USAID GLOBAL HEALTH SUPPLY CHAIN PROGRAM
Procurement and Supply Management

Open-Source Software Due Diligence Review

Traceability Interoperability Platform MVP & ARV Traceability Data Exchange Pilot

March 20, 2024

The USAID Global Health Supply Chain Program-Procurement and Supply Management (GHSC-PSM) project is funded under USAID Contract No. AID-OAA-I-15-0004. GHSC-PSM connects technical solutions and proven commercial processes to promote efficient and cost-effective health supply chains worldwide. Our goal is to ensure uninterrupted supplies of health commodities to save lives and create a healthier future for all. The project purchases and delivers health commodities, offers comprehensive technical assistance to strengthen national supply chain systems, and provides global supply chain leadership. GHSC-PSM is implemented by Chemonics International, in collaboration with Arbola Inc., Axios International Inc., IDA Foundation, IBM, IntraHealth International, Kuehne + Nagel Inc., McKinsey & Company, Panagora Group, Population Services International, SGS Nederland B.V., and University Research Co., LLC. To learn more, visit ghsupplychain.org

USAID FROM THE AMERICAN PEOPLE

PEPFAR U.S. President's Emergency Plan for AIDS Relief

Table of Contents

Table of Contents 1

Introduction 2

Product/Tool Due Diligence 4

OpenEPCIS Document Version Convertor and Format Convertor 4

Apache Superset 5

Elasticsearch & Kibana (ELK stack) 7

Pulse by NABP 9

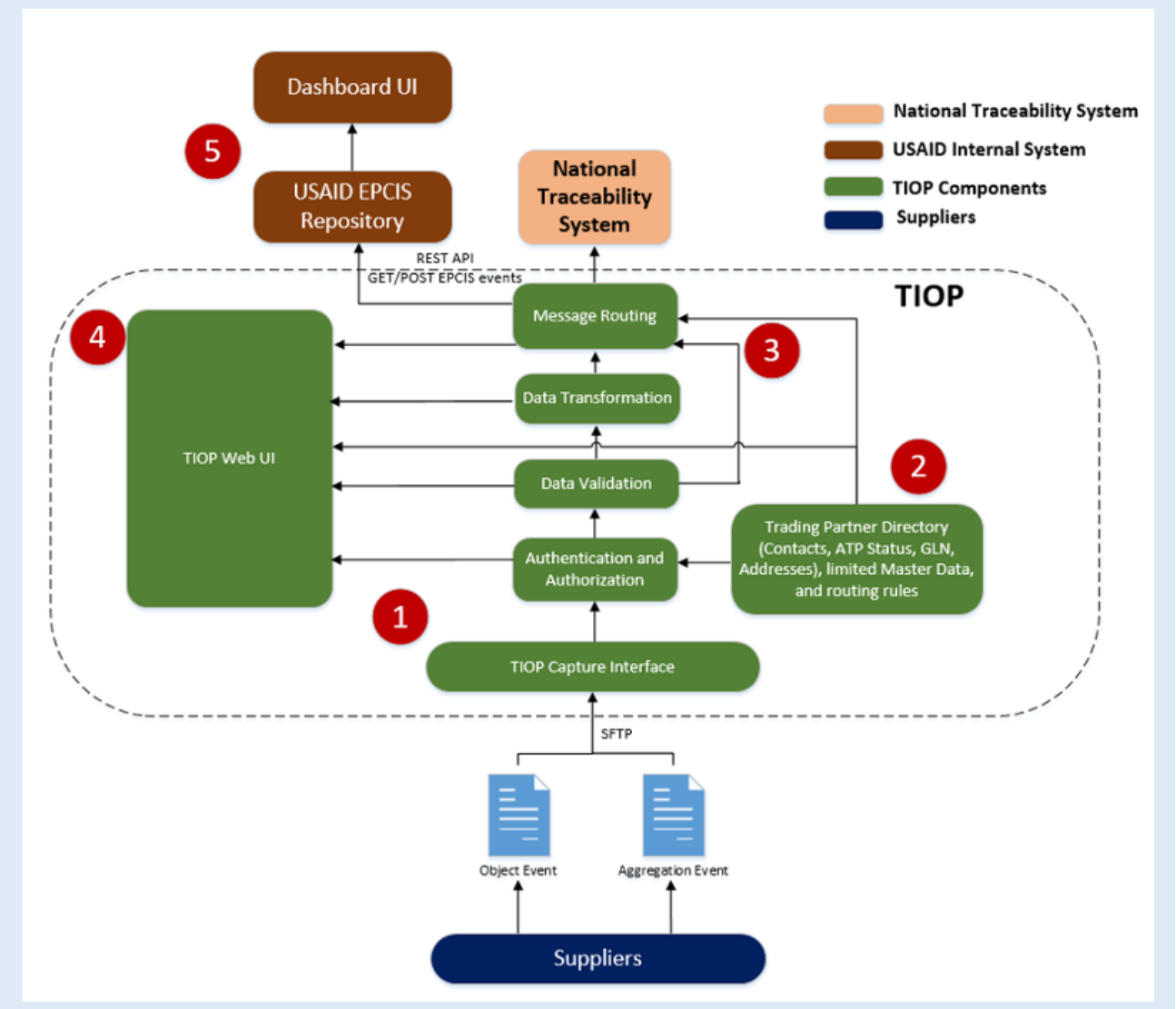
Pulse SWOT Analysis for TIOP 10

Analysis and Path Forward 10

Proposed Engagement Plan for Global Health GLN Registry 11

Introduction

Open-source products or tools are community developed, instead of commercial software company developed. Commercial software success depends on its product's security, support, pricing structure, legal and compliance adherence. While commercial software company products have licensing fees and/or subscription fees, they also have strict oversight mechanisms. Open-source software product development is self-governed. Open-source products have no licensing cost, and much of the governance responsibility falls onto the consumer of the open-source product. The TIOP and USAID EPCIS Repository technical team will utilize open-source products and tools. where available, for various components of the TIOP and EPCIS Repository. The document assesses the governance, security, ease of integration and other factors for consideration when utilizing these tools for the TIOP MVP. The following figure shows the TIOP architecture components at a high level and USAID EPCIS Repository and proposed open-source tools to develop a well-functioning solution responsive to the requirements and functional specifications. This document should be reviewed in consort with the TIOP Development and Resource Plan.



The following table shows the architecture elements utilizing proposed open-source products/tools:

