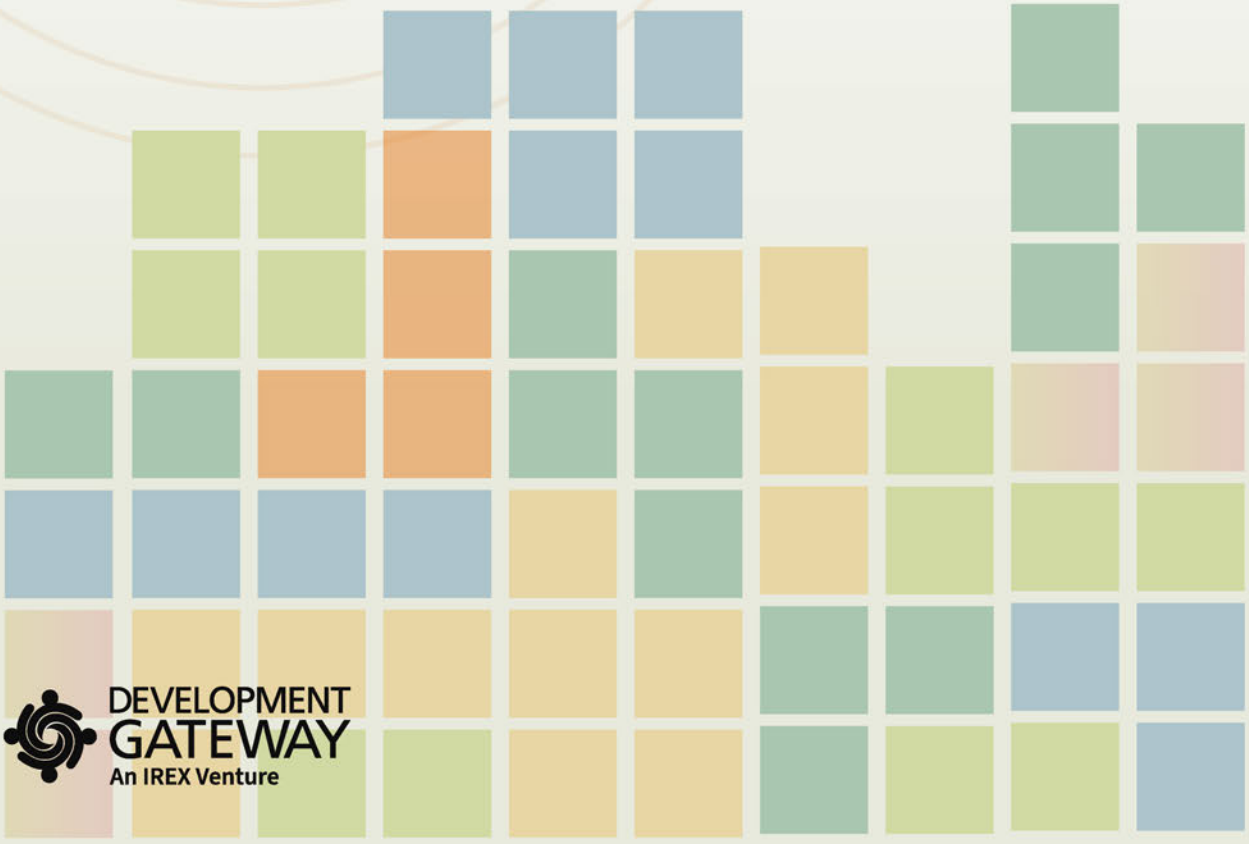
Demystifying Interoperability

Insights from our work in Ethiopia's agriculture sector



DEVELOPMENT GATEWAY
An IREX Venture

Acknowledgements

- Written by

Beverley Hatcher-Mbu
Fernando Ferreyra
Mihai Postelnicu
Viorel Cihai
Yosef Getachew
Yohannes FIkre
Selam Abraham Kassa (PhD)
Tom Orrell
Khyatee Tewari

- External Reviewers

Julianna Drinan
Josh Woodard

- Copy Edited by

Daniel Nyaga

- Designed by

Xinyu (Breeze) Hu

Contents

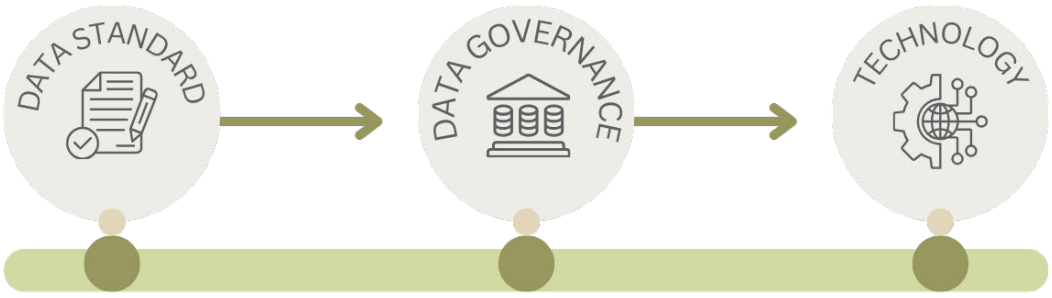
Executive Summary

Overview

Government offices are filled with siloed, sector-specific digital systems that strain their capacity to make decisions, use data effectively, and achieve ambitious sustainable development goals. Public investments in digital development, transformation, and infrastructure can only meet citizen needs if data and systems are consolidated and interoperable.

While interoperability is a sensible approach to building digital public infrastructure, transforming existing systems is easier said than done. Limited resources and expertise, lack of training, inadequate hardware and software infrastructure, and concerns over data privacy are some of the key factors that limit the impact of systems or lead to their discontinuation. The data that feeds these systems also needs to be standardized using global best practices and common languages and frameworks that can be understood across multiple platforms and adapted to the local context. In increasingly digital economies, systems and their data need to be relevant and responsive to citizens' specific conditions and needs With emerging challenges like climate change that require more complex data, and new technologies like artificial intelligence (AI) that require large volumes of data, systems need to be agile and scalable to advance solutions and fully harness innovation.

This paper discusses, in practical terms, what goes into implementing interoperable solutions in partnership with public administrations. Based on 20+ years of DG's experience, the paper demystifies key components needed to build robust, resilient, and interoperable data systems, focusing on the "how" of data standardization, data governance, and implementing technical infrastructure.



Setting up Data Standards

Data must be standardized for consistent and seamless exchange between different systems. Without standardization, systems cannot "talk" to one another, limiting users' ability to gain deeper insights

from combined data sources. Standardization can be a long process as not all data can be acted upon at once; certain datasets and systems need to be prioritized with a view to scale over time. Moreover, data standards are about the people who use them: it takes time for the people who know their data best to change data collection and usage methods over time. Inclusive data standards processes draw diverse stakeholders together for a win-win scenario: everybody can make better decisions that drive productivity when they can see the same picture together.

Standardization should combine the best of both worlds: international best practices, and local customization. Standards should align with global best practices that allow for cross-border and global linking of data down the road; they should also reflect local needs, such as accounting for indigenous breed varieties (in the agricultural/livestock context). For example, livestock movements are not confined to borders - as such, standardization methods will ideally become uniform across nations over time. Chapter 2 walks through the steps the aLIVE program has taken to date to develop and cascade data standards. Going forward, continual standard updates and additional financial resources will be needed to keep up with emerging data needs. Data owners will need to continue to "own" the process, to ensure changes are mainstreamed into data management practice.

Cultivating Adaptable Data Governance

Many countries do not have a clear "whole-of-government" policy on data sharing and access. This absence matters for interoperability because systems will struggle to remain usable without a consistent flow of quality data that can be shared. To tackle this challenge, Chapter 3 looks at the practical and legal frameworks that govern how key digital systems in the Ethiopian Ministry of Agriculture manage and control data availability, usability, integrity/quality, and security. Incorporating a data governance lens into the technical development of systems increases data accuracy and consistency by taking the guesswork out of data sharing.

From evaluating documents to engaging with system owners to understand their existing data governance practices, the aLIVE program collaboratively built templates grounded in reality, developing sharing and access approaches adapted to existing practice, but scaleable for use in national and multi-system governance in the future. Similar to data standardization, data owners need to see, in real-time, the shared value of data exchange to commit to new processes and approaches. When we say "interoperability requires robust data governance," this layered process of people, policies, and framework development is what we mean.

Building Up the Technology

What counts as "good" digital public infrastructure is hard to define, especially when there are several well-known problems. For example, there are systems that cannot be modified without re-engaging vendor support; platforms with rapidly outdated technology; and difficult-to-manage storage solutions that lead to data loss. Smart development of digital infrastructure boosts efficiency and builds trust, making services more responsive to citizen needs. Solving these challenges requires a holistic approach

that addresses software, hardware, and local IT infrastructure and capacity.

Software tools must be highly maintained, tested, and proven in production environments. They should have a history of stable releases and a reliable codebase to ensure key system functionalities stay intact after changes are made. Ideally, these tools need to be based on open standards and open source code, with permissive licenses such as Apache 2.0, BSD, or MIT. A vibrant community is also essential, as this ensures that operating costs remain low, IT specialists' expertise remains high, and the organization is not locked into proprietary, closed architecture.

BEST PRACTICE	IMPLEMENTATION
Vendor Management and Licensing	When adding or improving software components in your infrastructure, ensure the source code is well-documented and appropriately licensed. If proprietary, negotiate a special license that guarantees future access and flexibility. Aim to avoid vendor lock-in by ensuring your solution can be maintained or transferred without relying on a single vendor.
Cost Management and Sustainability	Conduct regular reviews of operational and maintenance costs to ensure the solution remains financially sustainable. Ensure that the platform can scale efficiently in terms of performance and price as usage grows.
Technology and Infrastructure Updates	Review and update the technology stack regularly to prevent technical debt and security vulnerabilities. Consider the platform's full lifecycle impact, from development to decommissioning.
Data Management and Governance	Implement data retention, archival, and disposal policies to manage costs and comply with changing regulations.
Documentation and Knowledge Transfer	Maintain up-to-date and accessible documentation for all components, processes, and decisions. If you update any component, ensure the vendor or participating team updates the related documentation.
Team Development and Training	Invest in continuous learning and upskilling so the team can keep pace with technological advancements.
Governance and Compliance	Ensure the platform complies with current and anticipated regulations, adjusting as needed.

Recommendations

One program cannot address all problems! The following recommendations from Chapter 5 reflect the systemic shifts required to build better technology across sectors as well as the forward-looking opportunities that set up public technology for the future.

1. View digital transformation on a maturity axis, rather than a transition from point A to B.
2. Understand what already exists to set priorities.
3. Use what exists to define the scope.
4. Institutionalize sustainability from the outset in every dimension: human, institutional, technical, and financial.
5. Take a portfolio view of digital transformation that puts interoperability at the center. Country digital roadmaps can guide priorities across multiple digital needs.
6. Recognize that transformation requires champions.
7. Ensure data standards stretch across sectors and regions.
8. Keep watching and actively collaborating with telecommunications outfits as the numbers of local cloud providers continue to grow.
9. Invest in national legal frameworks for responsible data sharing.
10. Do not forget the people.

Chapter 1

Setting the Scene for Interoperability

What is the Problem?

The significant investments in digital platforms over the last two decades have resulted in the widespread proliferation of siloed systems across multiple sectors.[1] This uncoordinated proliferation strains government capacity, reduces the impact of each system, and makes data-informed decision-making a massive challenge across every sector. For example, the COVID-19 pandemic exposed[2] how an inability to combine data across systems only exacerbates poor health outcomes for everyone involved. To bring disparate systems together and make them fit for purpose, both for citizens and policymakers, more consolidation and interoperability will be required going forward. Moreover, such fusion will help overcome the challenges associated with digital development, digital transformation, and the overhaul of digital public infrastructure.

Despite being a fairly established concept, the interconnection of systems faces complex challenges. The technical challenges of interoperability are well documented, with issues such as resource scarcity, limited expertise and training, inadequate digital infrastructure, and concerns over data privacy and security being prevalent.[3] However, political challenges are equally significant and warrant comprehensive exploration. As Carletto points out, "for governments and the international community to navigate this period of upheaval to protect vulnerable populations, a transformation within data systems will be required."[4]

Tensions between competing commercial interests, political power differences, and the collective good repeatedly arise as more governments pursue transformation. These tensions can lead to data silos in which valuable information is kept isolated due to proprietary concerns. They can also result in systems that cannot be effectively maintained due to licensing and intellectual property restrictions, hindering the potential for collaborative improvement and innovation.

---

1 Rucker, D., Hasan, A., Lewis, L., & Tao, D. (2020). *Advancing interoperability together globally.* Global Digital Health Partnership. https://gdhp.health/wp-content/uploads/2022/11/Advancing-Interoperability-Together-Globally-2.pdf

2 Greene, D. N., McClintock, D. S., & Durant, T. J. S. (2021). Interoperability: COVID-19 as an Impetus for Change. *Clinical Chemistry,* 67(4), 592–595.

3 Building an interoperable space for smart agriculture. (2023). *Digital Communications and Networks*, 9(1), 183–193.

4 Carletto, C. (2021). Better data, higher impact: improving agricultural data systems for societal change. *European Review of Agricultural Economics*, 48(4), 719–740.

Separately, there is a tension between the need for global standards on data interoperability and the need to adapt these standards to local contexts and specific national needs. The drive for global standards on data interoperability aims to create a common language and framework that enables seamless data exchange and integration across different systems and platforms. However, applying these global standards, without modification, often clashes with the national and citizen realities.

Agricultural practices, for example, vary significantly across regions, necessitating localized standards that address specific requirements and ensure data relevance and accuracy. From the naming of local cattle varieties to community expectations around data privacy, global standards may overlook cultural nuances, which can affect individual and collective trust in data-sharing processes.

On one hand, there is a need for interoperable data to support broad-scale decision-making and resource allocation. On the other hand, there is a pressing need to ensure that data systems are relevant and responsive to the specific conditions and challenges faced by citizens and increasingly digital economies. This balance is not easily achieved, as it requires ongoing dialog and collaboration between global standard-setting bodies and local stakeholders. Interoperability also, ultimately, needs to be future-looking: it needs to permit systems (and the data contained within them) to remain relevant for existing users while meeting the needs of new users and addressing emerging decision-making requirements.

