidad actualmente

Recientemente, el sector agro ha dado pasos importantes en materia de interoperabilidad de datos. En la actualidad, entidades como el Ministerio de Agricultura y Desarrollo Rural han adoptado el **uso del sistema X-Road para garantizar que la interoperabilidad de los datos**. Lo anterior, presenta beneficios importantes para el intercambio de datos:

- Facilita la cooperación entre instituciones gubernamentales para el intercambio de datos.
- Elimina barreras burocráticas para la compartición interna.
- Asegura la interoperabilidad de los datos mediante el establecimiento de estándares.
- Permite intercambiar datos de forma segura.

Si bien resulta en beneficios importantes, el uso de X-Road presenta limitaciones para facilitar la interoperabilidad e intercambio de datos entre los diferentes tipos de actores que participan en el sector agro. Esto se debe principalmente a que el uso de X-Road está limitado a ciertas entidades dentro del sector público. Por lo tanto, **no incluye a actores importantes como organizaciones gremiales e instituciones privadas de investigación**, entre otras.

Figura 6. Diagrama arquitectura de red X-Road



[5] Para mayor información referirse a sección 3.3.3.

Fuente: Ministerio de Agricultura

3. Solución de Data Commons

3.1. Sobre el modelo Data Commons

El concepto de *Data Commons* es parte de los modelos de gobernanza de intercambio de datos que han emergido en los últimos años. Se basa en la idea de que los **datos deben ser útiles para una comunidad o para el público en general**. Parte del argumento de que se debe garantizar que todos los agentes cuenten con las mismas condiciones en cuanto al acceso a los datos. Actualmente, existen dos principales categorías de Data Commons: **investigación** y **recurso común** (***common pool***). De acuerdo al *Mozilla Data Futures Lab,* "En un Data Commons los datos se comparten como un recurso común. A menudo, van acompañados de un alto grado de propiedad y liderazgo de la comunidad y tienen una causa de bien público" (Mozilla Foundation, 2022).

Dadas las características de la propuesta, se consideró que el modelo de gobernanza **más indicado para llevar a cabo el proyecto es el de Data Commons**. Este modelo de gobernanza permitiría que entidades líderes y miembros involucrados tengan el mismo acceso a los datos que se comparten. El Data Commons tiene además la finalidad de construir una comunidad de actores de distintas naturalezas que pretenden atender un problema común. Dentro de esta comunidad de actores se podrían incluir actores del sector privado, universidades, centros de investigación y ONGs, entre otros.

El Data Commons contribuirá a generar incentivos y normas para la cooperación entre distintas entidades. Esto **facilitará a las entidades involucradas mitigar y gestionar riesgos en temas agropecuarios**, fortaleciendo así el SIGRA. Asimismo, abriría la posibilidad a que otros actores relevantes como la sociedad civil o investigadores puedan acceder a estos datos en fases posteriores del proyecto (para hacer estudios de investigación, por ejemplo, sobre riesgos agropecuarios).

3.2. Data Commons para el SIGRA

Para el caso de uso del SIGRA, un Data Commons trae consigo valor agregado en comparación con otros sistemas de intercambio de información. En específico:

- Brinda acceso a los datos a actores externos.
- Permite a las entidades decidir qué actores participan - por lo tanto les brinda un

mayor márgen de acción en términos de confidencialidad.

- Incluye a más actores y por lo tanto más datos dentro del intercambio.
- Permite que se presenten datos agregados en una única plataforma.
- Da la oportunidad de presentar y analizar los datos en distintos formatos.
- Tiene el objetivo de que se compartan los datos con un propósito común, y por lo tanto permite que los miembros influencien aquellas acciones que se quieran llevar a cabo con los datos.

Esto no se podría conseguir con otros mecanismos de intercambio de datos. Un Data Marketplace no sería adecuado ya que el propósito del caso de uso no implica la monetización del intercambio de datos y no se pueden comercializar datos del gobierno. Un Data Trust tampoco sería el mejor mecanismo ya que el SIGRA no requiere la existencia de un fiduciario que garantice el uso adecuado de datos personales, si no fomentar la cooperación y confianza entre distintas entidades.

Adicionalmente, a diferencia de un portal de datos abiertos, los Data Commons cuentan con un sistema de membresía. Esto permite que los miembros tengan la capacidad de decidir quiénes tienen acceso a los datos; y que tengan mayor control sobre asuntos de privacidad. A su vez, esto brinda mayor confianza a los miembros.

Pero esto tampoco significa que algunas bases de datos o datos procesados del Data Commons no se puedan abrir al público en general. Se podría decidir dar acceso “Abierto” a algunos de estos, y mantener acceso a sólo los miembros para otros datos.

Otro beneficio del Data Commons es que si se incluyesen a otros actores como miembros del Data Commons en futuras fases, se podría crear valor adicional a través de nuevos casos de uso. Por ejemplo:

1. Si se diese acceso al Data Commons a aseguradoras del sector privado, estas podrían usar los datos que se comparten para mejorar o crear nuevos servicios. Estos servicios podrían ser utilizados por otros actores del sector agro, teniendo un impacto positivo en la gestión de riesgos agropecuarios.
2. Si se diese acceso al Data Commons a los mismos agricultores, estos podrían beneficiarse de los datos que se comparten para mejorar sus técnicas agrícolas y gestionar sus riesgos. Además, ellos mismos podrían contribuir al Data Commons con datos, de esta forma permitiendo que se pudieran hacer análisis más rigurosos de riesgos agropecuarios.
3. Si se diese acceso al Data Commons a centros de investigación, estos podrían usar los datos que se comparten para realizar estudios y contribuir a la definición de políticas públicas. Esto podría contribuir a que hubiese mayor innovación.

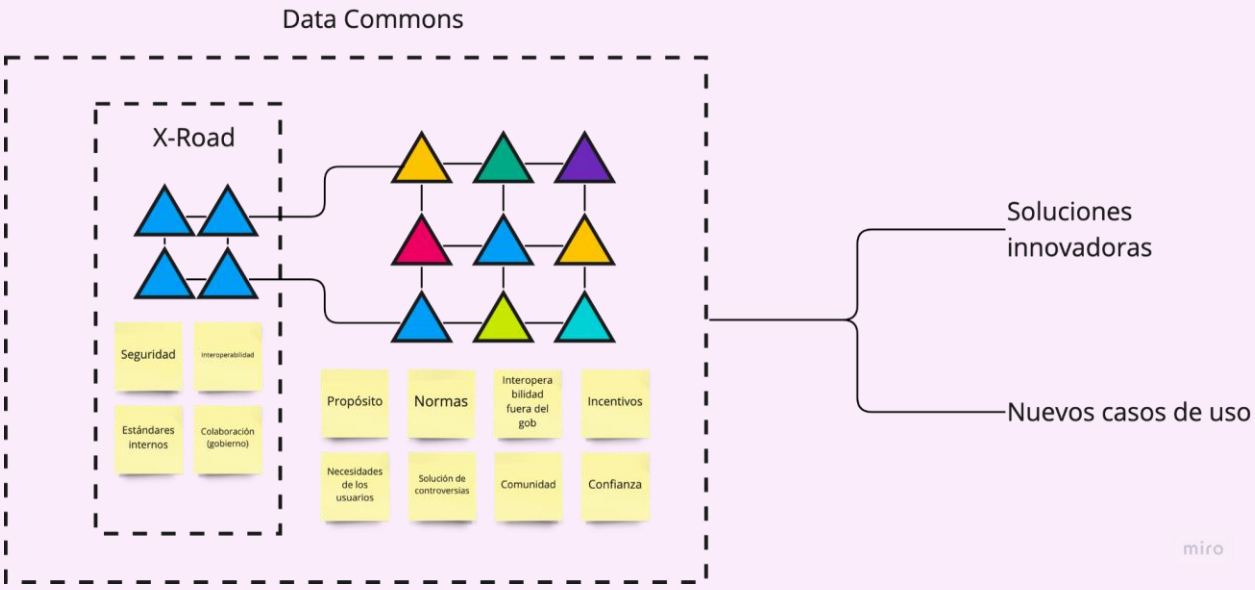
Finalmente, **es importante destacar que la implementación de un Data Commons no implicaría la sustitución del sistema X-ROAD como** sistema de interoperabilidad entre entidades públicas. Los Data Commons son modelos de gobernanza para el intercambio de datos. Es decir, son un conjunto de lineamientos sobre la interacción entre los agentes que intercambian datos. Por lo tanto, los **Data Commons no son sistemas tecnológicos**. Por supuesto requieren de una serie de habilitadores técnicos para funcionar, las cuales se ajustan a las características específicas del caso de uso y a las preferencias de los usuarios. En ese sentido, **los sistemas de interoperabilidad como X-Road son un ejemplo de los habilitadores técnicos requeridos**.