Architecture Component	Open-Source Product/Tool
Component 3 – Document Transformation	OpenEPCIS Document Version Convertor and Format Convertor
Component 4 – TIOP Web UI	Apache Superset
Component 5 – USAID EPCIS Repository and Dashboard	Elasticsearch/Kibana

The following factors are considered for the open-source product due diligence:

Due Diligence Factor	Analysis Description
License Compliance	Check the open-source license(s) under which the product is released. License compliance is essential for legal compliance.
Onboarding Requirements	Analyze the system requirements, approximate timeframe and technical development resources needed to onboard the open-source product.
Code Maintenance	Describe frequency of updates to code and response rate for issue resolutions.
Community and Support	Checks community participation and availability of support. Community blogs and active involvement will inform the extent to which solutions to the problems are addressed and understand the optimum usage patterns and adaptive management
Security Vulnerabilities	Identify potential security vulnerabilities or weaknesses.
External Dependencies	Understand the dependencies of the open-source product and tools to other tools. Dependencies may pose their own licensing, security, and maintenance risks
Quality of Documentation	Check the open-source software documentation. Check if it includes configuration instructions, installation guide, user manual, API documentation, etc.
Ease of Integration	Analyze the open-source product integration mechanism such as REST API, SDK
Cost of Ownership	Even though open-source products and tools have no licensing cost, the dependencies on other products or hosting infrastructure can incur cost
IP and Patent Risk	Understand the IP or patent clause associated with the open-source product. License documents typically contain the clauses regarding IP and Patent protection
Sustainability	Understand the support and maintainability concerns and long-term contract obligations.
Data Segregation	Understand if the tool stores the data. If the tool stores the data, then understand the nature of the data and concerns regarding data integrity.
Cost Estimate	Cost estimate for the software as well as the hosting environment.
Time Estimate	Time estimates to install and configure the software.

Product/Tool Due Diligence

OpenEPCIS Document Version Convertor and Format Convertor

Tool Description

OpenEPCIS Document Conversion tool is an open-source tool which converts EPCIS 1.2 documents to EPCIS 2.0 documents. This tool can also convert EPCIS documents in XML format to EPCIS documents in JSON format and vice versa.

Utilization within TIOP

TIOP will utilize the OpenEPCIS Document Version Convertor and Format Convertor to convert the EPCIS documents, which will be in version 1.2, from suppliers to version 2.0. The USAID EPCIS repository will store the EPCIS document in version 2.0 format.