Lastly, there is a lack of awareness and capacity regarding data management and the value of interoperable systems. Transformation requires greater awareness of standards, better compliance, improved policy and governance methods, up-to-date data science capacity, and a focus on safe and cost-effective data collection processes.[5] Collectively, these aspects can block the development of effective interoperable data systems that serve as the backbone for digital transformation.

5 Williamson, H. F., & Leonelli, S. (2022). *Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development.* Springer Nature.

How Are We Defining and Approaching Interoperability?

Put simply, "interoperability is the ability to link data from different sources in a standardized and contextualized way."[6] Interoperability allows for the seamless exchange and utilization of information, which is crucial for informed decision-making and strategic investments.[7] It enables different information systems, devices, and applications to access, exchange, integrate, and cooperatively use data in a coordinated manner, within and across organizational boundaries.[8]Ultimately, interoperability lets computers and people look at a variety of data sets together to benefit from their collective information.

Interoperability, when done well, can not only improve the efficiency of service delivery but also foster the creation of new ideas, collaborative efforts, and trends analytics which are vital for data-driven decisions and forecasting. For example, the African Soil Information Service (AfSIS) utilizes interoperable data systems to provide detailed soil health information that is vital for guiding soil management practices and enhancing crop yields tailored to specific regional conditions.[9] By sharing data across platforms, stakeholders can collaboratively develop strategies that simultaneously address complex issues affecting water, agriculture, health, and many other domains.

On a personal level, think about the number of smartphone applications that could help you decide what fruit to buy: if you have 50 different apps, does that make your buying decision harder or easier? How long would it take you to review different information about fruits from each individual app? What if you only had one or two apps that could help you sort through the thousands of options from many different fruit farmers? With fewer apps, you could more easily combine the information to generate specific information you need (rather than get tired and give up)! Farmers, business owners, and policymakers are no different, and this user-friendliness is what

6 Steele, L; Orrell, T (2017). *Defining and approaching interoperability.*

7 Investing in Interoperability : *Insights from a Landmark Accenture Study.* (n.d.). Workday. Retrieved October 23, 2024, from https://beyonderpdrivingthedigitalbusiness.cio.com/wp-content/uploads/sites/44/2024/06/investing-in-interoperability-guide-enus.pdf

8 Garcia, C. (2023). *The crucial role of interoperability in driving value-based care transformation.* https://www.athenahealth.com/knowledge-hub/value-based-care/interoperability-helps-identify-care-gaps

9 HIMSS. (n.d.). *Interoperability in Healthcare.* https://www.himss.org/resources/interoperability-healthcare

drives the need for, and value of, interoperable systems and data.

When systems are interoperable, individuals, communities, governments, development partners, and the private sector can have greater confidence in the impact of their investments. Integrated ecosystems, rather than silos, can build upon each other to better address complex needs; they can also allow stakeholders to share costs and insights and lower the barriers to innovation.

In fragmented information environments with different public and private actors, how can data interoperability be achieved for improved development outcomes? **A clear mandate, as well as transparent ownership, resource management, governance, and policy frameworks, are required to establish collaborative mechanisms for interoperability in digital transformation and digital public infrastructure.** Replacing entire systems can be expensive and time-consuming. By using interoperable frameworks to support and integrate legacy systems, organizations can leverage their existing investments to eliminate the need for extensive new expenditures on software, hardware, and training.

In many development contexts, interoperability - as a vehicle for digital transformation - presents significant governance and technical challenges and gaps that need to be understood, contextualized, and worked through before standards and guidance can be developed and applied. In such contexts, it can be helpful to categorize interoperability issues into four layers:

1. Institutional and organizational,
2. Human,
3. Data and format, and
4. Technology.[10]

Applying such categorizations can help identify and organize complex challenges into more manageable components. Tools such as the Joined-Up Data Assessment, developed primarily to support the interoperability needs of public statistical authorities, are designed to help categorized.[11]

---

10 ISRIC. (2016). *Africa Soil Information Service (AfSIS).* Www.isric.org. https://www.isric.org/projects/africa-soil-information-service-afsis

11 Morales, L. G., & Orrell, T. (2017). Data interoperability: *A practitioner's guide to joining up data in the development sector* (p. 10). https://www.data4sdgs.org/sites/default/files/services_files/Interoperability%20-%20A%20practitioner%E2%80%99s%20guide%20to%20joining-up%20data%20in%20the%20development%20sector.pdf

Table 1: Breaking down the interoperability "buckets" - what categories include[12]

This paper will break down the building blocks of interoperability using a detailed case study of Development Gateway: An IREX Venture (DG)'s work to support the transformation of Ethiopia's livestock data ecosystem under the aLIVE program.[13]

12 Orrell, T., & González, L. G. (2021). *Introducing the joined-up data maturity assessment.* https://www.data4sdgs.org/sites/default/files/file_uploads/Joined_Up_Data_Maturity_Assessment_draft5.pdf

13 Development Gateway. (2022, December 2). Announcing Development Gateway's aLIVE Program: Advancing Livestock Data Ethiopia. *Development Gateway.* https://developmentgateway.org/blog/announcing-development-gateways-alive-program-advancing-livestock-data-in-ethiopia/

Why This Paper?

Interoperability is the secret weapon that has never been more needed to tackle grand-scale challenges. With growing levels of hunger across the world, governments are increasingly under pressure to marshall resources effectively and respond to citizen needs carefully, quickly, and holistically.[14] In addition, growing interest in the use of artificial intelligence (AI) to address sustainable development means that the demand for large volumes of complex, quality data is more, not less,[15] than it has been in recent memory. However, AI applications cannot function without data; similarly, many analog-adjacent challenges spanning health, agriculture, sanitation, and other areas cannot be resolved with a lack of data.[16]

While a significant body of literature on interoperability – both as a concept and a data governance function – exists,[17] very little has been written by digital development practitioners seeking to leverage it as a vehicle for digital transformation. This is especially true when it comes to implementing interoperability solutions in partnership with public administrations replete with the budgetary, staffing, logistical, and other capacity challenges that define such contexts.

We may not call out "interoperability failures" specifically, but we know them when we see them: projects that build competing data systems to capture the same or similar data; projects that are announced to join up systems into one mega dashboard, and quietly disappear from view after dashboards are built, but little to no data feeds them for continued use; projects that focus on just one data standard, but without space built in for stakeholder validation; initiatives to build onsite server rooms that are silent on building the existing and emerging human capacity needed to maintain servers after they are built; other initiatives that produce cool visualizations, but not much else. A single project or digital transformation investment approach rarely spans all the examples above. So how do we move from failure to improvement and eventually success?

---

14 Arzoyan, L. (n.d.). *Governments Need a Citizen-Centered Digital Transformation*. https://www.gmfus.org/news/governments-need-citizen-centered-digital-transformation

15 Moses, B. (2024). *Survey Says: Data Quality Management Isn't Evolving Fast Enough for A.* https://barrmoses.medium.com/survey-says-data-quality-management-isnt-evolving-fast-enough-e51dc4de5315

16 Snaith, B. (2023). What do we mean by *"without data, there is no AI"?* https://theodi.org/news-and-events/blog/what-do-we-mean-by-without-data-there-is-no-ai/

17 Data Management Association. (2017). *DAMA-DMBOK: Data Management Body of Knowledge*

This paper will use DG's 20+ years of experience in government digital transformation, specifically its ongoing work in Ethiopia's livestock sector, to demystify the key components needed to build robust, resilient, and interoperable data systems. Standardization, fair and consistent governance, and technology, coupled with human capacity and participation, lie at the heart of each building block.

Chapter 2

Birds of a Feather

Standards Bring Data Together

The Need

The specific analytics and visualization needs of end users can be met by highly interoperable databases and information systems that support data integration. However, data must first be standardized if different systems are to consistently, and seamlessly, exchange data. Without standardization, system inter-communication and the exchange of data between systems would be fraught with challenges.

Given this context, the aLIVE[1] team laid the groundwork for system interoperability and data standardization by ensuring that the five databases in Ethiopia's livestock ecosystem (ET-LITS, ADGG, DOVAR2, ADNIS, and NLMIS),[2] as well as other existing and planned databases used by the Ethiopian Ministry of Agriculture (MoA), could be integrated at the outset. Ongoing work is happening to make the databases interoperable. All relevant data will be incorporated into a central Livestock Information System (LIS). The creation of strong data standard policies and guidelines will ensure a supportive environment for sustainable data flows exists along different government channels. Data standardization must be relevant and appropriate for a range of current and future systems.

The Process

The data standardization process described in this document is modeled after international best practices. Relevant stakeholders3 were engaged in the process, with the MoA playing a key role in planning and facilitating discussions.

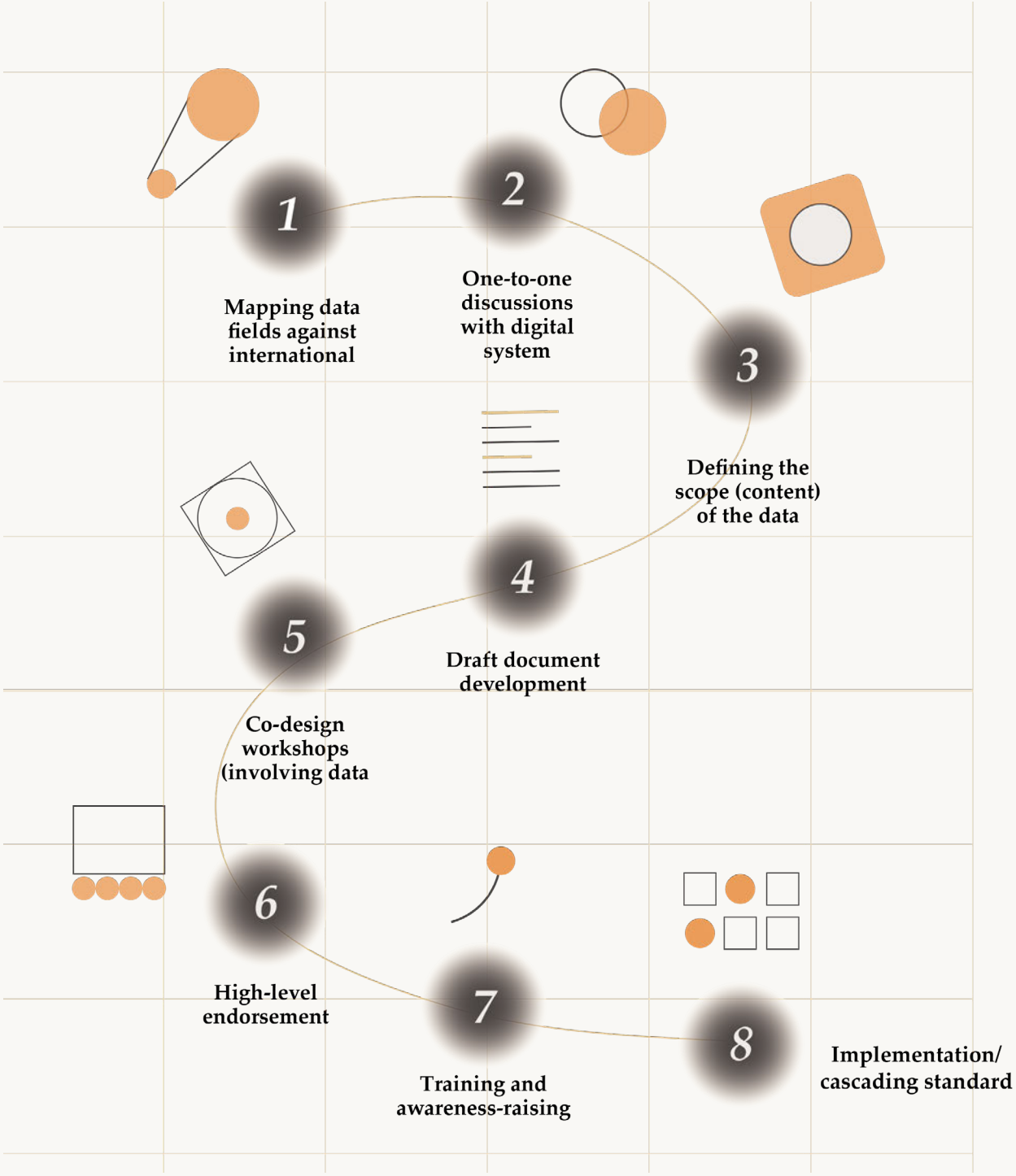
---

1 Development Gateway, with funding from The Bill & Melinda Gates Foundation (BMGF) and in partnership with Ethiopia's Ministry of Agriculture (MoA), is implementing a program known as a Livestock Information Vision for Ethiopia (aLIVE). This four-year, $5 million program aims to empower Ethiopia's stakeholders in the livestock sector to make data-informed decisions by providing relevant, accurate, timely, and digital livestock data and analytics. Learn more here: https://developmentgateway.org/blog/announcing-development-gateways-alive-program-advancing-livestock-data-in-ethiopia/

2 The 5 prioritized systems are as follows:
ADGG (African Dairy Genetic Gains): Data on genetic gains, collected by MoA in partnership with ILRI through an existing BMGF investment. LITS (Livestock Identification and Traceability System): Data on livestock population and registration, collected by MoA through an existing platform. DOVAR2 (Disease Outbreak & Vaccination Reporting System 2): Data on disease outbreaks and vaccinations, collected by MoA through an existing platform. ADNIS (Animal Disease Notification & Investigation System): Data on disease outbreaks, collected by MoA through an existing platform (complementary to DOVAR2) LMIS (Livestock Market Information System): Data on livestock markets, collected by MoA through an existing platform.