Como se presenta en la Figura 7, actualmente X-ROAD facilita el intercambio de datos entre instituciones públicas del sector agro, habilitando la interoperabilidad, seguridad, adopción de estándares y la colaboración entre entidades. Este ecosistema dinámico habilitado por X-Road puede ser integrado a una plataforma más amplia, como el Data Commons. La creación del Data Commons permitirá que el intercambio de datos trascienda en dos sentidos.

- En primer lugar, perimirá la **participación de entidades de carácter privado** —como organismos gremiales, instituciones de investigación y organizaciones no gubernmantales, entre otras—. Esto enriquecerá lantidad y variedad de datos que se intercambian y facilitará el intercambio de distintas perspectivas.
- Asimismo, el Data Commons contempla la creación de **un sistema de gobernanza que coordinará la interacción de los distintos actores involucrados**. Lo anterior resultará en varios beneficios como el establecimiento de un propósito común y la creación de normas que regulen la interacción. Dicho marco tiene el potencial de crear mayor mayor confianza entre los actores involucrados y generar incentivos para la colaboración.

Eventualmente, lo anterior resultará en **soluciones innovadoras a retos comunes** e incluso podría facilitar la **identificación de casos de uso adicionales** que sean de interés para el sector agro.

Figura 7. X-Road y Data Commons



Fuente: elaboración propia

3.3. Casos similares

Existen varios ejemplos de implementación de Data Commons en el sector agro. A continuación presentamos una selección para que puedan servir de referencia a la hora de construir el del SIGRA.

1. OS Climate

El “OS Climate Data Commons" se construyó con el objetivo de mitigar riesgos relacionados con el cambio climático. Los miembros tienen acceso a datos que les permiten enfocar sus inversiones hacia las tecnologías y proyectos más efectivos contra el cambio climático.

Este es un Data Commons al que solo pueden acceder sus miembros, que pertenecen en su mayoría al sector privado. Con esto se quiere influenciar el diseño de políticas que involucren eficazmente a los mercados de capitales en la lucha contra el cambio climático.

2. Ag Data Commons

El “Ag Data Commons” se creó a través de una colaboración entre distintas entidades del sector agrícola de Estados Unidos, con USDA (el departamento de agricultura), liderando la iniciativa.

El objetivo de este Data Commons es permitir que cualquier miembro del público acceda a los datos agrícolas que coleccionan estas entidades, para poder hacer estudios de investigación e innovación. Por lo tanto, casi todos los datos son de acceso abierto al público.

3. “A Data Commons for Food Security”

El Data Commons “A Data Commons for Food Security” es un proyecto que aún se está desarrollando. La idea detrás del proyecto es permitir que los agricultores obtengan beneficio de compartir sus datos con el gobierno. El Data Commons está construido de tal manera que los agricultores puedan tener control sobre la privacidad de sus datos, a la vez que puedan tener acceso a otros datos agregados anonimizados. Esto es para que puedan gestionar riesgos asociados a su producción, y que puedan incrementar su confianza con el gobierno.

3.4. Gobernanza del Data Commons

El modelo de gobernanza Data Commons se basa en la idea de que los datos son un bien común y, por lo tanto, se debe fomentar la igualdad de acceso a ellos. En ese sentido, los miembros de un Data Commons, comparten los datos entre ellos de forma igualitaria.

Algunos de los elementos que se deben evaluar al definir la gobernanza de un Data Commons como el que se pretende implementar son (entre otros): quiénes formarán parte; la forma de gobierno entre las partes interesadas; motivos e intereses de los actores; métodos para la toma de decisiones; incentivos. **Las respuestas a dichas cuestiones deberán ser resultado del proceso mismo del piloto y del intercambio de ideas entre los distintos actores involucrados**. A continuación, presentamos una guía para la definición de las dinámicas para gobernar el intercambio de datos en el contexto de un Data Commons en el sector agro colombiano.

Si bien, el acceso a los datos se da de forma igualitaria, el desarrollo del piloto requiere de entidades líderes. Será fundamental que las entidades líderes, MADR y UPRA, asuman el rol de liderazgo en el desarrollo del piloto. Esto quiere decir que dichas entidades tendrán la responsabilidad de coordinar acciones para definir la gobernanza

del piloto. Será fundamental que, como primer paso, MADR y UPRA definan cómo asumirán el rol de líderes y establezcan las responsabilidades de cada una de ellas.

Figura 8. Proceso de definición de dinámicas de gobernanza

Fuente: Elaboración propia

Como se ha mencionado anteriormente, los Data Commons parten de la idea de que los datos deben ser abordados como un bien común. Por lo tanto, **es importante que UPRA y MADR busquen involucrar a las demás partes interesadas en la definición de normas y demás procesos internos**. Las entidades líderes deben buscar el mayor grado de consenso para lograr un desarrollo efectivo del piloto.

En ese sentido, se recomienda que se aplique un enfoque de gestión de proyecto que sea colaborativo y que permita realizar ajustes durante su ejecución. Partiendo de lo anterior, el presente documento recomienda utilizar una metodología *Agile* para el desarrollo del proyecto. Agile es un concepto de gestión de proyectos desarrollado para proporcionar a los gestores de proyectos una forma más flexible y eficiente de comercializar productos más rápido. El significado de Agile (ágil) es la capacidad de moverse rápida y fácilmente. Por lo tanto, un enfoque Agile permite a los equipos del proyecto adaptarse más rápido y con más facilidad en comparación con otras metodologías de proyecto[6]. Los métodos ágiles animan a los equipos a construir rápidamente, a testear lo que han construido y a iterar su trabajo basándose en la retroalimentación regular y en otros datos útiles. Los métodos ágiles pueden parecer muy diferentes para las personas acostumbradas a los métodos tradicionales de desarrollo de software en cascada.