3 Development Gateway. (2024). aLIVE Program Reaches Milestone: Livestock Data Standards Endorsed by Ethiopia's Ministry of Agriculture. https://developmentgateway.org/blog/alive-program-reaches-milestone-livestock-data-standards-endorsed-by-ethiopias-ministry-of-agriculture/

The standardization process involved multiple iterative steps that involved various stakeholders with varying expertise and levels of involvement, including stakeholders with livestock subject matter knowledge, software development/programming expertise, and data management, analysis, and usage capabilities. The data standards development process had the following eight steps:



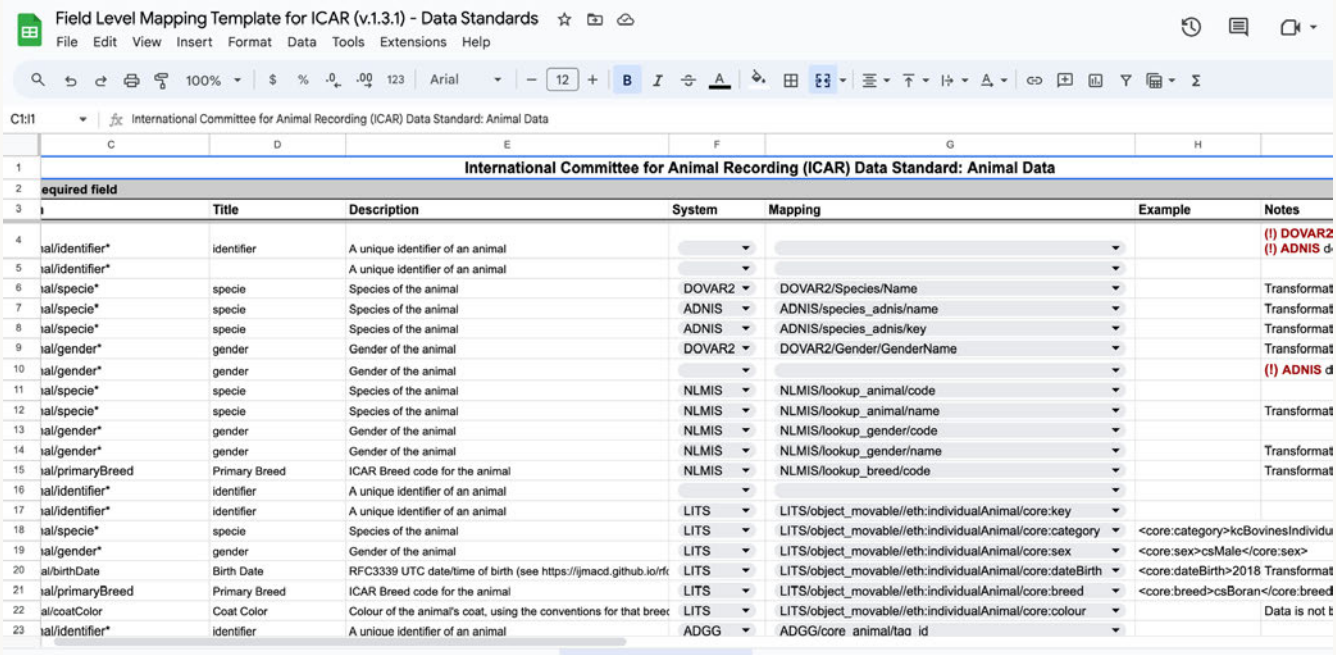
Eight steps can feel like a lot, especially when moving from differing terminology to a shared standard! It is worth noting that several steps may run at the same time (as they did in this case).

1. Field-level mapping

Field-level mapping was the initial step in developing the data standard, where the data fields in the prioritized systems were compared and mapped with the International Committee for Animal Recording (ICAR) Animal Data Exchange (ADE) standard. The ICAR-ADE standard defines common data concepts and structures that support various aspects of livestock management, such as animal identification, health, reproduction, and recording of events. To give some perspective, the ICAR standard has 134 required fields in the data schema and 23 enumerated fields/codelists. The prioritized systems had the following number of fields:

- ADGG - 373 fields
- ADNIS - 170 fields
- DOVAR - 158 fields
- ETLITS - 1,041 fields
- NLMIS - 270 fields

Following adoption of the ICAR-ADE standard, Ethiopian livestock datasets were aligned with those used in other countries, ensuring cross-country compatibility to facilitate trade and cross-border traceability of animals and products. This process helped identify similarities and differences between each system's data fields and the ICAR standard. The field-level mapping served as the basis for understanding how the data standard would affect each system within the interoperability plan.



Field Level Mapping Template for ICAR (v.1.3.1) - Data Standards

File Edit View Insert Format Data Tools Extensions Help

C1:1 | International Committee for Animal Recording (ICAR) Data Standard: Animal Data

	C	D	E	F	G	H
1	International Committee for Animal Recording (ICAR) Data Standard: Animal Data					
2	equired field					
3		Title	Description	System	Mapping	Example	Notes
4	al/identifier*	identifier	A unique identifier of an animal				(!) DOVAR2 (!) ADNIS d
5	al/identifier*		A unique identifier of an animal				
6	al/specie*	specie	Species of the animal	DOVAR2	DOVAR2/Species/Name		Transformat
7	al/specie*	specie	Species of the animal	ADNIS	ADNIS/species_adnis/name		Transformat
8	al/specie*	specie	Species of the animal	ADNIS	ADNIS/species_adnis/key		Transformat
9	al/gender*	gender	Gender of the animal	DOVAR2	DOVAR2/Gender/GenderName		Transformat
10	al/gender*	gender	Gender of the animal				(!) ADNIS d
11	al/specie*	specie	Species of the animal	NLMIS	NLMIS/lookup_animal/code		
12	al/specie*	specie	Species of the animal	NLMIS	NLMIS/lookup_animal/name		Transformat
13	al/gender*	gender	Gender of the animal	NLMIS	NLMIS/lookup_gender/code		
14	al/gender*	gender	Gender of the animal	NLMIS	NLMIS/lookup_gender/name		Transformat
15	al/primaryBreed	Primary Breed	ICAR Breed code for the animal	NLMIS	NLMIS/lookup_breed/code		Transformat
16	al/identifier*	identifier	A unique identifier of an animal				
17	al/identifier*	identifier	A unique identifier of an animal	LITS	LITS/object_movable//eth:individualAnimal/core:key		
18	al/specie*	specie	Species of the animal	LITS	LITS/object_movable//eth:individualAnimal/core:category	<core:category>kcBovinesIndividu	
19	al/gender*	gender	Gender of the animal	LITS	LITS/object_movable//eth:individualAnimal/core:sex	<core:sex>csMale</core:sex>	
20	al/birthDate	Birth Date	RFC3339 UTC date/time of birth (see https://ijmacd.github.io/rf	LITS	LITS/object_movable//eth:individualAnimal/core:dateBirth	<core:dateBirth>2018	Transformat
21	al/primaryBreed	Primary Breed	ICAR Breed code for the animal	LITS	LITS/object_movable//eth:individualAnimal/core:breed	<core:breed>csBoran</core:breed	
22	al/coatColor	Coat Color	Colour of the animal's coat, using the conventions for that breed	LITS	LITS/object_movable//eth:individualAnimal/core:colour		Data is not b
23	al/identifier*	identifier	A unique identifier of an animal	ADGG	ADGG/core_animal/tag_id		

2. Discussions with system owners

Following the field-level mapping stage, the aLIVe team held discussions with all five system owners to address the differences between the ICAR data standard and each system (including other prioritized systems). The team used a mix of one-to-one discussions (for the five system owners) and small group sessions (for the two health systems managed by a single directorate) to review the various points unveiled by the initial field-mapping process. These discussions aimed to establish a consensus on the necessary changes required for achieving interoperability at an individual system level (ahead of other larger group discussions where consensus across systems would be required). The team highlighted that some changes would be easier to implement than others. For instance, one significant change discussed was the transition from group to individual animal data collection in certain systems, which would facilitate analysis and interoperability. In practice, this transition will require changes to data collection forms or data entry tools and the structure of databases to accommodate individual animal data collection – not an insignificant task. Legacy data collected at the group level will be kept in its original format, while new tools and forms will be designed to capture/enter data at the individual level. However, this transition will require owners to collectively agree that it would be worth the effort to change their processes to make individual-level data analysis a reality.

3. Defining the scope

After multiple rounds of discussions with system owners (through ad-hoc meetings and two Technical Implementation Committee[4] [TIC] workshops) and several internal debates at the project level, the scope of the data standard document under development was defined. Consensus was reached on which animal species would be initially covered (the systems to include had already been identified during the inception of the project). As a result, the data standard includes four species - cattle, goats, sheep, and camels - across four main sections: animal registration, events, health-related data, and additional information (such as animal grading, market information, and location). The four species were prioritized because at least four existing systems already recorded data about their economic lifespans. The species covered play a significant role in Ethiopia's economy. Cattle are especially useful for draft power, milk, meat, and dung for fuel. Together, these species contribute to around 40% of Ethiopia's agricultural gross domestic product (GDP), nearly 20% of its total GDP, and 20% of its national foreign exchange earnings.[5] In contrast, only one or two systems separately collect data about other species. This inclusion ensures that the data standard covers the most relevant and widely recorded species. Additionally, market information is included to provide data on the entire livestock value chain, from birth up to market and trade, complementing the other sections on animal registration, events, and health.

4 See Annex 2 to learn more about the aLIVE Program governing structure.

5 Feed the Future Innovation Lab for Livestock Systems. (2021). Ethiopia's Livestock Systems: Overview and Areas of Inquiry. The U.S. Agency for International Development. https://livestocklab.ifas.ufl.edu/media/livestocklabifasufledu/pdf-/LSIL_Livestock_Systems_Overview_Ethiopia_2021_08.pdf

4. Development of draft data standard document

The aLIVE team developed a draft data standard document based on inputs gathered from discussions with system owners, two rounds of TIC meetings, and the field-level mapping document. The document focused on four core topics: animal data recording, animal events recording, animal health-related data, and additional attributes. Its primary purpose was to establish initial standards to guide discussions planned for the next major workshop in May 2023.

The topic that was the hardest to obtain consensus on was the transition from group data collection to individual animal data collection. There was initial resistance because this change not only impacts the technical aspects of the systems but also affects the overall data collection approach (as it requires enumerators to start collecting data at an individual animal level). In addition, several key field-level standardizations were done. For example, the age group categories across species were standardized and the naming and codes of breed lists for indigenous, exotic, and crossbreeds (per species) were harmonized, including the methods of measuring breed fractions. Geographic divisions were also aligned with official administrative boundaries. Standardized approaches to animal grades and farm types - based on production system, investment size, and farm sizes (small, medium, large) - were similarly implemented. However, unifying the fragmented systems was additionally complex as data had to be standardized across and within the different categories.

5. Workshops

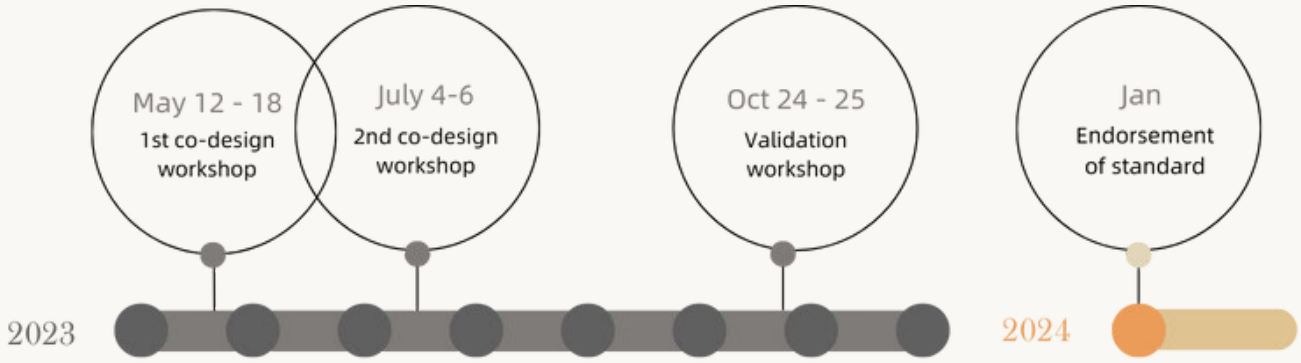
Three major co-design workshops were conducted to develop, review, and validate the data standard document. These workshops involved over 20 participants from the prioritized systems, the MoA's ICT/data team, and other stakeholder groups, including the International Center for Tropical Agriculture (CIAT). The first data standard co-design workshop, held from May 8th to 12th 2023, thoroughly reviewed the draft standard document and provided feedback across the four sections presented. Thereafter, the aLIVE team refined the document based on feedback ahead of the second workshop held from July 4th to 6th, 2023. Further feedback from this workshop helped with the finalization of the document. During another workshop held in October 2023, the document underwent an additional review and validation by the data standard taskforce.

6. Endorsement

The data standard document was finalized and endorsed by the governing committee of the project, which included the senior management of the MoA, on Jan 25, 2024. This marked a significant milestone in the project, as the governing committee's endorsement signified official approval and recognition of the document's importance and adherence to the necessary standards.[6] The endorsement was crucial because it confirms that the document is ready to be used as a guiding principle for all livestock data systems currently

6 For more information about the endorsement process please see https://developmentgateway.org/blog/alive-program-reaches-milestone-livestock-data-standards-endorsed-by-ethiopias-ministry-of-agriculture/

in operation and future ministerial systems at the federal and regional levels. Following the endorsement, electronic and physical copies were distributed to key stakeholders at the national and regional levels to increase awareness and ensure implementation.



7. Training and data standard awareness:

Following the distribution of hard and soft copies of the document, the next step was to organize a workshop to create awareness and introduce stakeholders to the data standard. The Program invited key federal and regional focal persons for priority systems as well as representatives from the regional bureaus where much of the data collection by enumerators takes place. This event introduced the document and secured buy-in for the cascading and use of the data standard. Additionally, it served as a foundation for the implementation and application of the standard, facilitating preparations for data collection according to the standard and informing resource allocation decisions for direct training support at the ground level.

8. Implementation:

After the finalization of the data standard document and its endorsement by the MoA, the next step was to ensure its proper implementation across the initial prioritized systems and the broader livestock sector ecosystem. From national and state ministers to officials in different departments, the MoA's active engagement in joining working sessions and providing detailed feedback was essential for the development and endorsement of the data standard. The MoA's continued, active ownership of the standards implementation process was critical.

In the future, any data systems introduced to capture data within the livestock sector must adhere to the data standard if they are to become part of a wider interoperable network. Accordingly, guidelines for implementing the data standard, carrying out capacity-building efforts, modifying database systems and data collection tools (including mobile applications and paper-based forms), cascading the data standard across regions, zones, and woredas (when applicable), and monitoring implementation progress were developed. These guidelines also mandated an analysis of potential changes and their impact. Additionally, a detailed plan

outlining timelines and responsibilities for implementation activities was established. A GitHub repository[7] of the data standard targeted at developers was also created to help with system-level modifications. Currently, implementation of the data standard has started within one of the five prioritized systems, namely ETLITS.

7 Ministry of Agriculture. (n.d.). Ethiopia-Livestock-data-standard. Github. Retrieved October 24, 2024, from https://github.com/moa-et/Ethiopia-Livestock-data-standard

The Outcome

Data standardization is a long[8] process needed to support interoperability between datasets and information systems. Everything cannot be standardized all at once, but knowing which datasets (and data systems) to prioritize provides a manageable scope for starting the process (and leaving room for scale over time). Moreover, people must be included as they know their data the best - indeed,much of the data standards process is documenting and formalizing the gaps and opportunities that data collectors and users instinctively know can help (or hurt) their ability to merge data. Inclusive data standards processes draw a diverse range of stakeholders together for a win-win scenario: everybody can make better decisions that drive productivity and health when they can see the same picture together.

Standardization occurs within systems, between systems, and at local, national, regional, and international levels. It is just as important to understand which fields need customization to fit local contexts (e.g., accounting for indigenous breed varieties) as it is to identify international standards that can serve as models for localized standards. Livestock know no borders, and neither do their data. Room for the "big picture" must be left when planning data standards as better outcomes for livestock, farmers, and policymakers can only be achieved when data across nations is amalgamated.

Ultimately, the development of data standards is an agile process that requires regular tweaks, updates, and adjustments to reflect how data and systems evolve. However, if momentum is to be maintained, system owners must take ownership of the process and financial resources must be committed to training and adapting staff, processes, and systems to merging needs. High-level commitments from sector leaders are also essential if any development effort is to be worth the effort.

8 It has taken the aLIVE program two years to date to conclude steps 1-7. Step 8, implementation, is ongoing and expected to continue until the end of the four-year program.

Chapter 3

Sharing is Caring

With or Without a National Legal Framework

The Need

Many countries do not have a clear "whole-of-government" policy on data sharing and access. In the absence of a national mandate that encourages or requires data sharing, data fragmentation and siloes occur, even in the United States.[1] In Ethiopia, for example, there is no national-level data-sharing policy that facilitates or compels data exchange across ministries and within ministerial directorates. According to an official of the Ministry of Innovation and Technology (MiNT), a draft policy was in the pipeline but the government instead decided to enact the Personal Data Proclamation 1321/2024. The Personal Data Proclamation incorporates many of the relevant concepts from the draft policy and includes additional data-sharing features as a directive instead of a policy, but falls short of setting cross-ministerial mandates for data exchange.[2] Why does this matter when looking at interoperability? Put simply, systems can be made interoperable, but they will struggle to remain usable without a consistent flow of quality data.

Tackling this common blocker to sustainability and data usage means incorporating a lens that looks at the practical as well as the legal frameworks for sharing data safely, routinely, and effectively (where data and systems are linked). Data governance refers to the process of managing and controlling data availability, usability, integrity/quality, and security within a system or systems.[3] It involves the development and implementation of policies, procedures, and/or guidelines to ensure that data is effectively and responsibly used.[4] Incorporating a data governance lens into the technical development of systems improves decision-making and increases data accuracy and consistency by taking the guesswork out of data sharing. It also enhances data security by ensuring compliance with legal and regulatory requirements. This reduces data-related risks and can promote data usage.[5]

Even where piecemeal national data policies exist, many laws do not address the day-to-day "how" of shifting from an individual to an enterprise-type data-sharing model. Cognizant of these complex challenges, the aLIVE

---

1 Diebold, G. (2023, September 25). Overcoming Barriers to Data Sharing in the United States. Data Innovation https://www2.datainnovation.org/2023-data-sharing-barriers.pdf

2 Policies are statements of the government's plans. They lie outside the hierarchy of Ethiopian laws because they do not have the same legal status as proclamations, regulations, and directives in terms of justiciable in a court of law. They are still related. Policies are statements of overall purpose that set out goals and provide principles that should be followed to achieve those goals. Policy goals and principles are made into laws through proclamations and regulations. Directives are at the lowest level in the Ethiopian legislation hierarchy. They describe how regulations should be implemented and are usually developed by a ministry or a department within a ministry. For more see Create, O. (2016). Study Session 15 National Policy Context in Ethiopia. https://www.open.edu/openlearncreate/mod/oucontent/view.php?id=79982&printable=1

3 Fothergill, B. T., Knight, W., Stahl, B. C., & Ulnicane, I. (2019). Responsible Data Governance of Neuroscience Big Data. Frontiers in Neuroinformatics, 13, 28.

4 Data governance key components: What to know in 2024. (2023, May 5). https://atlan.com/data-governance-key-components/

5 BrennaP. (2023, September 22). 7 benefits of data governance for your organization. DataGalaxy. https://www.datagalaxy.com/en/blog/7-benefits-of-data-governance-for-your-organization/

program[6] has built on work started in Ethiopia's health sector[7] to create a suite of data governance documents and policies for the Livestock Information System (LIS) under development, which is to be endorsed and implemented by the ministries and units that manage each data system.[8]

The Overall Approach

The five prioritized systems[9] that feed into the LIS were developed at different times and with different objectives. They all have specific rules defining data access and security. Each system has a unique method of managing data access aligned with its user needs and security requirements. Except NLMIS which openly publishes the vast majority of its data, each system controls how users access specific data fields, retrieve reports, download data, and upload data. Most systems, however, do publish aggregated data, or statistics derived from their data, for public consumption.[10]

Once built, the LIS will track data across the indicators identified through the Ethiopia Livestock System Roadmap.[11] As the roadmap details, the poor quality of data, coupled with fragmented database systems and limited analytical capacity, has limited the government's ability to plan, implement, and monitor livestock growth strategies. Given the expense, time, and complexity in sunsetting multiple systems and replacing them with one centralized system, the best approach focuses on interoperability; the key to this (eco) system is knowing that the prioritized systems will continue to exist separately and remain autonomous, with their own data management, permissions, protocols, and practices. However, it is essential to establish an overarching framework and set rules for how data access permissions are granted and maintained so that individual management does not prevent the gathering and linking of data through the LIS. An overarching

6 Announcing development gateway's aLIVE program: Advancing livestock data in Ethiopia. (2022, December 2). Development Gateway: An IREX Venture. https://developmentgateway.org/blog/announcing-development-gateways-alive-program-advancing-livestock-data-in-ethiopia/

7 Ethiopia Data Use Partnership. (2022, August 11). Ethiopia Data Use Partnership (DUP). https://ethiopiadup.jsi.com/

8 Data subjects (their role, needs, data sharing, and engagement) are an important piece of data governance. Data subjects are not covered in this section because the aLIVE program is several steps removed, working with the system owners who in turn work with data collectors (and in turn directly with data subjects). You can learn more about the role of individual data subjects in agriculture here; Farmer-centric data governance: Towards A new paradigm. (2023, January 30). Development Gateway: An IREX Venture. https://developmentgateway.org/publication_landing/farmer-centric-data-governance-towards-a-new-paradigm/

9 The 5 prioritized systems are as follows:
ADGG (African Dairy Genetic Gains): Data on genetic gains, collected by MoA in partnership with ILRI through an existing BMGF investment. LITS (Livestock Identification and Traceability System): Data on livestock population and registration, collected by MoA through an existing platform. DOVAR2 (Disease Outbreak & Vaccination Reporting System 2): Data on disease outbreaks and vaccinations, collected by MoA through an existing platform. ADNIS (Animal Disease Notification & Investigation System): Data on disease outbreaks, collected by MoA through an existing platform (complementary to DOVAR2) LMIS (Livestock Market Information System): Data on livestock markets, collected by MoA through an existing platform.

10 These insights come from an internal study conducted by the aLIVE program, which analyzed existing documentation and interviewed each system owner to understand their documented (and undocumented) arrangements.

11 Alemayehu, S. (2021). A livestock information system roadmap for Ethiopia. Alliance Bioversity & CIAT. https://alliancebioversityciat.org/publications-data/livestock-information-system-roadmap-ethiopia

framework also prepares LIS for future ministry-wide changes. For example, a new governing body that has been proposed will link up the management of many ministry-wide systems and be aligned with the digital agriculture roadmap of the ministry.[12]

The Process

The suite of data governance documents needed covers three pillars: strategic governance, regulatory compliance, and cross-cutting governance. These components are aligned with ongoing efforts to make the five prioritized systems interoperable. The details of what the pillars contain can be viewed at the end of the chapter in Annex 1.

Building off its experience with tackling data governance frameworks in complex settings,[13] DG has taken a 10-step approach to document development, with steps ranging from in-depth policy analysis and system analysis to extensive buy-in from key system owners and leadership.

To design this approach, the aLIVE program team worked (and continues to work) with the Technical Implementation Committee (TIC) and the Data Standards and Governance Task Force (DSGTF), two structures under the program's governance framework. You can learn more about the program's structure (institutionalizing how we maintain buy-in with a broad cross-section of stakeholders) in Annex 2.

Let's get to the ten steps!

12 At the time of publication, the DAR is under development and soon to be made public.

13 What We Know So Far: Best Practices in Developing Data Governance Frameworks. (2022, August 18). Development Gateway: An IREX Venture. https://developmentgateway.org/blog/what-we-know-so-far-best-practices-in-developing-data-governance-frameworks/

1 Project-level data governance policies operate under national and local legal contexts. It is relevant to ensure that such policies are aligned with such contexts. Thus, the aLIVE governance workstream leverages national data-sharing policies whenever relevant and possible. Accordingly, the initial step in setting priorities and developing the data governance documents was identifying and reviewing National Strategies, Policies, Plans, and Legislations relevant to aLIVE/LIS. This overview served as the basis for understanding the government's priorities for the livestock sector. Moreover, it has provided a comprehensive understanding of the legal environment set up to regulate digital and data-related initiatives.

2 following the official launch of the aLIVE project various governmental and non-governmental stakeholders were identified and contacted. These stakeholders included regulatory bodies and entities that administer complex information systems. The aLIVE team members had discussions with the Ministry of Innovation and Technology, the National ID Program (NIDP), the Ministry of Health- Health Information System (MoH-HIS), the Ethiopian Statistics Service (ESS), Information Network Security Agency (INSA), Agricultural Transformation Agency(ATA), and Precision for Development (PxD) - Farmstack, and CABI. These interactions have provided input for the aLIVE team to set priorities in the governance workstream and clarified the challenges and opportunities of data governance in the country.

3 Various documents were also gathered from the stakeholders mentioned above and others. The documents reviewed were gathered from MinT, INSA, ESS, NIDP, MoH-HIS, Ministry of Transport and Logistics, and Digital Agricultural Advisory Services (DAAS). Then, focusing more on the core systems, materials that served as training manuals, reports, requirement documents, guidelines, and strategic documents were gathered, reviewed, and analyzed. The documents from the five core systems were reviewed and analyzed. This provided an in-depth and comprehensive understanding of each system and its respective data governance, in particular.

4 following the stakeholder engagement and desk research, one-on-one discussions were held with each system owner to obtain up-to-date information. These discussions focused on the roles of different stakeholders in each system; their data collection, quality, and validation processes, and how they share and protect their data. These discussions also shed light on the major challenges they face and the opportunities available. These discussions were particularly necessary because many of the documents reviewed were outdated. It was pertinent to gather up-to-date information regarding the systems before building something new that will be based on these five systems.

5 this document was prepared as an output of the desk research and discussions held with the system owners. It provides a description of the existing frameworks provided to safeguard the quality, availability, and accessibility of data within the five core databases. It aims to avail a compiled/unified, summarized, and comprehensive document that describes the current data management arrangements deployed in the five core systems. Thus, it serves as a background document that identifies the data management and governance strengths as well as limitations of each system that need to be addressed to smoothen and facilitate the integration with the (LIS).

1. CABI collaborated with the MoA in developing the Soil and Agronomy Data Management Use and Sharing Directive No. 974/2023.
2. These documents include the following: MOH - Data Access and Sharing Guideline, National HIS Governance Framework, National HIS Governance, HIS strategic plan, NIPD Policy-and-Governance document, NIPD Data sharing protocol, MiNT Policy-and-Governance document, MinT EeGIF-Governance-and-Compliance - interoperability document, ESS Data Quality Assessment Framework, National Administrative Data Management Guideline, Ministry of Transport Access to Information Manual, Soil and Agronomic Data Sharing Directive, the Ethiopian AAS Ecosystem Data Actors and Standards Inventory

6 through multiple rounds of discussions with system owners and project-level internal discussions, it was decided that the scope of the data governance work would focus on LIS. The documents that will be prepared under the governance work (e.g. data sharing protocol, data governance framework, etc) will focus on the data that will be shared through LIS. This decision was reached because of the following four major reasons.

- There are national and ministry-level data governance-related policies. Some of these policies or legal documents are still under development.
- There are other livestock databases that are not currently included in the LIS. The governance documents need to address specific data governance issues as the five core databases will feed data to the LIS. For instance, the data sharing and access protocol need to set specific rules for data access permissions for various types of data and various types of users of the LIS.
- The five core databases will remain autonomous and they will continue to govern their respective databases and share their own data as they see fit. It would be beyond the program's reach to make changes to the structural and governance-related arrangements of these systems.

Nonetheless, the governance documents developed can serve as templates for the five core systems, the Ministry, and others as needed.

7 before delving into drafting the data access and sharing protocol for the LIS a Data access mapping/matrix of the five core systems was developed. This mapping was done to show the existing data-sharing practices of the systems. Based on these practices the Data access matrix for LIS was developed to show the desired data access in LIS. The LIS data access matrix was shared with the system owners. Based on their feedback, the team is drafting the data access protocol.

8 two major co-design workshops were conducted that provided significant input for the data access protocol. These co-design workshops were organized in order to harness a better understanding of the users' and system owners' needs. Such workshops also enhance engagement, buy-in, collaboration, and alignment of the stakeholders who have different interests. More innovative, effective, user-centered, and creative solutions emerge in the discussions.