Una vez establecidas las dinámicas para la gestión del proyecto, será necesario definir el marco interno de gobernanza. Una de las metodologías más avanzadas para establecer los lineamientos de gobernanza de un Data Commons son los principios de Elinor Ostrom. El presente documento recomienda basarse en dichos principios para definir las dinámicas específicas que guiarán la interacción entre las entidades líderes y demás usuarios y proveedores de datos.

[6] https://www.wrike.com/es/project-management-guide/fundamentos-de-la-metodologia-agile/

En la siguiente tabla, encontrarán los principios de Ostrom aplicados al caso de uso seleccionado. Asimismo, se incluyen una serie de preguntas que servirán para evaluar los lineamientos de gobernanza una vez que el piloto comience a operar.

Figura 9. Principios para un Data Commons

Principio	Descripción	Preguntas de monitoreo
1. Barreras bien definidas	Es fundamental que se defina de forma muy clara el propósito y la finalidad del Data Commons. De esa forma será factible definir los actores que pueden contribuir, usar y acceder a los datos, así como los actores que intervienen en la toma de decisiones.	¿Existe un propósito declarado para la creación del Data Commons y para la recopilación y el uso de los datos? ¿Están definidos los contribuyentes y los beneficiarios? ¿Cómo utilizarán y se beneficiarán las partes interesadas de los datos?
2. Reglas apropiadas	Se deben establecer reglas sobre el uso de los recursos involucrados en el Data Commons: datos, financiamiento, tiempo de las personas, equipo técnico, etc. Dichas reglas deben definirse a partir del propósito del proyecto y buscar el mejor interés de las entidades involucradas.	¿Existen normas que regulen qué métodos y herramientas se utilizan para almacenar, asegurar y proteger la integridad de los datos? ¿Existen normas que regulan quién puede acceder a los datos, de qué manera, con qué finalidad y durante qué tiempo? ¿Existe una diferencia entre el modo en que los contribuyentes, otros actores interesados (si las hay) y el público en general pueden acceder a los datos, es decir, derechos de exclusión?
3. Proceso de definición de reglas	Definir las reglas de forma colectiva será benéfico para el proyecto. Es importante que quienes estén sujetos a las normas también participen en la creación de éstas.	¿Existen medios que permitan articular las perspectivas de los actores interesados (incluidos los usuarios de datos, los productores de datos y, potencialmente, los sujetos de los datos) y representar sus intereses?¿Existe un proceso claro que describa cómo se toman y actualizan las decisiones sobre la recogida, el almacenamiento, el acceso y el uso de los datos?
4. Monitoreo	Se deben establecer mecanismos que permitan monitorear el uso correcto de los datos y calidad de éstos, así como cuestiones de certificación y seguridad. Se	¿Existen mecanismos para supervisar la seguridad de la arquitectura? ¿Existen mecanismos para supervisar las violaciones de la seguridad, los pirateos

	recomienda un enfoque colectivo, donde todas las entidades puedan involucrarse de forma similar.	y los usos indebidos?¿Está claro cuándo las auditorías serán realizadas por auditores externos o internos?
5. Sanciones	Es importante establecer sanciones a las entidades que violen las normas acordadas. Se recomienda aplicar un enfoque de sanciones de escala. Es decir, no todas las violaciones a las normas deben ser tratadas igual.	¿Las normas relacionadas con la seguridad e integridad de los datos tienen su correspondiente conjunto de medidas de responsabilidad?
6. Mecanismos para la resolución de conflictos	Es necesario que el Data Commons cuente con un mecanismo para resolver controversias entre las distintas entidades. En caso de conflictos de gravedad mayor, se recomiendan ser tratados por instancias externas. Es fundamental definir qué controversias se tratarán de forma interna y cuáles se llevarán a otras instancias.	¿Existen mecanismos internos para resolver los conflictos? ¿Existe un procedimiento de apelación a las decisiones tomadas (especialmente en torno a las sanciones)?
7. Derechos a la autogobernanza	Los lineamientos del Data Commons deben reconocer la facultad de cada entidad de gobernarse a sí misma. La definición de normas no deberá contradecir las reglas internas de cada entidad en materia de protección de datos y demás regulaciones.	¿Existe un organismo responsable de promover el modelo de administración de datos que rinda cuentas a los reguladores y autoridades externas? ¿Hay algún mecanismo o infraestructura que pueda poner en marcha para mejorar su libertad de decisión y organización en torno a los datos?
8. Entidades implantadas dentro de un sistema	Todos los principios anteriores serán aplicados a un conjunto de entidades independientes con un objetivo común. En algunos casos, será útil dividir a las entidades en grupos dentro del Data Commons (basado en temas como fases de la cadena de valor, productos, etc.). Esto puede llevar a una toma de decisiones más eficiente en casos donde aborden temas muy específicos.	¿Está claro cómo encaja la organización o el proyecto en sistemas más amplios?

Fuente: Elaboración propia, basada en Mozilla, “A Practical Framework for Applying Ostrom’s Principles to Data Commons Governance,” 2021

3.4.1. Roles y responsabilidades en el Data Commons

Como se mencionó anteriormente, durante la fase piloto, las entidades líderes tendrán que definir qué otras instituciones formarán parte del Data Commons. Estas entidades podrán participar como proveedores y/o usuarios de los datos. Asimismo, tendrán que seguir las reglas, lineamientos y formatos establecidos. Es importante recalcar que, si bien las entidades líderes coordinan acciones, en términos de nivel de acceso a los datos, deberán estar sujetas a las mismas condiciones que los demás miembros.

A continuación se detallan las entidades líderes así como instituciones que se han identificado como potenciales miembros del Data Commons y aquellas que brindarán apoyo adicional.

Figura 10. Roles y responsabilidades sugeridas

Rol	Responsabilidad	Entidades
Entidades líderes	Se encargarán de coordinar el proyecto piloto. Esto involucra la definición de las dinámicas de gobernanza, así como la planeación de la arquitectura tecnológica que se va a utilizar. Asimismo, participarán como recolectores, proveedores y usuarios de datos una vez que el Data Commons comience a operar.	FINAGRO IDEAM ICA BANCO AGRARIO
Miembros del Data Commons	Participarán como recolectores, proveedores y usuarios de datos una vez que el Data Commons comience a operar. También podrán ser involucrados, a consideración de las entidades líderes, en la fase de pilotaje para definir cuestiones de gobernanza y arquitectura tecnológica.	FINAGRO IDEAM ICA BANCO AGRARIO BOLSA MERCANTIL DANE FASECOLDA
Apoyo adicional	Estas entidades podrán apoyar con conocimiento especializado que contribuya a un mejor funcionamiento tanto del proyecto piloto como del Data Commons una vez que esté operando formalmente. Podrán aportar conocimiento en áreas como marco de política pública, infraestructura tecnológica, mejores prácticas internacionales y regulación.	MinTIC DNP DAPRE C4RI MinCIENCIAS

Fuente: Elaboración propia

Asimismo, existe la posibilidad de que algunos de los datos se compartan con actores que no son proveedores, ni participan como miembros del Data Commons. Esto podría hacerse mediante la **publicación de los datos en plataformas de acceso público**. Esto deberá definirse como parte de los lineamientos de gobernanza.

3.5. Requerimientos técnicos

3.5.1. Tecnología

Los Data Commons funcionan mejor en situaciones en las que hay mucha confianza entre los actores interesados, ya que todos los participantes reciben beneficios basados en el recurso común compartido. Un aspecto importante del flujo de datos en un Data Commons es que todos los datos se comparten por igual, al igual que los derechos sobre los mismos. Esto ayuda a facilitar la igualdad de acceso a los datos por parte de todos los participantes en un Data Commons.