8 The first one was the LIS Co-design Workshop was held from November 23-24, 2023. In this workshop, the types of data that will be included in the LIS were presented to the stakeholders. The second round of the LIS Indicators and Data Governance Workshop was held from February 6-7, 2024. In this workshop, the types of data (Indicators, pillars, and sub-pillars that will be displayed in the LIS Dashboard) were refined and on the second day, the LIS Data Access Mapping was presented.

9 The zero draft outline for the frameworks has been developed and the work will continue once the data sharing and access protocol, along with its annexing documents, are completed. The framework will be an overarching document that will serve as a guideline to ensure accountability, transparency, coordination, and partnerships in governing the LIS. As the development of the LIS progresses, the questions that need to be answered in the governance framework will manifest more clearly. So far, the data standard, the resource mobilization strategy, the architecture, and the data sharing policy have been developed by various work streams of the aLIVE program. Thus, preliminarily, it is anticipated that the governance framework will cover the following specific issues:

- Which body will be responsible for setting user tiers and permissions?
- What detailed procedure will be used by the responsible body for making tier and permission decisions?
- What data access procedure will be followed when integrating new systems into LIS?
- How frequently will permissions and tiers be reviewed and updated?
- Which body will be responsible for oversight to ensure the sustainability of LIS, implementation of new Data Standards, and governance guidelines?

10 The Technical Implementation Committee, which also includes all the system owners, will review and validate the data governance documents. Before they can be officially implemented, the Governing Committee will also review and endorse the documents. To ensure the implementation of the data governance documents, they need to be distributed to all the relevant stakeholders. Training may also be organized, as necessary.

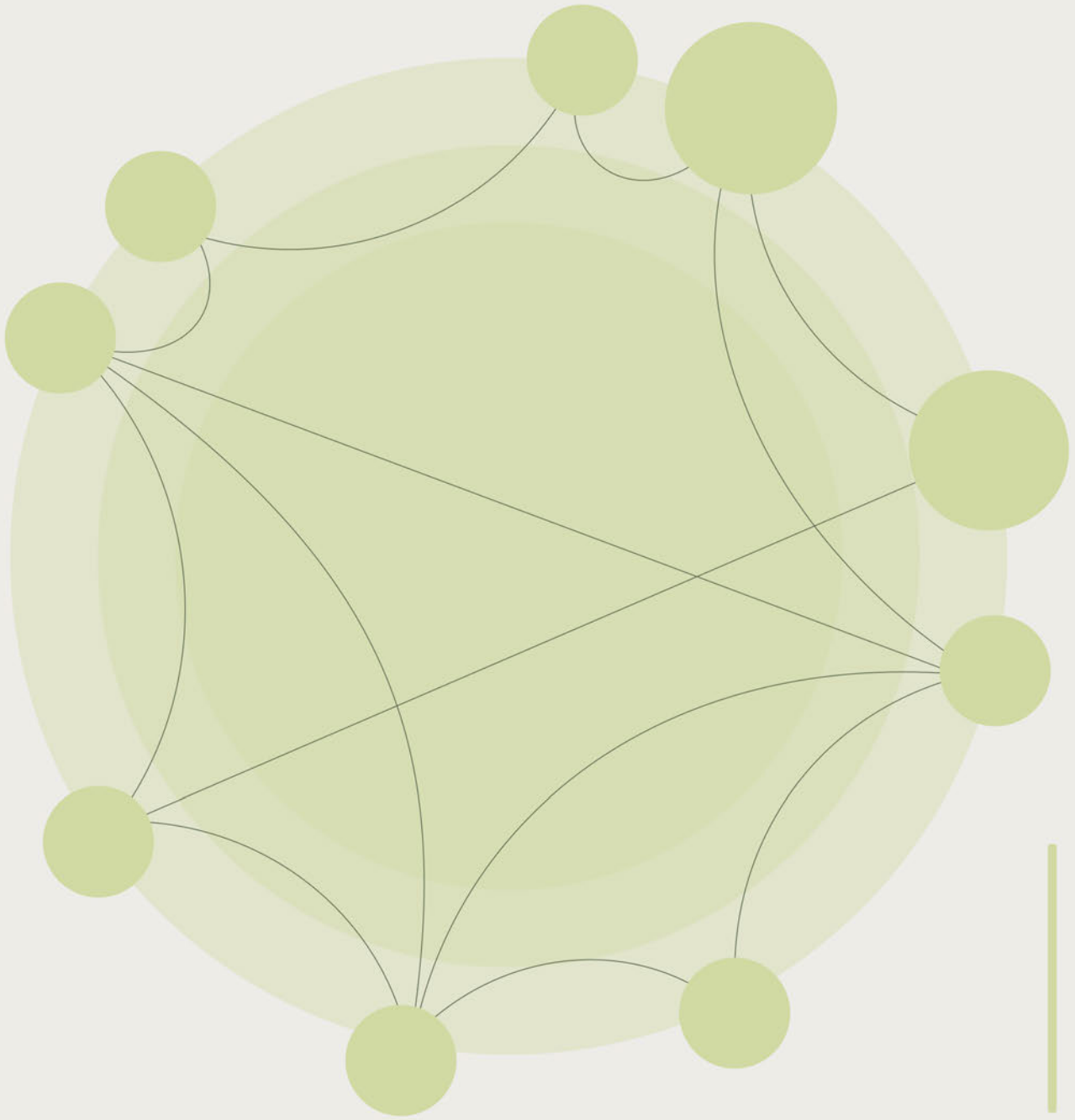
The Challenge

The biggest overarching challenge is the absence of a clear framework that mandates data sharing within and across government ministries. As a consequence, the five priority systems have limited data governance arrangements. Due to these gaps, defining the scope of the LIS data governance work required various discussions with the system owners, MoA leadership, and other concerned stakeholders (e.g., government agencies outside the ministry that give and receive data) to set a reasonable scope given the time-scale of the aLIVE program. Ultimately, the aLIVE team decided that the data governance scope should be focused on the LIS, effectively making the LIS platform's documents a template that any government system, including the priority systems, could adapt to their respective data governance protocols. This approach helps lay the foundation for a customizable, but uniform, level of data sharing across the MoA and heralds a shift from ad-hoc to consistent data access.

The Outcome

Making data available is not just a question of standardizing what exists, as in Chapter 1; it is also about governing how existing data is shared. This is especially important in policy environments where there is a gray area about how data should be consistently shared with internal departments/agencies and external partners/stakeholders. This chapter outlines how to move from evaluating the need for data governance to reviewing the existing policy framework, engaging with data owners to understand their data-sharing starting points, reviewing and discussing data owners' documentation, and defining the scope of sharing and access protocols to feed into a future shared framework.

After getting to know system owners and their current data governance structures, the aLIVE team acknowledged that data sharing and use do not occur in vacuums - approaches to both already exist, even though they may be imperfect. The team members realized that the building of a shared framework - based on the characteristics of each system - allows for national or multi-system governance approaches. Such approaches simultaneously harness what works from individual systems and build buy-in from stakeholders. However, the involvement of human actors will be essential, even if detailed laws can set the stage for data sharing. Data owners need to see, in real-time, the shared value of data exchange to commit to new processes and approaches. When we say "interoperability requires robust data governance," this layered process of people, policies, and framework development is what we mean.



Chapter *4*

The Breakdown

What Good Technology Infrastructure Can Look Like

The Need

What counts as "good" digital public infrastructure is notoriously hard to define. Given the absence of experts who can break down tech requirements and make them understandable, coupled with a non-specialist audience that is trying to quickly learn what software, hardware, and cloud infrastructure is needed to solve problems, it is no wonder the public sector struggles to develop sustainable, future-proof technology. Without expert advice, results can produce more problems than they solve: specifically, systems that cannot be modified to fit local needs without continuously re-engaging vendor support, technology that quickly becomes outdated, and difficult-to-manage storage solutions that can easily delete important data.

Smart development of digital infrastructure not only enables resource-constrained governments to operate more efficiently but also builds trust by improving transparency and responsiveness in citizen services. A dashboard going offline because of improper maintenance is not just a minor inconvenience: it can damage citizen or key stakeholder trust and discourage the use of data to support more informed decision-making.

If a government website you need is repeatedly down, would you continue to use it? This is the critical question that underlies thinking through and investing in better digital public infrastructure at scale.

The Objective

Several technical aspects must be considered to implement interoperability, with deployment of a sustainable, scalable, locally supported, future-proof solution being the overall objective. The integration of multiple systems (two at a minimum) is the critical factor to consider. These systems must remain autonomous and continue to function independently, even as they seamlessly work together. Each system has its quantitative properties, be it in the number of records or the amount of data processed. Centralized platforms need to ingest data from all of these systems and not fail in the process. **A fully interoperable system is a sum of all its individual system parts, and hardware, software, and sustainability/expansion choices must reflect this realization.**

Implementing interoperability is one of the most critical aspects of creating an efficient and sustainable technological environment. This section delves into the essential technical aspects of implementation, from **choosing hardware** and its related **software platforms** to detailing the pros and cons of local hosting versus cloud hosting. While each approach has advantages and disadvantages, it is important to consider the context: what are the existing systems that need to be brought along with a vision for the future? What procurements,

plans, or budget restrictions exist that might impact how hardware and software decisions are made? There is no "right" or "wrong" answer. The goal of this section is to lay out common options and spell out the key factors that should be considered when making a decision.

Choosing Hardware Infrastructure

In modern information systems, hardware infrastructure decisions play a key role in determining the efficiency, scalability, and security of any government's or organization's data operations.1 Whether opting for on-site hosting or cloud-based solutions, each choice has unique advantages and challenges. Understanding the nuances of these options is crucial for making informed decisions that align with specific needs, resources, and long-term goals. This section delves into the critical considerations that come into play when choosing the appropriate hardware infrastructure, examining factors such as sustainability, local expertise, access to foreign currency, control and customization, latency, scalability, security, and compliance. Governments and organizations should develop a robust infrastructure strategy that supports their operational objectives and enhances their overall data management capabilities.

1 Sirkemaa, S. (2018). Information Systems Infrastructure – Importance of Robustness. *Information Technology Science, 241–244.*

What About Cloud or Local? That is the IT Question

Pros of Self-Hosting

Suppose the interoperable solution needs to operate within national borders due to laws or regulations on data localization and sovereignty. In that case, on-site hosting[2] almost always offers superior bandwidth[3][4] as it leverages local internet nodes and traffic routes and resides close to data production nodes within the country. In contrast, cloud solutions – despite being highly scalable and flexible – may not always offer the same bandwidth efficiency as interoperable solutions need to operate within national borders. Cloud services typically rely on geographically dispersed data centers,[5] potentially leading to increased latency and slower data transfer rates compared to on-site hosting. This dispersion can affect the performance and responsiveness of applications that require real-time data processing and high bandwidth.

Local hosting contributes to data sovereignty by giving governments complete control over the associated networks and servers, where financial resources and technical personnel are available. Data is subject to local regulations when it is stored on servers within a country's borders and managed directly by the government using available enforcement mechanisms. This control reduces, among other things, the risk of unauthorized access and the reliance on third-party providers. It also supports the enforcement of strict data protection measures. By keeping data within a controlled, physical environment, governments can better protect sensitive information, ensure compliance with national regulations, and prevent potential data sovereignty breaches.[6]

Cons of Self-Hosting

The main disadvantage with self-hosting is the need to have a team to manage hardware and software infrastructure.[7] This team handles all management aspects, including keeping hardware updated at service locations and within the network where servers are connected. The team is also responsible for controlling

2 Wikipedia contributors. (2024, July 24). *On-premises software*. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=On-premises_software&oldid=1236371187

3 Shim, T. (2024, April 5). *What is latency: Impact of server locations to your site performance. HostScore.* https://hostscore.net/learn/what-is-latency-impact-of-server-locations-to-your-website-performance/

4 Wikipedia contributors. (2024, August 17). Latency (engineering). Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Latency_(engineering)&oldid=1240837898

5 Wikipedia contributors. (2024, October 27). *Cloud computing*. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Cloud_computing&oldid=1253652807

6 Chander, A., & Sun, H. (2023). Data Sovereignty: *From the Digital Silk Road to the Return of the State*. Oxford University Press.

7 Krombholz, K. (n.d.). To cloud or not to cloud: A qualitative study on self-Hosters' motivation, operation, and security mindset. Retrieved October 28, 2024, from https://www.usenix.org/system/files/usenixsecurity23-grober.pdf

physical server access, maintaining strict licensing control, and implementing thorough recovery plans. Additionally, team members schedule external backups, implement redundancy measures, monitor the temperature and humidity of the server room, and ensure redundancies for both the network and power supply. In general, most governments and large organizations should already have this team in place. As such, it is important to leverage the available options in the deployment context.

Pros of Cloud Computing

Generally, cloud services excel in scalability terms, enabling quick adjustments in response to resource demands, while on-site hosting is limited by physical constraints and slower expansion capabilities. Compliance with regulatory requirements, such as mandates on where government data needs to be stored, is another key benefit. The choice depends on the needs and resources available, balancing control, cost, security, scalability, and compliance considerations.

Cons of Cloud Computing

Cloud services providers are interested in selling access. The older a machine gets, the higher the cost; as such, providers try to keep all their hardware infrastructure as new as possible. Thus, if an organization rents a server for some years (which is roughly what cloud computing is), at some point, its running cost will be much higher than it was initially because it will be considered "legacy." So, even if an application runs perfectly fine and does not need any particular hardware upgrades, organizations are forced to upgrade because of the cost of using the old machine, which may eventually become unsupported.[8]

Although most offerings for cloud solutions leverage the advantages of automatically maintained third-party services (such as AWS S3, AWS Cognito, AWS Glue), paying for essential services priced in foreign currency presents significant challenges that affect the sustainability of the infrastructure and operations. Volatile exchange rates can lead to unpredictable costs, making budgeting and financial planning very difficult. Unless billing and the strategy for scaling up is strictly controlled, scalability can become a liability since the costs may increase faster than they can be detected, optimized, and controlled. Limited access to foreign currency reserves can exacerbate these issues, forcing governments to prioritize spending and potentially compromising the continuity of service delivery. Finally, dealing with these technical and financial complexities undermines any platform's sustainability, which is particularly important for essential government services.

8 Mann, T. (2022, May 4). Concerned about cloud costs? Have you tried using newer virtual machines? The Register. https://www.theregister.com/2022/05/04/aws_amd_intel/

Free or Low-Cost Tiers Have Downsides as Well

Most cloud providers at the moment offer low-cost or free tiers of their Infrastructure-as-a-Service (IaaS)[9] offerings. However, there are several reasons why adoption of such tiers is a bad idea. For example, there are:

- Security and compliance risks, as governments often need to follow strict regulations and data protection policies that free trial environments may not support;
- Data privacy concerns, as data under a free tier is usually is subject to different terms and conditions which may not align with government regulations;
- Limited features;
- Possible hidden costs, as most IaaS offerings let users go over the prescribed thresholds or activate services10 not covered by the free tier, leading to unexpected expenditures; 11
- Vendor lock-in concerns, as any advanced solution developed on proprietary IaaS platforms would be difficult to migrate to alternative hardware, largely due to differences in the technical implementation of some services (e.g., Google's Cloud Functions12 versus AWS' Lambda13);
- Support issues, as free tiers usually come with lower service-level agreements as well as limited to no technical support. Free tiers are sometimes used as marketing tools to lure clients into a hosting ecosystem.

Below is a list of the factors that should be considered when deciding between on-site hosting and cloud computing:

9 Wikipedia contributors. (2024, October 12). Infrastructure as a service. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Infrastructure_as_a_service&oldid=1250725639

10 AMAZON. (n.d.). FAQ sur l'offre gratuite d'AWS. Aws Amazon. Retrieved October 28, 2024, from https://aws.amazon.com/free/free-tier-faqs/

11 Makhlouf, R. (2020). Cloudy transaction costs: a dive into cloud computing economics. Journal of Cloud Computing Advances Systems and Applications, 9(1). https://doi.org/10.1186/s13677-019-0149-4

12 Cloud run functions. (n.d.). Google Cloud. Retrieved October 28, 2024, from https://cloud.google.com/functions

13 AMAZON. (n.d.). AWS Lambda. AMAZON. Retrieved October 28, 2024, from https://aws.amazon.com/lambda/

ASPECT	ON-SITE HOSTING	CLOUD SERVICES
 Scalability	Limited by physical constraints and slower expansion capabilities.	Highly scalable, enabling rapid adjustments to resource demands.
 Flexibility and Management	Requires significant investment in local capacity building and infrastructure management.	Offers flexibility and ease of management, with rapid resource allocation and scaling capabilities.
 Performance and Responsiveness	Enhanced performance and responsiveness for real-time data processing applications.	Performance may vary based on the location of data centers and the specific requirements of applications.
 Data Sovereignty Compliance	Easier compliance with national data sovereignty regulations as data stays within national borders.	Complexities in complying with data sovereignty regulations as data may be stored and processed in multiple geographical locations.
 Latency	Reduced latency by being close to the source of data production.	Potentially higher latency due to data centers being located in various geographical locations.
 Bandwidth Efficiency	Superior bandwidth leveraging local internet nodes and traffic routes close to data production nodes within the country.	May have increased latency and slower data transfer rates due to geographically dispersed data centers.
 Foreign Currency Considerations	Not directly affected by foreign currency fluctuations; services denominated in local currency are more likely to be sustained in time.	Prices in foreign currency can lead to unpredictable costs due to exchange rate volatility.
 Vendor Lock-in	Using open-source platforms allows to migrate to new infrastructure eventually (both on-site and cloud).	Cloud platforms have their own proprietary SaaS, which makes it difficult to migrate to any other provider or on-site hosting.

Choosing Software Platforms

Similar to the hardware implementation issues mentioned above, choosing a sustainable software ecosystem for any interoperability initiative is very important. Before any interoperable system is deployed, IT leaders must ensure that the underlying software ecosystem supports best practices and covers all aspects of the software development life cycle. A strong, well-supported software ecosystem with clear processes that not only enable the evolution of interoperable platforms but also support the inclusion of future systems must be established.

Open Versus Proprietary Software

Although open-source software may be a cost-effective option for governments looking to minimize capital and operating expenses, significant costs are associated with implementation, customization, and ongoing maintenance. In short, while the Principles for Digital Development[14] encourage the use of open-source software, this type of software is not free.[15] It can, however, be managed internally or externally through community support, particularly by leveraging local vendors or professionals. The advantage of open-source software is the abundance of community-generated documentation (especially in well-supported projects). In contrast, proprietary software usually involves significant upfront licensing fees, recurring subscription costs, and potential additional expenses for updates, support, and feature enhancements. These costs can escalate, especially if the software or platform is key to operations and requires regular updates and support.

One example of a strong open-source community that provides strong options for interoperable architectures is the Apache Software Foundation[16] (ASF) has played a critical role in developing sustainable solutions. ASF is a key organization in the open-source community that plays a crucial role in developing and maintaining high-quality open-source software.[17] ASF is known for its governance model, which emphasizes community-driven development and ensures that contributions are valued based on quality rather than source[18]. The foundation maintains and incubates some of the most widely used and influential open-source projects, such as Apache HTTP Server, Apache Hadoop, Apache Spark, Apache Kafka, Apache Ozone, and Apache Iceberg. ASF's commitment to open standards and transparency promotes interoperability and trust, making it easier

14 Principles for Digital Development. (2014). Principles for digital development. Principles for Digital Development. https://digitalprinciples.org/

15 Warner, B. (2024, September 17). Mythbusters: Digital public infrastructure & digital public goods. Development Gateway: An IREX Venture. https://developmentgateway.org/blog/mythbusters-digital-public-infrastructure-digital-public-goods/

16 Committer, B. a. (n.d.). Welcome to the Apache software foundation! Retrieved October 28, 2024, from https://www.apache.org/

17 DG uses products developed by the Apache Software Foundation, but we are not affiliated with or endorsed by the Apache Software Foundation in any way

18 Committer, B. a. (n.d.). Briefing: The Apache way. Retrieved October 28, 2024, from https://www.apache.org/theapacheway/

for organizations to adopt and integrate their software. Additionally, ASF's strong emphasis on licensing and legal frameworks ensures that projects remain freely available and sustainable, protecting the interests of both developers and users.

Consider Vendor Lock-In - Flexibility in the Source Code

Vendor lock-in is one of the most important factors to consider when deciding which platform to use and a key difference between proprietary and open-source software. Open-source software typically minimizes vendor lock-in due to its transparent and accessible nature, and various stakeholders or vendors have the legal right to maintain and augment a system's code base. Governments have full access to the source code and, depending on the license type, this access can allow for the modification, distribution, and even "forking" the repository as needed (i.e., if there is a customization that only the government is interested in). This ensures that governments are not dependent on a single vendor for updates, support, or improvements – they can easily switch to another provider or bring the solution in-house if necessary. Under a proprietary license, governments rely heavily on the vendor for updates, support, and continued access to the software, making it challenging to migrate to alternative solutions.

Proprietary software vendors recognize the concerns about vendor lock-in.[19] They offer several options that need to be considered, such as a more permissive license agreement/contract, a strong application programming interface (API), support for standards, and support for data portability and migration.[20]

One major concern regarding the financial implications of proprietary software is the denomination of any support and license fees in foreign currency (unless the vendor of a certain platform has local partners). As mentioned previously when talking about cloud and on-site infrastructure, this denomination may impact sustainability. If a government cannot pay for services, it risks having a key system that is unsupported.

Governments and organizations may hire a local or international consultant or developer to create a bespoke software product that avoids any type of vendor lock-in. Such consultants can ensure the future sustainability of any solution with explicit licenses, open-source code and related documentation, and clear intellectual property ownership rules.

Below is a list of aspects to be considered when selecting a sustainable platform:

19 Cloudflare. (n.d.). What is vendor lock-in? | Vendor lock-in and cloud computing. Cloudflare. Retrieved October 28, 2024, from https://www.cloudflare.com/learning/cloud/what-is-vendor-lock-in/

20 Opara-Martins, J., Sahandi, R., & Tian, F. (2016). Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective. Journal of Cloud Computing, 5(1), 1–18.

ASPECT	OPEN SOURCE SOLUTIONS	PROPRIETARY SOLUTIONS
Cost	Typically free or low-cost with permissive licenses (e.g., Apache 2.0, BSD, MIT).	Often involves significant licensing fees and recurring costs.
Customization	High level of customization and flexibility (via forking/branching)	Limited customization, depending on vendor policies and available features.
Support	Strong community support with active development and extensive documentation.	Vendor-provided support, often requiring additional support contracts.
Vendor Lock-in	Minimal vendor lock-in due to open standards and code availability.	There is a high potential for vendor lock-in, making it difficult to switch to other solutions without significant effort and cost.
Security	Transparent security features with code visibility, allowing for independent audits.	Security features managed by the vendor, with limited visibility into the source code (depending on solution)
Innovation	Rapid innovation driven by community contributions and open collaboration.	Innovation dependent on vendor's development cycle and priorities.
Scalability	Highly scalable with the ability to modify and optimize for specific needs; this requires higher in-house expertise.	Scalability often limited to the vendor's predefined capabilities and upgrade paths, although many solutions can be prepared to scale by design.
Compliance	Easier to ensure compliance with specific regulatory requirements via source code modifications.	Compliance features are dependent on the vendor's implementation and may require additional customization with more costs.
Support and Maintenance	Requires in-house expertise for support and maintenance, with community assistance.	Vendor-provided support and maintenance, often at an additional cost denominated in foreign currency.
Updates and Patches	Updates and patches are community-driven, potentially leading to quicker releases, but with a risk that a project or a component may be abandoned.	The vendor controls updates and patches, with scheduled release cycles, while absorbing the costs of regression testing, deployment, etc.
License Restrictions	Few or no license restrictions, allowing for broad usage and redistribution.	Usage governed by strict license agreements, with restrictions on redistribution and modification.

Digital Public Goods

Digital Public Goods (DPGs) offer numerous advantages, particularly because they must adhere to the Digital Public Goods Standard[Digital Public Goods Standard[21] which mandates the use of open-source licenses. Unlike proprietary solutions, which can lock organizations into restrictive agreements and limit access to source code, DPGs are typically free or available at a low cost. They promote transparency, clear ownership, and compliance with privacy and legal standards. Additionally, DPGs align with 'do no harm,' data privacy, and security principles. In contrast, proprietary solutions have limited flexibility, a lack of transparency, and potential conflicts over data ownership and control.

APIs and Standards in Interoperability

APIs live at the core of most interoperability processes. They are the rules and protocols that allow platforms to communicate with each other. They play a key role in this respect, working as intermediaries between disparate systems and external or internal services. By leveraging standards such as RESTful[22],GraphQL,[23] or SOAP[24], APIs ensure flexibility, compatibility, and scalability among very different systems. They are used to overcome differences in architecture, platforms, languages, and locations.

For this reason, APIs must adhere to established security, authentication, and data exchange standards. For security, standards such as OAuth[25] are essential for enabling secure access to resources. For data exchange, JSON[26] or XML[27] (and their schemas) ensure that the information being exchanged can be interpreted by all parties involved, regardless of underlying technologies. There is no "one size fits all" approach to the implementation of APIs. The cost, upsides, pitfalls, and blockers to API implementation will depend on the landscape of the systems to be connected.

---

21 Digital Public Goods Standard. (2020, September 21). Digital Public Goods Alliance - Promoting Digital Public Goods to Create a More Equitable World; Digital Public Goods Alliance. https://digitalpublicgoods.net/standard/

22 Amazon, A. W. S. (n.d.). What is a RESTful API? AWS Amazon. Retrieved October 28, 2024, from https://aws.amazon.com/what-is/restful-api/?nc1=h_ls

23 GraphQL. (n.d.). GraphQL. GraphQL. Retrieved October 28, 2024, from https://graphql.org/

24 SOAP vs REST APIs: Which is right for you? (n.d.). Retrieved October 28, 2024, from https://www.soapui.org/learn/api/soap-vs-rest-api/

25 OAuth. OAuth 2.0. Retrieved October 28, 2024, from https://oauth.net/2/

26 w3schools. (n.d.). What is JSON. w3schools. Retrieved October 28, 2024, from https://www.w3schools.com/whatis/whatis_json.asp

27 Schnier, J. (2008). XML Introduction. In Flash XML Applications (pp. 3–9). Elsevier.

Supporting Existing Systems

Working With Legacy Systems

In most government contexts where interoperability is to be implemented, there are challenges related to the existing technology ecosystem. Certain requirements and good practices should be uniform since these systems are the source of raw or processed data that is exchanged or extracted. Naturally, these systems evolve; they contain valuable data, workflows, and functionalities that have been refined over time to meet specific business needs, and this impacts any further transformations needed to share information between systems. Many governments rely on legacy systems that may be outdated but still perform key functions. Replacing existing systems entirely can be expensive and time-consuming, so supporting and integrating them into a new interoperability framework allows organizations to leverage their existing investments and reduce the need for extensive new expenditures on software, hardware, and training. Existing systems may pose complex problems themselves due to their legacy code, age, various technologies used across the decades with distinct implementers, possible expired or non-existent technical support from the original vendors, and lack of documentation.

As a starting point for any digital roadmap or interoperability strategy, taking an appropriate census or taking stock of the existing technology systems is critical. Often, systems are built in different programming languages, using different components, different licensing terms, different hosting environments, and so on. A holistic understanding of the current state of a ministry's or organization's technology ecosystem is thus crucial for identifying the systems that should be decommissioned or made interoperable. Such an understanding is also essential for determining the technical and human resources needed to provide continual support and ensure the availability of legacy systems relied upon for interoperability (e.g., via data lakes or data exchange tools).

The Critical Importance of Code Repositories and Issue Trackers

A version control system is one of the key aspects of any software ecosystem. Code repositories record every change made to files, and each change should be accompanied by a narrative explanation or a support ticket number. Paired with issue trackers, code repositories not only capture the history of code changes but also record the context behind them, including the change requests and collaborative discussions. Over years of development and maintenance, this practice creates a 'written story' that documents the evolution of each file. This comprehensive history can explain how a software system has progressed from its initial state to its current form, detailing the refactoring, bug fixes, architectural decisions, and release dates that have shaped its development.