Debido a lo anterior, **la tecnología debe asegurar que no se impongan demasiadas restricciones a los miembros del Data Commons**. La tecnología no debe ser restrictiva. Por lo tanto, se puede prestar menos atención a cuestiones de verificación y auditoría de los datos. En ese sentido, el Data Commons debería tener un entorno más parecido al de una plataforma de datos abiertos o al de los portales de datos en general. Por otro lado, es importante contar con APIs que permitan el intercambio de información.

Para definir la arquitectura tecnológica que requiere un Data Commons es fundamental pensar en las principales necesidades de los usuarios que conocemos actualmente y convertirlas en "historias de usuario". Una historia de usuario es una explicación informal y general de una función de software escrita desde la perspectiva del usuario final. Su propósito es articular cómo una característica del software proporcionará valor al cliente y, por lo tanto, asegurarse de que se está construyendo un software basado en las necesidades reales del usuario. Se escribe en el siguiente formato:

Como… > necesito… > para que....

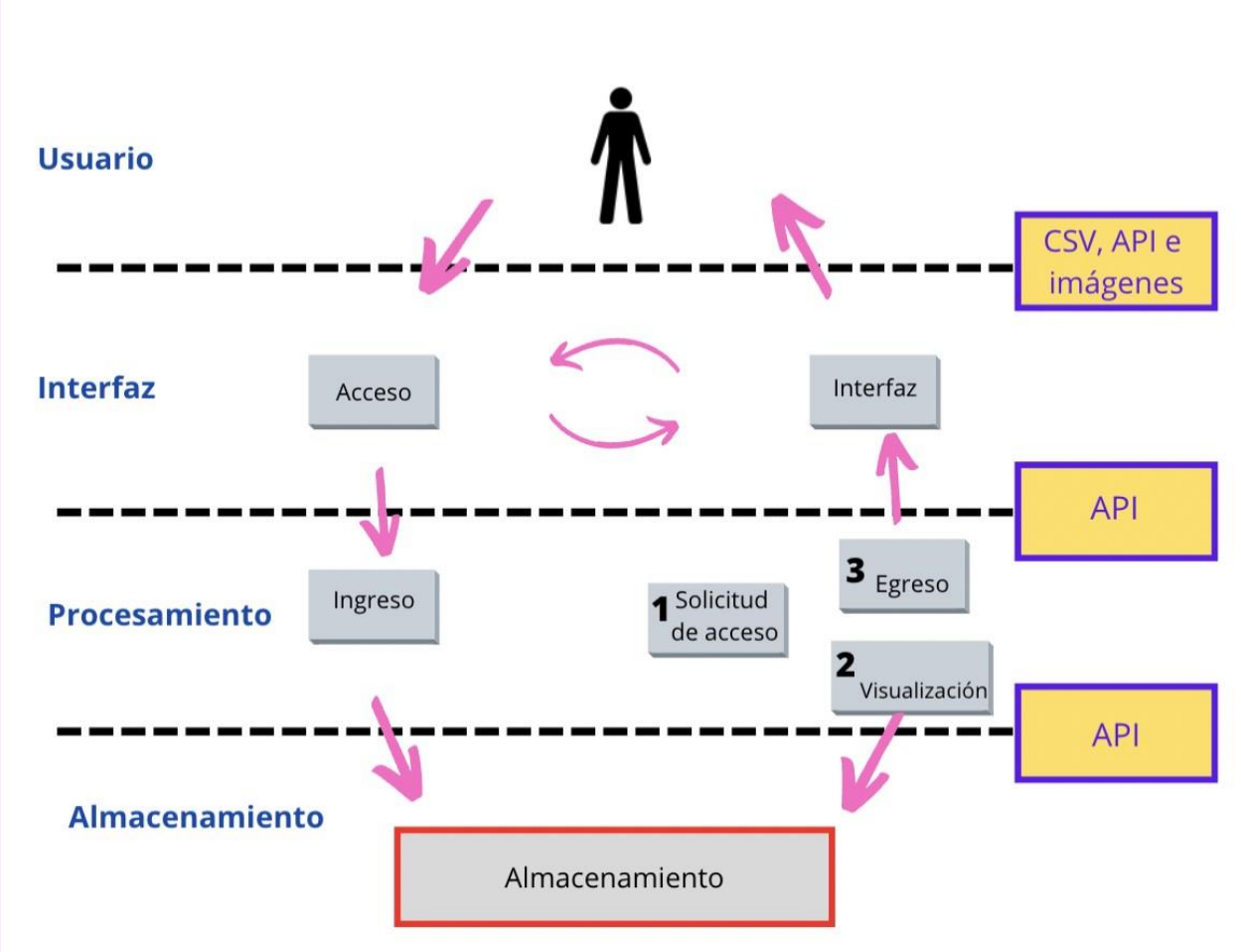
Basado en la investigación realizada, se han identificado las siguientes historias de usuario:

- Como Ministerio de Agricultura, necesito proporcionar datos en el Data Commons, para que las otras entidades tengan acceso a ellos.
- Como UPRA, necesito proporcionar datos en el Data Commons, para que las otras entidades tengan acceso a ellos.
- Como una entidad gubernamental del sector agrícola, necesito que se me conceda acceso al Data Commons, para poder consumir los datos que han sido proporcionados por el Ministerio de Agricultura y la UPRA.
- Como proveedor de datos, necesito poder proporcionar datos al Data Commons desde la ubicación actual en la que están almacenados, para que otros puedan acceder a ellos.
- Como consumidor de datos, necesito poder acceder a los datos a través del Data Commons, para poder utilizarlos.
- Como usuario de Data Commons, necesito poder descargar los datos en formato reutilizable, para poder crear visualizaciones con ellos.
- Como entidad gobernante del Data Commons, necesito establecer requisitos de autenticación y registro, de modo que sólo los usuarios con permiso puedan acceder a él.
- Como usuario del Data Commons, necesito saber que los datos no se filtrarán a actores que no son miembros de este, para confiar en la integridad del Data Commons.

3.5.2. Arquitectura técnica

Basándose en los relatos de los usuarios mencionados anteriormente, se puede deducir la siguiente infraestructura técnica general para la implementación del Data Commons. **Es importante mencionar que la siguiente propuesta tiene la finalidad de servir como referencia para el diseño de piloto**. No obstante, la arquitectura definitiva tendrá que ser resultado de las etapas de pilotaje propuestas en la sección de Hoja de Ruta.

Figura 11. Arquitectura técnica de referencia



Fuente: Elaboración propia

La siguiente arquitectura técnica supone que las entidades utilizan servicios de nube. Sin embargo, **no hay requisitos específicos sobre qué tipo de almacenamiento se debe utilizar**. Se pueden usar servicios de nube o bien almacenamiento físico. Esta decisión tendrá que tomarse a partir de la preferencia de las partes involucradas.

A continuación las distintas capas que conforman la arquitectura técnica. Es importante destacar que **entre cada capa y entre cada bloque, se deben establecer APIs.**

Usuario: se asume que hay un usuario en la primera capa que quiere proporcionar, almacenar y/o obtener acceso a los datos en el Data Commons.