The following is a summary of the best practices and baselines that improve the sustainability of an ecosystem, although the actual implementation may need time and investments:

BEST PRACTICE	IMPLEMENTATION
Documentation and Knowledge Sharing	Create and maintain a knowledge base or wiki for documentation, guidelines, and best practices. Incentivize knowledge sharing through team meetings, workshops, and internal presentations
Version control and source management	Implement Version Control Systems like Git or SVN to track changes, manage code versions, and facilitate collaboration. Also, implement clear branching strategies (e.g., Git Flow, Git Forking Flow) to effectively manage features, fixes, and releases.
Coding standards and Code reviews	Establish and enforce code reviews, coding standards, and guidelines to ensure consistency and maintainability. Code reviews can be especially helpful in catching issues early.
Automated testing	Leverage automated testing frameworks integrated into the software release cycle to ensure that the different parts of your ecosystem function seamlessly.
Continuous Integration/ Deployments	Set up CI/CD pipelines using tools like Jenkins, GitLab CI, or GitHub Actions to automate build, test, and deployment processes. Automating the build process ensures consistency and reduces manual errors.
Collaboration	Leverage tools like Slack, Microsoft Teams, or Zoom for real-time communication and collaboration.
Apply agile	Adopt agile practices (e.g., Scrum, Kanban) to enhance team collaboration, flexibility, and project management.
Continuous Learning and Improvement	Invest in training programs to keep the team updated with the latest technologies and best practices while conducting project retrospectives to identify improvement areas.

Building for the Future

Solving the aforementioned sustainability challenges requires a holistic approach addressing software, hardware, and local IT infrastructure and capacity. Leveraging local expertise, both private and public, is crucial to ensure that a system is well-supported and maintained.

Software tools must be highly maintained, tested, and proven in production environments. They should have a history of stable releases and a reliable codebase to ensure key system functionalities stay intact after changes are made. Ideally, these tools need to be based on open standards and open source code, with permissive licenses such as Apache 2.0, BSD, or MIT. A vibrant community is also essential, as this ensures that operating costs remain low, the IT specialists' expertise remains high, and the organization is not locked into proprietary, closed architecture.

An Example of the Future: Increased Livestock Data in Ethiopia

The current platform DG is building in Ethiopia is focused on existing cattle records (on weight, height, productivity, etc.). As the government's capacity to gather data increases in the future, there will be a need to add more "animal events" (such as weighing animals, the quantity and quality of milk, etc.) . With multiple events per cow, multiple cows per farm, and multiple farms per area, the volume of data points will easily grow from the thousands to the millions going forward. An interoperable system thus needs to have "strong bones" and the agility to accommodate new, decision-critical data points in the future.

DG is setting up the LIS to support millions of records. The existing individual systems do not currently support that volume of records; neither do they have the hardware and software infrastructure needed to absorb increases in new data (which is one of the stated goals of the MoA). Accordingly, DG is setting up appropriate infrastructure to make the LIS ready for expansive data volumes.

Designing with scalability in mind means avoiding bottlenecks in data collection and processing and favoring large-scale data ingestion to handle potential future surges in data generation. *In this case, it means recognizing that the external systems integrated with the aLIVE LIS can go offline for days (or even months). Fortunately, the LIS has been designed in a manner that allows it to process all data generated by systems that previously went offline. Building for this kind of asynchronous data exchange makes the overall interoperable system more resilient. The system can also act as a funnel so that the subsequent mapping (the addition of new events or data points) does not create a choke point.*

Creating interoperable frameworks is not just about solving current and past problems but also preparing for the future. New systems will inevitably be added after the interoperable framework project concludes; therefore, extensibility and pluggability[28] must be inherent in the design. This approach ensures that incorporating new systems will not require extensive changes to the existing framework.

28 Extensibility and modularity. (n.d.). Adobe Developer. Retrieved October 28, 2024, from https://developer.adobe.com/commerce/php/architecture/modules/

To summarize:

BEST PRACTICE	IMPLEMENTATION
Vendor Management and Licensing	When adding or improving software components in your infrastructure, ensure the source code is well-documented and appropriately licensed. If proprietary, negotiate a special license that guarantees future access and flexibility. Aim to avoid vendor lock-in by ensuring your solution can be maintained or transferred without relying on a single vendor.
Cost Management and Sustainability	Conduct regular reviews of operational and maintenance costs to ensure the solution remains financially sustainable. Ensure that the platform can scale efficiently in terms of performance and price as usage grows.
Technology and Infrastructure Updates	Review and update the technology stack regularly to prevent technical debt and security vulnerabilities. Consider the platform's full lifecycle impact, from development to decommissioning.
Data Management and Governance	Implement data retention, archival, and disposal policies to manage costs and comply with changing regulations.
Documentation and Knowledge Transfer	Maintain up-to-date and accessible documentation for all components, processes, and decisions. If you update any component, ensure the vendor or participating team updates the related documentation.
Team Development and Training	Invest in continuous learning and upskilling so the team can keep pace with technological advancements.
Governance and Compliance	Ensure the platform complies with current and anticipated regulations, adjusting as needed.

The solution itself should be scalable. Interoperability frameworks aim to interconnect multiple systems, each capable of producing data. At the same time, the architecture should leave room for future systems to be added.

Chapter 5

Looking Forward to Better Interoperability

Creating interoperable systems is complex, requiring content and context specificity. In this paper, we have used an emerging example from DG's work in Ethiopia's livestock sector to demonstrate some of the key technical, human, and data considerations that should shape interoperable systems. However, we have not considered everything. Several factors still need to be considered, implemented and shared with the community of users, implementers, and funders seeking improved service delivery and decision-making.

Recommendations

1. View digital transformation on a maturity axis, rather than a transition from point A to B.

What Ethiopia needs to accelerate digital transformation is different from Zambia, and different still from India. The broader digital ecosystem and the targeted systems must be assessed from the outset to allow for the development of reasonable digitization and system interoperability metrics tailored to each country's context.

2. Understand what already exists to set priorities.

No one likes audits, but they exist for a reason. It is essential to understand – at a technical level – what systems are already in place, how they are constructed, and how they operate. With a holistic understanding of the existing state of a ministry's or organization's technology ecosystem, leaders can identify the systems that need to be joined or decommissioned; they can also ascertain the resources needed for continuity (see point 3).

3. Use what exists to define the scope.

Not all systems can (or should) be integrated at one time. Accordingly, the aLIVE program works with five out of more than three dozen systems used by the MoA. The program built off the livestock roadmap development process to prioritize the data (and systems) reflecting the most data points featured in policymaker decision-making. What systems are likely to have the most impact on policies that affect farmer productivity? Use this framing, as an example, to set realistic interoperability goals.

4. Institutionalize sustainability from the outset in every dimension: human, institutional, technical, and financial.

Interoperability is not only for datasets; it is for people,1 finances, and hardware as well. Interoperable systems do not exist in a vacuum, and initiatives to build this type of infrastructure must think through how individual components scale together.

5. Take a portfolio view of digital transformation that puts interoperability at the center.

Not every project or initiative can include standards development, technical infrastructure building, and capacity building at multiple levels (even though such an expansive scope is definitely beneficial!). Public institution working groups and donor coordination groups have power in this respect. By first setting a roadmap that views these specific components as necessary for the entire system, government agencies, the private sector, civil society, and development partners can take a "divide and conquer" approach to the building blocks of interoperability (as resources and partnerships allow) that is aligned with a central vision. While approaches covering the building blocks of interoperability are ideal, targeted approaches aligned to an overarching vision would be the next best alternative.

a. *The latter approach is not impossible to reach. Several countries have digital roadmaps[2] that outline their plans for digital investments, sometimes across whole sectors. Funders and implementers can use these roadmaps as a guide. In cases where roadmaps do not exist or do not address the building blocks and layers necessary for successful interoperability, the approaches outlined in the paper can be used to reorient existing and upcoming activities.*

6. Recognize that transformation requires champions.

The aLIVE program's ability to co-design approaches and meet data standards, governance, and technical needs was due in large part to champions who contributed to the vision of a more effective livestock data ecosystem for Ethiopia. Champions at the funder, ministerial, system owner, and regional/field levels were and continue to be critical to the program's success, and more importantly, they can identify blockers and work collectively to find cross-departmental, and even cross-border, solutions in real-time as the work progresses.

7. Ensure data standards stretch across sectors and regions.

Developing standards with accompanying stakeholder buy-in is a time-consuming process. It takes years to develop standards for every possible sub-sector and sector. While working with smaller subsets of actors and data (such as livestock or soil health data) may make sense when developing standard prototypes, the scaling of such efforts will prove to be difficult. International initiatives such as the International

1 Carfi, N. (2023, November 27). To achieve data interoperability, we need to start with "people interoperability." World Bank Blogs. https://blogs.worldbank.org/en/opendata/achieve-data-interoperability-we-need-start-people-interoperability

2 For example, Rwanda, Nigeria, and Kenya, to name some examples, each have a digital master plan, roadmap, or strategy that helps dictate digital agriculture priorities across the government.

Committee on Animal Recording (ICAR)3 provide some hope in terms of standards alignment, but more effort is needed. Farmers have cattle, crops, soil, and fertilizer demands on the same plot of land. Service delivery and decision-making need to work toward a better-unified data picture to holistically boost farmer productivity.

8. Keep watching and actively collaborating with telecommunications outfits as the growth of local cloud provider options continues.

The growth of local, regional, and international cloud options will continue to reshape what is possible in terms of the technical infrastructure required to house large volumes of interoperable data. Investments in consistent, cost-effective, and accessible cloud storage options throughout the Global South must go hand in hand with pushes to make more data and systems interoperable.

9. Invest in national legal frameworks for responsible data sharing.

Open data policies, interoperability protocols, and other similar documentation allow for the reinvention of data infrastructure, thereby addressing issues of data asymmetry and enabling a more equitable distribution of information. These frameworks are essential for enabling public, private, and civil society actors to share and integrate data across various platforms. When system owners are forced to operate piecemeal, one Memorandum of Understanding at a time, it slows down integration, access, and use, for everyone. ACoupled with personnel changes, changing mandates, and competing demands, the paper trail establishing data sharing becomes increasingly difficult to follow. There remains a strong advantage to national laws mandating data sharing that take the "guesswork" out of determining what to share, and when.

10. Do not forget the people.

Farmers, pastoralists, business owners, and citizens are the primary producers of data, yet their role in the interoperability of data systems often goes unrecognized. As interoperability becomes more common, systems will need to ensure that the voices and experiences of different types of users are heard and integrated into the decision-making processes, incorporating the gaps as well as the opportunities that interoperable data can create for people. Improved decision-making through joined-up data still requires that decisions are connected to real user needs and rights.

The four previous chapters have described the intricacies of data and system interoperability and emphasized their importance for an AI-driven future filled with complex development challenges. Chapter 2 detailed how a data standard for livestock came together, involving many people, workshops, and opportunities for ownership and adaptation for livestock system owners in Ethiopia. Chapter 3 picked up from where the data standard

3 ICAR Guidelines. (n.d.). Retrieved October 24, 2024, from https://www.icar.org/index.php/icar-recording-guidelines/

left off, diving into the importance of creating a data governance ecosystem in which data sharing lies the foundation of interoperable systems, walking through how to work collectively with partners to develop data management, access, and sharing protocols that can honor individual system processes while maximizing the potential for safe data sharing for future centralized systems. Lastly, Chapter 4 broke down the elements of "good" technology, walking through the pros and cons of software (open and proprietary), hardware, and storage (on-site and cloud) considerations that must underpin interoperable data and systems.

Ultimately, developing inclusive, sustainable, and interoperable systems is a lot of work. Nonetheless, it is worth it! The approach we have laid out is designed with a future view in mind, focused on laying solid foundations for emerging technologies and solutions that cannot yet be seen. Globally, we need digital infrastructure that can adapt with us: infrastructure that is ready for more data; and for complex challenges, like climate change, that require drawing on a broad variety of data to understand a dynamic landscape. Why retrofit when we can start smart?

Annex 1.
What Is in the Pillars?

I. Strategic Governance: Focuses on the development of the LIS Data Governance Framework. The framework will encompass the overall structure put in place to ensure the sustainability of LIS at the MoA (committees, teams, personnel). It will also define the roles and responsibilities of stakeholders, including those beyond the five core system owners. These stakeholders include regional entities as well.

II. Regulatory Compliance: Focuses on the development of core documents such as the data access and sharing protocol,[1] data sharing agreement template,[2] data request form,[3] data privacy and cookie policy,[4] and data breach protocol and open data license statement.[5]

III. Cross-cutting governance: In the process of making the five databases interoperable and building the LIS, livestock experts, software engineers, lawyers, and other experts will carry out various activities. Hence, there are cross-cutting governance issues that must be taken into consideration in the development of the governance framework:

- *The technical aspect of data governance: Addresses how permissions/authorisation systems will be built into the back-end of the LIS and specify what security protocols will be in place to protect the data in the system.*

- *Data quality and lifecycle management: Ensures documentation of processes and rules for data cleaning, data validation, and ongoing monitoring of data quality. Furthermore, outlines rules for data retention and disposal policies, as well as processes for data archiving and backup.*

- *Technical coordination and system management - Focuses on the coordination and management of the distributed architecture of the LIS. This component takes into account the higher-level considerations that need to be made in the future. (e.g., How are data and system security protocols standardized across the core systems? How frequently are reviews of international standards undertaken to ensure that the most up-to-date and appropriate ones are being used? Who decides what standards to use and how they will be applied across the core systems? How frequently are reviews of the software and hardware (both cloud and local server-based) undertaken and who decides what updates are made?)*

---

1 LIS Data Access and Sharing Protocol sets out the data access permissions for users of the Ethiopian LIS. It aims to establish transparent and clear guidance for prospective and active system users on how LIS data access is managed and controlled.

2 The LIS data sharing agreement is an agreement between the LIS administrator/MoA/ and another person that outlines which data will be shared and, most importantly, how the data can be used. It is hoped that it will prevent data misuse, data abuse, and unregulated data dissemination.

3 LIS Data Request form is a form in which the requesting party asks the LIS administrator (MoA) for a specified level of access of data in the LIS and a specific format.

4 LIS Data Privacy Policy includes information about all the data LIS collects, processes, stores, and/or displays. LIS cookie policy looks strictly at the cookies that track user data.