Interfaz: en esta capa se asigna un mecanismo de inicio de sesión para que el usuario pueda acceder al Data Commons. **En esta etapa, X-ROAD se puede aprovechar como mecanismo de autenticación**. El interfaz es también el *front-end* del servicio, y por tanto lo que los usuarios ven y con lo que interactúan.

Procesamiento: en esta capa, hay un Ingreso que es el punto en el que se introducen los datos. En este punto, es probable que se lleve a cabo una **validación de seguridad de los datos permitidos en el Data Commons**, para garantizar que todos los datos del sistema sean de buena calidad. Es importante validar si la entrada es correcta o no.

Almacenamiento: debe haber un mecanismo de almacenamiento para los datos que se están compartiendo. Se puede elegir entre diferentes opciones, y la mejor opción para este Data Commons en específico dependerá de cómo el usuario vaya a solicitar los datos. Los posibles mecanismos de almacenamiento incluyen:

- **Blob storage o bases de datos NoSQL**: Este tipo de almacenamiento en bloques está optimizado para almacenar cantidades masivas de datos no estructurados. Los datos no estructurados son datos que no se adhieren a un modelo o definición de datos particular, como los datos de texto o binarios.
- **Querying storage o bases de datos SQL:** SQL es un lenguaje estándar para acceder y manipular bases de datos. SQL significa Structured Query Language (lenguaje de consulta estructurado) y permite acceder y manipular las bases de datos.
- **Almacenamiento triple (triple storage), tecnologías semánticas (semantic technologies):** tecnología de bases de datos que procede de una rama de la inteligencia artificial. La tecnología semántica es un conjunto de métodos y herramientas que proporcionan medios avanzados para categorizar y procesar datos, así como para descubrir relaciones dentro de conjuntos de datos variados. Esta opción suele ser algo lenta y bastante costosa.
- **Blockchain:** No se recomienda Blockchain como forma de almacenar datos en un Data Commons. Es poco probable que el rendimiento sea bueno, además suele ser muy lento y costoso.

Después de este punto, el flujo de datos vuelve al usuario a través de las mismas capas mencionadas anteriormente, pero con diferencias específicas.

Procesamiento: Aquí habrá una **solicitud de acceso**, ya que el usuario pedirá que se extraiga un subconjunto de datos de la base de datos. Probablemente se necesitará una forma de visualizar los datos cuando se extraigan del sistema. Asimismo, habrá un **egreso**. Aquí es cuando, justo antes de que los datos vuelvan a salir hacia el usuario, se valida que el proceso se ha hecho correctamente y que el usuario tiene derecho a ver aquello que ha solicitado.

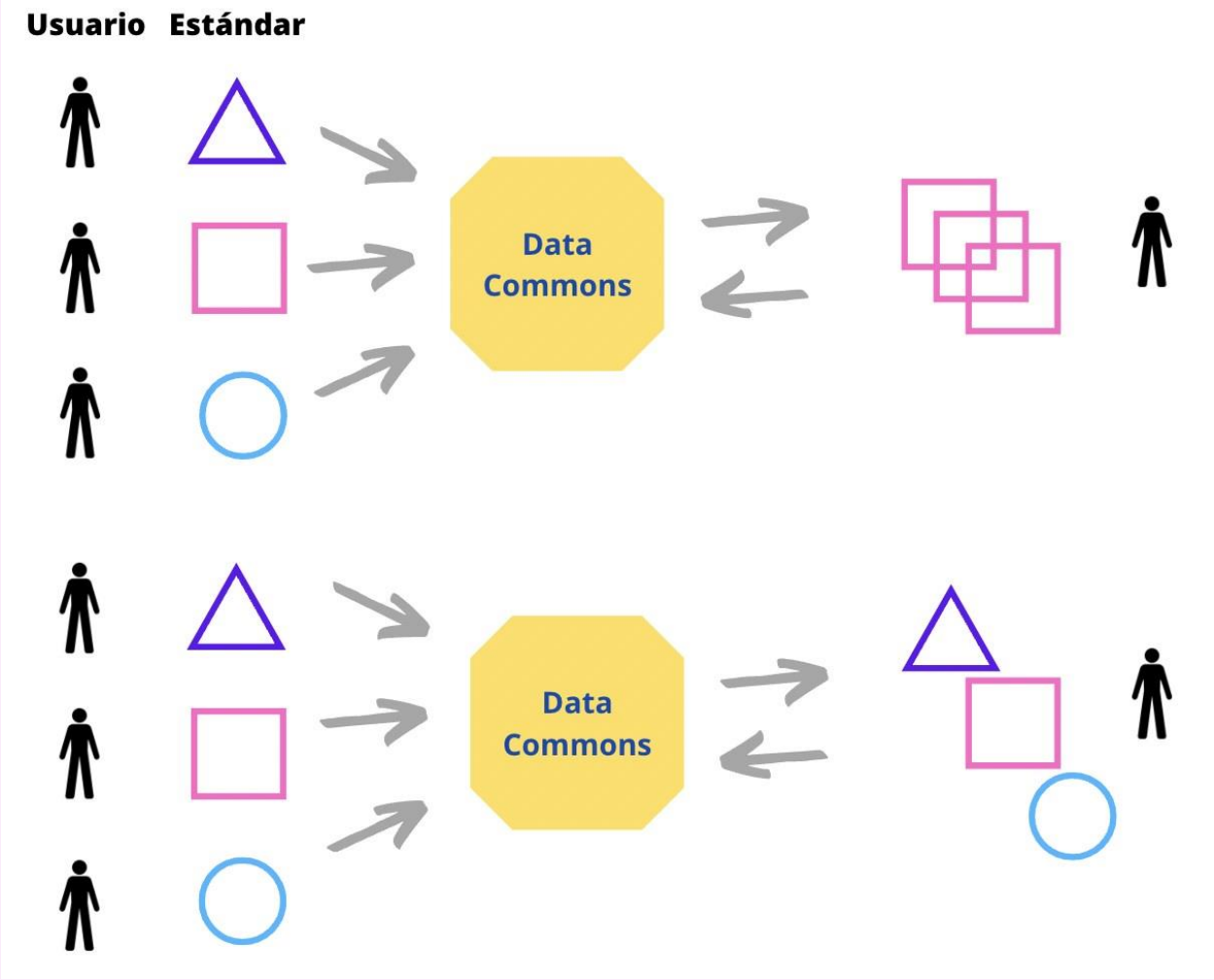
Interfaz: aquí el usuario podrá interactuar con los datos que solicitó.

Usuario: en esta etapa es importante decidir cómo los usuarios pueden obtener los datos. Algunos posibles formatos son CSV (archivo plano), API o en formato de imagen (visualización). La forma en que el usuario quiera que salgan los datos definirá el estándar necesario para los mismos. Al considerar cómo salen los datos, hay que tener en cuenta estos tres casos de uso propuestos (API, CSV, imagen).

Los estándares de datos dependerán de cómo se espera que salgan los datos del sistema. No hay normas de datos intrínsecas a un Data Commons, todo depende del caso de uso. Sin embargo, existen normas de datos en torno a casos de uso específicos, como los geodatos.

Como se describe en la siguiente figura, las necesidades del usuario informarán las elecciones de la primera capa (capa de usuario), y la capa de usuario (cómo quiere obtener los datos del sistema) determinará el resto de las capas. **Las elecciones tecnológicas deben basarse en las necesidades del usuario**.

Figura 12. Estándares y necesidades de usuario



Fuente: Elaboración propia

3.5.3. Seguridad

En cuanto a la seguridad de la información, es fundamental contar con mecanismos de verificación. En este caso, **X-Road podría utilizarse como sistema de autenticación en el Data Commons**.

Además, es crucial que todos los datos que se compartan en el Data Commons se encuentren anonimizados. Asimismo se debe evitar el uso de información personal, sensible y de casos específicos. En este contexto, se debe cumplir con la

guía para la anonimización de bases de datos en el Sistema Estadístico Nacional (DANE, 2018).

Además de la seguridad proporcionada por el protocolo de intercambio de información X-Road, la protección de los datos personales por parte de todos los actores del Data Commons, en especial por los recolectores, debe responder a las buenas prácticas de ingeniería en software y hardware. Asimismo, debe responder a la **normativa actual de protección de datos personales, al Modelo de Seguridad y Privacidad de la Información publicada por MinTIC** (MinTIC, 2016) y a los **capítulos de seguridad dentro de las guías del Marco de Referencia de Arquitectura** (MinTIC, 2019).

En este caso, hay que asegurarse de que se está construyendo un sistema con integridad. **Se recomienda una arquitectura sencilla que pueda ampliarse posteriormente si es necesario**. Lo más probable en el presente caso de uso es que inicialmente no se compartirán datos personales. Aunque sí se recomienda incluir datos personales en el futuro si se comprueba que añadiría valor hacerlo.

Sin embargo, hay una decisión importante que debe tomarse en cuanto a la seguridad: la anonimización de los datos**. La anonimización de los datos puede realizarse fuera de Data Commons**, lo que significaría que cada Ministerio tendría que anonimizar sus propios datos y sólo se podrían cargar en la plataforma los datos ya anonimizados. Asimismo, la anonimización **podría realizarse dentro de la plataforma**. En el segundo caso, la información personal se almacenará dentro de la plataforma y **se anonimizaría en la fase de solicitud de acceso**. Las respuestas a las solicitudes, por tanto, se darían en formato agregado. Es importante añadir que si los datos se anonimizan fuera de la plataforma en lugar de dentro de Data Commons, resulta más difícil consultar los datos.

3.5.4. Equipo: recursos humanos

Al desarrollar y poner en marcha un Data Commons, es **fundamental contar con un equipo multidisciplinario** que pueda crear y hacer funcionar el servicio de forma sostenible. Es importante que el equipo esté formado por personas con una variedad de habilidades y conocimientos[7]. También es importante que las personas que participan en la toma de decisiones se involucren y formen parte del equipo. Esta es

[7] Basado en *GOV UK Service Standard Manual*. https://www.gov.uk/service-manual/service-standard/point-6-have-a-multidisciplinary-team

una lista de posibles funciones para formar un equipo para el diseño e implementación del Data Commons.

Es importante mencionar, sin embargo, que se trata de información de referencia que podrá ser ajustada de acuerdo a las necesidades de entidades líderes y usuarios. Los recursos humanos y financieros disponibles, así como las necesidades reales de las entidades, determinarán la selección final del equipo. Se recomienda que, sea cual sea la selección final, el equipo cuente con servidores públicos que conozcan el sector agropecuario y las entidades participantes en profundidad.

Figura 13. Roles dentro de un equipo para diseñar e implementar el Data Commons

Rol	Responsabilidades
Líderes del gobierno	Apoyar las iniciativas como la puesta en marcha de un Data Commons. Son clave para obtener fuentes de financiación y asegurar que el servicio se adapta a las prioridades de política pública.
Reguladores	Crear los marcos regulatorios bajo los que operan los actores involucrados.
Titular/Diseñador del servicio	Asegurarse de que lo que está construyendo encaja con las prioridades de las entidades, y contribuye a definir cuál es el objetivo futuro del Data Commons. Asegurarse de que el Data Commons satisface las necesidades de los usuarios y de que el servicio es accesible para todos.
Director de la entrega/ Delivery manager	Establecer el entorno ágil que el equipo necesita para construir e iterar un Data Commons centrado en el usuario, eliminando los obstáculos o "bloqueos" al progreso. Ayudar al equipo de servicios a organizar mejor su propio trabajo de forma autónoma y asegurarse de que la accesibilidad se tiene en cuenta en cada función o actividad en la que trabaja el equipo.
Analista de negocios/ Business Analyst	Definir las competencias que necesitará un servicio y determinar dónde pueden encontrarse (por ejemplo, en un contratista externo). Comprobar que hay presupuesto para cubrir el enfoque propuesto.

	Analizar cuánto dinero necesita el proyecto para su funcionamiento continuo, determinar los riesgos y proponer soluciones.
Investigador usuario/ User Experience Researcher (UX)/	Llevar a cabo actividades de investigación de usuarios para ayudar a los equipos técnicos a conocer a profundidad las necesidades de los usuarios.
Diseñador de experiencia de usuario (UX)	Usar la investigación del usuario para definir los procesos que llevarán al usuario a cubrir su necesidad.
Desarrollador Front-end (UI)/ Front-end developer (UI)	Codificar la interfaz gráfica con la que interactúan los usuarios cuando visitan un sitio web o utilizan una aplicación. Los desarrolladores de front-end pueden especializarse en JavaScript, CSS, HTML o todo lo anterior. Los desarrolladores de front-end son los responsables del aspecto de un sitio web o una aplicación.
Desarrollador back-end/ Back-end developer	Crear componentes y funciones en el lado del servidor, al que accede indirectamente un usuario a través de una aplicación o sistema de front-end.
Arquitecto técnico	Proporcionar liderazgo técnico y diseño arquitectónico de sistemas. Trabajar con equipos de entrega y terceros para decidir los requisitos técnicos y las mejoras para el desarrollo de software y el funcionamiento de la web. Asegurarse de que las plataformas, los productos, las transacciones y las arquitecturas de sistemas sean sólidos, escalables, abiertos y seguros.
Encargado de medición de impacto y control de calidad	Encargarse de supervisar el éxito del Data Commons y de medir su impacto, así como de garantizar que cumpla los objetivos establecidos. También puede trabajar a corto plazo con su equipo para incorporar la calidad a todo lo que hacen.
Analista de datos	Encargarse de obtener información útil a partir del análisis de los datos. Debe conocer y haber utilizado lenguajes como R, Python, SQL o Scala.
Artista de datos	Encargarse de crear visualizaciones con los datos para facilitar su interpretación, análisis y comunicación.

Fuente: Elaboración propia (basada en GOV UK, 2022)

4. Análisis de costos

El presupuesto requerido para el desarrollo de la siguiente fase del proyecto se basará principalmente en el costo de los recursos humanos. Para ello, se recomienda tener en cuenta los roles recomendados para cada fase de proyecto presentados en la sección 5[8]. Dicha propuesta considera los **roles necesarios para construir todos los elementos del Data Commons**, tales como arquitectura técnica, investigación de usuario, diseño de interfaz, entre varios otros.

A continuación, un desglose de los recursos humanos requeridos para cada fase. Se consideran 22 días de trabajo promedio por mes.