5 These two documents are likely to become eventual annexes to the LIS's overarching Data Governance Framework document.

Annex 2.
Program Governance Structure

An overarching aLIVE governance structure will ensure the scale-up and implementation of the LIS strategic and operational plans. The following committees have been established with various roles and responsibilities, the details of which are explicitly narrated below:

1. Governing Committee (GC)
2. Technical Implementation Committee (TIC)
3. External Advisory Group (EAG)
4. Task Forces (TF)

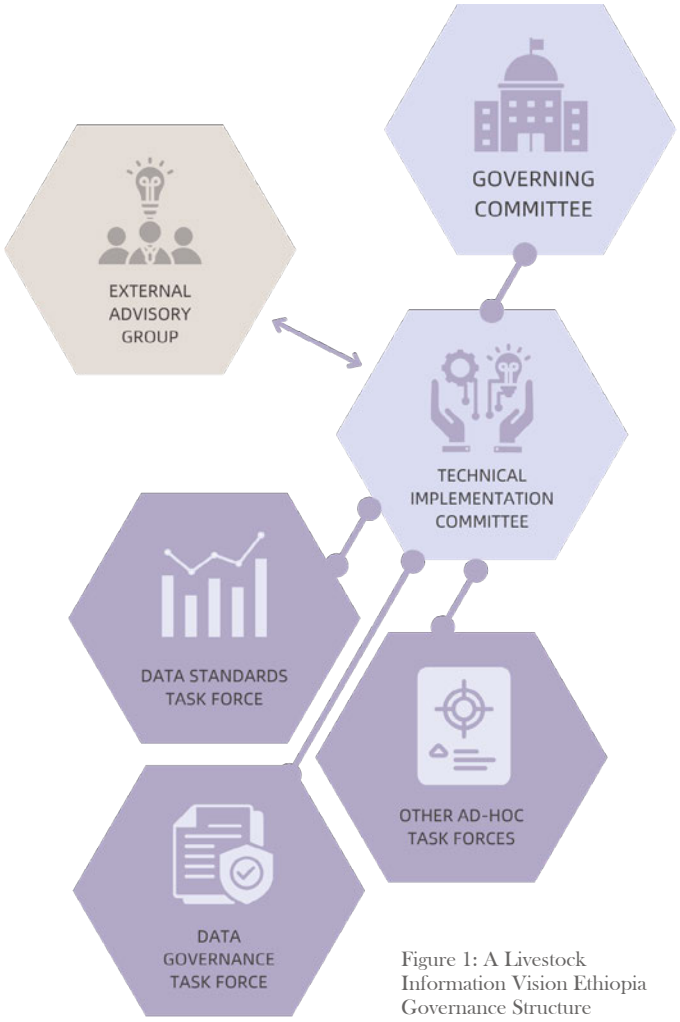


Figure 1: A Livestock Information Vision Ethiopia Governance Structure

1. Governance Committee :

The aLIVE GC functions as an oversight and final decision-making body, and will provide overall governance, strategic direction, and support for its sustainability. The GC will be chaired and co-chaired by the Minister for Agriculture and the State Minister for Livestock. It will also primarily comprise some of the relevant directorates in the ministry as well as representatives from donor and implementing partners.

2. Technical Implementation Committee

The overall objective of the TIC is to provide technical guidance, support, and expert advice to the project team.

3. External Advisory Group

The EAG will be identified and engaged periodically and will not have a formal governance function. The EAG will bring local livestock expertise to advise the program and ensure that potential external data users are continuously engaged and familiarized with the program and LIS data. The EAG will ensure that the latest developments in data standards and technical innovations in livestock data are considered; it will also guarantee that the needs of external users are continuously captured and incorporated into the planning and design processes of the program. Detailed aLIVE technical issues will be dealt with by the TF and technical experts whenever needed.

4. Task Force

This ad-hoc committee will be established when needed to provide guidance on specific issues, such as the Data Standard and Governance Task Force. Other ad-hoc TFs will be established for specific tasks as needed under the TIC.

Bibliography

1. Alemayehu, S. (2021). A livestock information system roadmap for Ethiopia. Alliance Bioversity & CIAT. https://alliancebioversityciat.org/publications-data/livestock-information-system-roadmap-ethiopia

2. Amazon, A. W. S. (n.d.). What is a RESTful API? AWS Amazon. Retrieved October 28, 2024, from https://aws.amazon.com/what-is/restful-api/?nc1=h_ls

3. AMAZON. (n.d.). AWS Lambda. AMAZON. Retrieved October 28, 2024, from https://aws.amazon.com/lambda/

4. AMAZON. (n.d.). FAQ sur l'offre gratuite d'AWS. Aws Amazon. Retrieved October 28, 2024, from https://aws.amazon.com/free/free-tier-faqs

5. Announcing development gateway's aLIVE program: Advancing livestock data in Ethiopia. (2022, December 2). Development Gateway: An IREX Venture. https://developmentgateway.org/blog/announcing-development-gateways-alive-program-advancing-livestock-data-in-ethiopia/

6. Arzoyan, L. (n.d.). Governments Need a Citizen-Centered Digital Transformation. https://www.gmfus.org/news/governments-need-citizen-centered-digital-transformation

7. BrennaP. (2023, September 22). 7 benefits of data governance for your organization. DataGalaxy. https://www.datagalaxy.com/en/blog/7-benefits-of-data-governance-for-your-organization/

8. Building an interoperable space for smart agriculture. (2023). Digital Communications and Networks, 9(1), 183–193.

9. CABI collaborated with the MoA in developing the Soil and Agronomy Data Management Use and Sharing Directive No. 974/2023. No-974-2023 - GeoNode. Retrieved October 28, 2024, from https://nsis.moa.gov.et/documents/134

10. Carfi, N. (2023, November 27). To achieve data interoperability, we need to start with "people interoperability." World Bank Blogs. https://blogs.worldbank.org/en/opendata/achieve-data-interoperability-we-need-start-people-interoperability

11. Carletto, C. (2021). Better data, higher impact: improving agricultural data systems for societal change. European Review of Agricultural Economics, 48(4), 719–740.

12. Chander, A., & Sun, H. (2023). Data Sovereignty: From the Digital Silk Road to the Return of the State. Oxford University Press.

13. Cloud run functions. (n.d.). Google Cloud. Retrieved October 28, 2024, from https://cloud.google.com/functions

14. Cloudflare. (n.d.). What is vendor lock-in? | Vendor lock-in and cloud computing. Cloudflare. Retrieved October 28, 2024, from https://www.cloudflare.com/learning/cloud/what-is-vendor-lock-in/

15. Committer, B. a. (n.d.). Briefing: The Apache way. Retrieved October 28, 2024, from https://www.apache.org/theapacheway/

16. Committer, B. a. (n.d.). Welcome to the Apache software foundation! Retrieved October 28, 2024, from https://www.apache.org/

17. Data governance key components: What to know in 2024. (2023, May 5). https://atlan.com/data-governance-key-components/

18. Data Management Association. (2017). DAMA-DMBOK: Data Management Body of Knowledge.

19. Development Gateway. (2022, December 2). Announcing Development Gateway's aLIVE Program: Advancing Livestock Data Ethiopia. Development Gateway. https://developmentgateway.org/blog/announcing-development-gateways-alive-program-advancing-livestock-data-in-ethiopia/

20. Development Gateway. (2024). aLIVE Program Reaches Milestone: Livestock Data Standards Endorsed by Ethiopia's Ministry of Agriculture. https://developmentgateway.org/blog/alive-program-reaches-milestone-livestock-data-standards-endorsed-by-ethiopias-ministry-of-agriculture/

21. Diebold, G. (2023, September 25). Overcoming Barriers to Data Sharing in the United States. Data Innovation https://www2.datainnovation.org/2023-data-sharing-barriers.pdf

22. Digital Public Goods Standard. (2020, September 21). Digital Public Goods Alliance - Promoting Digital Public Goods to Create a More Equitable World; Digital Public Goods Alliance. https://digitalpublicgoods.net/standard/

23. Ethiopia Data Use Partnership. (2022, August 11). Ethiopia Data Use Partnership (DUP). https://ethiopiadup.jsi.com/

24. Extensibility and modularity. (n.d.). Adobe Developer. Retrieved October 28, 2024, from https://developer.adobe.com/commerce/php/architecture/modules/

25. Feed the Future Innovation Lab for Livestock Systems. (2021). Ethiopia's Livestock Systems: Overview and Areas of Inquiry. The U.S. Agency for International Development. https://livestocklab.ifas.ufl.edu/media/livestocklabifasufledu/pdf-/LSIL_Livestock_Systems_Overview_Ethiopia_2021_08.pdf

26. Fothergill, B. T., Knight, W., Stahl, B. C., & Ulnicane, I. (2019). Responsible Data Governance of Neuroscience Big Data. Frontiers in Neuroinformatics, 13, 28.

27. Garcia, C. (2023). The crucial role of interoperability in driving value-based care transformation. https://www.athenahealth.com/knowledge-hub/value-based-care/interoperability-helps-identify-care-gaps

28. GraphQL. (n.d.). GraphQL. GraphQL. Retrieved October 28, 2024, from https://graphql.org/

29. Greene, D. N., McClintock, D. S., & Durant, T. J. S. (2021). Interoperability: COVID-19 as an Impetus for Change. Clinical Chemistry, 67(4), 592–595.

30. HIMSS. (n.d.). Interoperability in Healthcare. https://www.himss.org/resources/interoperability-healthcare

31. ICAR Guidelines. (n.d.). Retrieved October 24, 2024, from https://www.icar.org/index.php/icar-recording-guidelines/

32. Investing in Interoperability : Insights from a Landmark Accenture Study. (n.d.). Workday. Retrieved October 23, 2024, from https://beyonderpdrivingthedigitalbusiness.cio.com/wp-content/uploads/sites/44/2024/06/investing-in-interoperability-guide-enus.pdf

33. ISRIC. (2016). Africa Soil Information Service (AfSIS). https://www.isric.org/projects/africa-soil-information-service-afsis

34. Krombholz, K. (n.d.). To cloud or not to cloud: A qualitative study on self-Hosters' motivation, operation, and security mindset. Retrieved October 28, 2024, from https://www.usenix.org/system/files/usenixsecurity23-grober.pdf

35. Makhlouf, R. (2020). Cloudy transaction costs: a dive into cloud computing economics. Journal of Cloud Computing Advances Systems and Applications, 9(1). https://doi.org/10.1186/s13677-019-0149-4

36. Mann, T. (2022, May 4). Concerned about cloud costs? Have you tried using newer virtual machines? The Register. https://www.theregister.com/2022/05/04/aws_amd_intel/

37. Ministry of Agriculture. (n.d.). Ethiopia-Livestock-data-standard. Github. Retrieved October 24, 2024, from https://github.com/moa-et/Ethiopia-Livestock-data-standard

38. Morales, L. G., & Orrell, T. (2017). Data interoperability: A practitioner's guide to joining up data in the development sector (p. 10). https://www.data4sdgs.org/sites/default/files/services_files/Interoperability%20-%20A%20practitioner%E2%80%99s%20guide%20to%20joining-up%20data%20in%20the%20development%20sector.pdf

39. Moses, B. (2024). Survey Says: Data Quality Management Isn't Evolving Fast Enough for AI. https://barrmoses.medium.com/survey-says-data-quality-management-isnt-evolving-fast-enough-e51dc4de5315

40. OAuth. OAuth 2.0. Retrieved October 28, 2024, from https://oauth.net/2/

41. Opara-Martins, J., Sahandi, R., & Tian, F. (2016). Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective. Journal of Cloud Computing, 5(1), 1–18.

42. Orrell, T., & González, L. G. (2021). Introducing the joined-up data maturity assessment. https://www.data4sdgs.org/sites/default/files/file_uploads/Joined_Up_Data_Maturity_Assessment_draft5.pdf

43. Principles for Digital Development. (2014). Principles for digital development. Principles for Digital Development. https://digitalprinciples.org/

44. Rucker, D., Hasan, A., Lewis, L., & Tao, D. (2020). Advancing interoperability together globally.

Global Digital Health Partnership. https://gdhp.health/wp-content/uploads/2022/11/Advancing-Interoperability-Together-Globally-2.pdf

45. Schnier, J. (2008). XML Introduction. In Flash XML Applications (pp. 3–9). Elsevier.

46. Shim, T. (2024, April 5). What is latency: Impact of server locations to your site performance. HostScore. https://hostscore.net/learn/what-is-latency-impact-of-server-locations-to-your-website-performance/

47. Sirkemaa, S. (2018). Information Systems Infrastructure – Importance of Robustness. Information Technology Science, 241–244.

48. Snaith, B. (2023). What do we mean by "without data, there is no AI"? https://theodi.org/news-and-events/blog/what-do-we-mean-by-without-data-there-is-no-ai/

49. SOAP vs REST APIs: Which is right for you? (n.d.). Retrieved October 28, 2024, from https://www.soapui.org/learn/api/soap-vs-rest-api/

50. Steele, L; Orrell, T (2017). Defining and approaching interoperability.

51. W3schools. (n.d.). What is JSON. w3schools. Retrieved October 28, 2024, from https://www.w3schools.com/whatis/whatis_json.asp

52. Warner, B. (2024, September 17). Mythbusters: Digital public infrastructure & digital public goods. Development Gateway: An IREX Venture. https://developmentgateway.org/blog/mythbusters-digital-public-infrastructure-digital-public-goods/

53. What We Know So Far: Best Practices in Developing Data Governance Frameworks. (2022, August 18). Development Gateway: An IREX Venture. https://developmentgateway.org/blog/what-we-know-so-far-best-practices-in-developing-data-governance-frameworks/

54. Wikipedia contributors. (2024, August 17). Latency (engineering). Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Latency_(engineering)&oldid=1240837898

55. Wikipedia contributors. (2024, July 24). On-premises software. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=On-premises_software&oldid=1236371187

56. Wikipedia contributors. (2024, October 12). Infrastructure as a service. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Infrastructure_as_a_service&oldid=1250725639

57. Wikipedia contributors. (2024, October 27). Cloud computing. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Cloud_computing&oldid=1253652807

58. Williamson, H. F., & Leonelli, S. (2022). Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development. Springer Nature.