Figura 14. Desglose de roles y días requeridos

	Descubrimiento	Alfa	Beta privada	Total de días
Titular del servicio	22	33	77	132
Gestor de entrega	22	33	77	132
Investigador de usuario	44	66	154	264
Analista de negocio	22	33	77	132
Arquitecto técnico	11	33	77	121
Encargado de impacto y control de calidad	22	33	77	132
Desarrolladores			154	154
Diseñador de experiencia			77	77
Total				**1144**

Fuente: elaboración propia

Como referencia, las tarifas para un equipo de proyecto como el que se propone podrían girar en torno a los 500 USD por día. **No obstante, el cálculo deberá ajustarse a los estándares laborales en Colombia**.

Asimismo, para llegar a un presupuesto preciso son necesarias dos acciones por parte de las entidades líderes:

1. **Definir qué recursos humanos aportará cada una al equipo**. Lo anterior deberá basarse en el presupuesto y recursos disponibles en cada entidad.

2. **Decidir si el desarrollo del proyecto se llevará a cabo únicamente por servidores públicos o si se contratarán proveedores externos**. Es posible que las entidades no cuenten con personal para el desarrollo de algunos

[8] Para mayor información sobre las fases del proyecto, referirse a la sección Hoja de Ruta.

elementos del proyecto (como interfaz, programación, diseño de experiencia, etc.) y se requiera contratar servicios.

a. Por ejemplo: el equipo propuesto en la sección 5 contempla la integración de diseñadores de experiencia y desarrolladores. Dichos elementos participarán en la construcción de la interfaz del Data Commons. Para ello, existe la posibilidad de contratar proveedores privados o llevar a cabo el proceso de forma interna. Esta decisión deberá basarse en las necesidades de usuario identificadas y recursos disponibles. Lo anterior influirá en el costo total del proyecto.

En cuanto a infraestructura, se puede asumir que los datos del presente caso de uso serán almacenados en la nube. En ese sentido, **es posible trabajar con el servicio de la nube con el que se cuenta en este momento**. **Por lo tanto, no se requiere contratar servicios adicionales**. En el caso de que se quisiera cambiar de proveedor de servicios de nube (cualquier proveedor de nube que ofrezca uno de los sistemas de almacenamiento propuestos en la sección de tecnología sería válido), entonces esto implicaría un costo adicional, cuyo valor dependería del servicio que se seleccione.

5. Producto Mínimo Viable

Para construir una plataforma como un Data Commons, es necesario plantear el Producto Mínimo Viable previo al inicio de la fase de pilotaje. Es decir, es importante tener en cuenta cuáles deben ser las **funcionalidades mínimas** con las que debe contar el primer prototipo del servicio para ser utilizado por los usuarios. En ese sentido, se han identificado las siguientes funcionalidades:

1. Crear un **sistema de autenticación de los usuarios** que contribuya al intercambio seguro de información. Tiene la finalidad de crear un sistema de inicio de sesión para poder hacer uso del servicio, así como funciones adicionales para corroborar la identidad del usuario.

2. Los **datos compartidos deben seguir las reglas y estándares establecidos** para asegurar su interoperabilidad y seguridad.

3. El prototipo debe permitir a los usuarios **ingresar datos siguiendo los estándares establecidos**.

4. El prototipo debe permitir a los usuarios **solicitar el acceso a los datos y posteriormente extraerlos** en los formatos que se hayan establecido previamente.

Idealmente, estos procesos deberían realizarse de forma automatizada a través de una interfaz diseñada a partir de investigación de usuario. No obstante, para efectos de tiempos, existe la posibilidad de que el primer prototipo funcione provisionalmente de forma manual. Es decir, se puede realizar el intercambio por otros medios (como plataformas de nube o correo electrónico) sin la necesidad de una interfaz construida específicamente para el caso de uso.

6. Hoja de Ruta

6.1. Enfoque

Para la fase de implementación de Data Commons, se sugiere que se siga una metodología de entrega ágil (*agile*). Los métodos ágiles permiten a los equipos construir rápidamente, probar lo construido e iterar su trabajo basándose en retroalimentación constante y en otros datos útiles. Los métodos ágiles pueden ser muy diferentes para las personas acostumbradas a los métodos tradicionales de desarrollo de software en cascada.

Con los métodos en cascada, el proceso es secuencial. Se empieza por reunir los requisitos, hacer planes y pasar por los procesos de adquisición. Luego se diseña el producto y se construye. En la fase final se prueba y se pone a disposición del público. Sólo en esta fase final se obtiene retroalimentación y se descubre si el producto funciona para los usuarios. Sólo se tiene una oportunidad para obtener un resultado efectivo en cada etapa del proyecto. Esto debido a que, una vez terminados, no se vuelve a trabajar en los pasos anteriores. En ese sentido, el enfoque ágil es diferente porque las tareas de investigación, planificación, diseño, construcción y pruebas se llevan a cabo de forma simultánea.

Un proyecto ágil tiene tres fases: descubrimiento, alfa y beta. A continuación se describe más concretamente cada una de estas fases y lo que hay que hacer para sacar adelante la implantación del Data Commons.

Algo importante a tener en cuenta es que, como en cualquier otro proyecto, la realidad puede cambiar las ideas o planes iniciales. Debido a que se aprende sobre la marcha,

es posible que se descubran cuestiones que obliguen a cambiar el curso de la acción. De igual manera, es posible que el equipo advierta que se necesitan determinadas habilidades o personas en distintos momentos del proyecto. Aquí es donde entra la pertinencia del modelo ágil: **se pueden tomar decisiones basadas en la información que se obtiene sobre la marcha, lo que lleva a enfocar el trabajo en necesidades reales**.

6.2. Cronograma

Un proyecto como el que se pretende llevar a cabo debe tener una duración aproximada de **6 meses dividida en tres distintas fases**. De igual manera, el **equipo deberá ser multidisciplinario y dinámico**. Es decir, se debe contar con personas con perfiles variados, las cuales participarán en etapas diferentes del proyecto. A continuación, un cronograma general sugerido para la implementación del Data Commons:

Figura 15. Cronograma

Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6
Descubrimiento	Alfa	Alfa/Privada Beta	Beta Privada	Beta Privada	Beta Privada

Fuente: Elaboración propia

Descubrimiento (Mes 1)

Durante la fase de descubrimiento, el **equipo comprenderá a un nivel más profundo el problema que se pretende resolver con la creación del Data Commons**. Esto implica anotar preguntas y suposiciones sobre cómo funcionaría el Data Commons y el contexto actual que se quiere cambiar.

Aquí hay que plantearse preguntas realmente detalladas sobre la implementación: **analizar en detalle la variedad de formas en que se tienen actualmente los datos y cuáles son los diferentes estándares de datos existentes (mapeo detallado)**. La forma de hacerlo (aprender sobre los estándares de datos) es **comprender con mucho más detalle cómo es el usuario que consulta los datos y cómo es el usuario que los introduce**. En resumen, la fase de descubrimiento debe centrarse en comprender realmente cómo van a utilizar el sistema los usuarios.

Se sugiere un descubrimiento de 4 a 6 semanas de duración.

El equipo debería estar formado aproximadamente por los siguientes roles:

- **1x Titular del servicio**: para entender el panorama general y la complejidad del problema y posteriormente definir estrategias para abordarlo.
- **2x Investigadores de usuarios**: para llevar a cabo investigación cualitativa y cuantitativa sobre los usuarios para averiguar necesidades y entender la forma como van a interactuar con el Data Commons.
- **1x Analista de negocio**: para convertir lo abstracto en historias de usuario detalladas y presentar un caso de negocio (que incluye análisis de costos y riesgos).
- **1x Arquitecto técnico**: para brindar una perspectiva técnica en la fase de descubrimiento y aconsejar sobre la arquitectura a construir. En esta fase, no es necesario que el arquitecto técnico trabaje a tiempo completo.
- **1x Gestor de entregas**: para asegurarse de que el equipo es eficaz y facilitar el flujo de trabajo.

Alfa (Mes 2 / Primera mitad del Mes 3)

En la fase alfa se prueban diferentes soluciones a los problemas identificados en la fase de descubrimiento. La fase alfa debe enfocarse en construir prototipos y probar diferentes ideas. La fase alfa es una oportunidad para explorar nuevos enfoques. Se deben construir soluciones que sean lo suficientemente complejas para poder ponerlas a prueba. Es probable que muchas de las ideas que prueben, se desechen al final de la fase alfa. Al final de la fase, el equipo debería estar en condiciones de decidir cuáles entre las ideas que se hayan probado merecen ser abordadas en la fase beta. Una fase alfa debería durar entre 6 y 8 semanas.

En este caso, la fase de alfa podría servir para **explorar las diferentes soluciones técnicas o interfaces de usuario que podrían funcionar para el Data Commons** que se pretende implementar. Se podrían explorar cuestiones como:

- ¿Qué interfaz de usuario debería utilizar este Data Commons? ¿Va a ser suficiente con compartir algunos enlaces y visualizaciones de datos o va a construir un sitio web o servicio específico para ello?
- ¿Cómo se va a integrar X-ROAD en el Data commons?

Al final de Alfa, el equipo debería tener una idea clara sobre las características del servicio y sobre el equipo necesario para la implementación.

El equipo debería estar formado por los siguientes roles:

- **1x Titular/Diseñador del servicio**: para entender el panorama general y la complejidad del problema y posteriormente definir estrategias para abordarlo.
- **2x Investigadores de usuarios**: para llevar a cabo investigación cualitativa y cuantitativa sobre los usuarios para averiguar necesidades y entender la forma como van a interactuar con el Data Commons.
- **1x Analista de negocio**: para convertir lo abstracto en historias de usuario detalladas y presentar un caso de negocio.
- **1x Arquitecto técnico**: para tener una perspectiva técnica en la fase de descubrimiento y aconsejar sobre la arquitectura a construir.
- **1x Gestor de entregas**: para asegurarse de que el equipo es eficaz y facilitar el flujo de trabajo.
- **1x Encargado de medición de impacto y control de calidad**: Estas personas prueban el servicio para comprobar su funcionamiento. Son personas metódicas. Su trabajo consiste en comprobar la integridad del servicio.

Mitad del Mes 3 - Final del Mes 6 - Beta privada

Durante la beta privada es el momento en el que se toma la mejor idea de la alfa y se construye el servicio para preparar su puesta en marcha. Durante la beta, se debe asegurar que la solución elegida funciona lo mejor posible, llevando a cabo investigación de usuarios y empezando a recopilar datos sobre el éxito del servicio en función de las métricas de éxito que se identificaron en la fase alfa. **Es crucial iterar el servicio en función de lo que se aprenda**. Asimismo, se debe asegurar que el servicio es accesible para los usuarios.

En este caso, la **fase beta privada es el momento en el que se construirá el Data Commons**. El equipo habrá identificado la solución técnica que mejor funciona para este caso concreto y tendrá objetivos que cumplir —así como métricas para evaluar si los está cumpliendo. El equipo deberá considerar los comentarios de los usuarios sobre el funcionamiento del Data Commons y ajustarlo de acuerdo a éstos. Esta fase debería durar unos 3 meses. Al final de la misma, se debería tener un Producto Mínimo Viable (MVP) para un Data Commons que permita ser utilizado por los usuarios. Se trata de una versión básica de partida para el Data Commons. Siempre existe la posibilidad de ajustar o iterar esta versión en futuras fases.

El equipo beta debería incluir los siguientes roles:

- **1x Titular/Diseñador de Servicio**: para entender el panorama general y la complejidad que lo rodea y decidir cómo reducirlo.
- **2x Investigadores de usuarios**: para centrarse en los usuarios y averiguar sus necesidades, y cómo van a interactuar exactamente con el Data Commons.

- **1x Analista de negocio**: para poder convertir lo abstracto en historias de usuario detalladas y presentar un caso de negocio.
- **1x Arquitecto técnico**: para tener una perspectiva técnica en la fase de descubrimiento y aconsejar sobre la arquitectura a construir.
- **1x Gestor de entregas**: para asegurarse de que el equipo es eficaz y facilitar el flujo de trabajo.
- **2x Desarrolladores (front end y back end)**: para programar el servicio.
- **1x Encargado de medición de impacto y control de calidad**: Estas personas prueban el servicio para comprobar su funcionamiento. Son personas metódicas. Su trabajo consiste en comprobar la integridad del servicio.
- **1x Diseñador de experiencia de usuario (UX)**: para centrarse en el diseño del proceso que permitirá al usuario cubrir su necesidad.

Es importante tener en cuenta que, a medida que vaya descubriendo más cuestiones sobre el Data Commons que se quiere construir, **es posible que se requiera incorporar más funciones de la lista anterior** (por ejemplo, un artista de datos o un analista de datos si tiene necesidades de visualización o interpretación de datos).

Después de Beta Privada

Una vez terminado de definirse el producto mínimo viable (MVP), se podrán ir construyendo y definiendo versiones más complejas y completas del Data Commons según las necesidades de usuarios que vayan surgiendo.

Llegados a este punto, todos los usuarios podrán acceder e interactuar de forma normal con el Data Commons. Es decir, el Data Commons comenzará a operar formalmente. No obstante, la iteración deberá seguir siendo una constante. Será necesario proporcionar mecanismos para que el usuario proporcione retroalimentación y se puedan solucionar posibles problemas que puedan surgir.

Este también será el momento en el que se decida si se desea ampliar el acceso a otros usuarios, o incorporar otros conjuntos de datos —quizás incluyendo datos personales— al Data Commons.

7. Bibliografía

GOV UK Service Standard Manual. (2022). Have a multidisciplinary team. Disponible en: https://www.gov.uk/service-manual/service-standard/point-6-have-a-multidisciplinary-team

Mozilla Foundation. (20210. A Practical Framework for Applying Ostrom's Principles to Data Commons Governance. Disponible en https://foundation.mozilla.org/en/blog/a-practical-framework-for-applying-ostroms-principles-to-data-commons-governance/

OCDE. (2022). Purchasing power parities (PPP) (indicator). Disponible en: doi: 10.1787/1290ee5a-en

ODI. (2019). Data Ecosystems Mapping Tools. Disponible en https://theodi.org/article/data-ecosystem-mapping-tool/

ODI. (2019). Data Ecosystems Mapping. Disponible en: http://theodi.org/wp-content/uploads/2019/06/ODI-Data-Ecosystem-Mapping-%E2%80%93-print-at-home-guide-A4-%E2%80%93-2019-06-26.pdf

ODI. (2019). Data Ecosystems Mapping Tools. Disponible en: https://theodi.org/article/data-ecosystem-mapping-tool/

Wrike. (2022). Fundamentos de la metodología Agile. Disponible en: https://www.wrike.com/es/project-management-guide/fundamentos-de-la-metodologia-agile/