Due diligence Factor	Due diligence Output
License Compliance	OpenEPCIS Document Conversion tool is a free open-source tool distributed with Apache 2.0 license. OpenEPCIS License
Onboarding Requirements	The benelog GmbH & Co: team provides a set of documentation to support the onboarding process. OpenEPCIS Document Conversion Git Repository
Code Maintenance	The benelog GmbH & Co: team primarily maintains and contributes to the Document Conversion code repository. The benelog monitors the issues log. OpenEPCIS Document Conversion Git Repository
Community and Support	The benelog GmbH & Co: OpenEPCIS team liaises closely with GS1 Global on current EPCIS trends and have informed the deployment of EPCIS 2.0. The benelog GmbH & Co: OpenEPCIS team will support TIOP team through email, and over the phone. This team also monitors the issues logged in GitHub.
Security Vulnerabilities	The OpenEPCIS team runs a static code analysis tool on all the source code to find out the security vulnerabilities in the source code. The OpenEPCIS team approves the code contribution to the Git repository before adding it in the distribution.
External Dependencies	The external dependencies are included in the build and deployment framework. No external software or tools need to be installed separately. All the dependencies are open source. The only dependency which will incur cost is the hosting environment. Hosting environment can be containerized, Cloud VM or local machine.
Quality of Documentation	The benelog GmbH & Co: OpenEPCIS team has provided the configuration and usage guide for EPCIS Document Conversion tool. OpenEPCIS Document Conversion Tool
Ease of Integration	The Document Conversion tool can be included in the Java project as a dependency and use the java classes in the logic. The Document Conversion tool can also be exposed as an API. The OpenEPCIS

	team provided a Docker image to run the Document Conversion tool as a service in a container.
Cost of Ownership	Since the Document Conversion tool is an open-source tool, there is no cost of ownership for the tool. The hosting environment for this tool will incur the cost. This tool can be hosted in a Containerized VM, Cloud VM or local server.
IP and Patent Risk	OpenEPCIS Document Conversion tool is protected under Apache License 2.0. IP and Patent protection clauses are included in the license. OpenEPCIS License
Sustainability	Since the OpenEPCIS Document Conversion tool is an open-source project, there is no restriction on bundling and distributing the development artifacts. It is advisable that this tool is installed from the latest release in the GitHub repository, instead of from the external distribution. GitHub repository has the latest code fixes and additional features. Besides the GitHub issues log and documentation, OpenEPCIS team will provide support through email and phone, when needed. Developers and system admins can keep track of updates on GitHub repository by setting the email notification on GitHub.
Data Segregation	OpenEPCIS Document Convertor does not store data. Since this is an open-source tool, processing happens within the boundary of organization infrastructure.
Cost Estimate (approx.)	There are costs associated with the hosting environment: VM - $9-$12 per month Containerized Service - $30-$35 per month No separate cost if included within a microservice. Serverless Function - $10 per month Team will decide the mechanism by consulting with OpenEPCIS team.
Time Estimate (approx.)	12 hrs. to 16 hrs.

Apache Superset

Product Description:

Apache Superset is an open-source data visualization platform. Superset comes with over 40 prebuilt visualizations, including charts and graphs. Superset allows developers to build custom visualizations as well. Supersets can connect with RDBMS NoSQL databases and document databases.

Utilization within TIOP:

TIOP will utilize the Apache Superset to develop the TIOP operational dashboard, with various types of graphs and charts, to visualize the TIOP operational data.

Due diligence Factor	Due diligence Output
License Compliance	Superset is licensed under Apache License. Apache License contains comprehensive clauses related to Copyrights, Patents, Warranty, Liability and more, See the link below for Apache License details.

	Apache License
Onboarding Requirements	Apache Superset can be installed and configured on containerized VM or VPC (Virtual Private Cloud) VM. To keep the installation generic, TIOP will utilize VPC VM. There is a detailed installation and configuration guide on Apache Superset web site. Apache Superset Installation Guide
Code Maintenance	Apache Superset Git project has more than 1000 contributors and more than 14000 commits. This project is active and regularly updated. See below for the link to the Git repository. Apache Superset Git repository
Community and Support	Apache Superset has active communities on GitHub, Slack and Stack Overflow, along with a mailing list. See below for the link to the Apache Superset Community home page. Apache Superset Community home page
Security Vulnerabilities	Apache Superset supports role-based authentication and authorization. Apache Superset utilizes Flask-AppBuilder, which is an application development framework. Flask-AppBuilder supports various security mechanisms such as LDAP, OAuth, etc. Row level security can also be implemented in Superset. Developers can use Flask Talisman plug-in for mitigating cross site scripting attacks. Apache Superset Security
External Dependencies	Apache Superset has several external dependencies. All the dependencies are open-source, so no cost incurred, but hosting the dependencies on either cloud or on-prem infrastructure will have cost associated. Superset has multiple configuration opinions with the following external dependencies: Docker Kubernetes Using Python package Apache Superset on a Linux VM Apache Superset utilizes Flask-Caching for caching the data. Flask-Caching requires one of the following caching tools. Redis Open-source Redis needs a Linux VM to install on. Redis can be installed in Docker. Apache Superset recommends Redis caching. Local File System Since Redis is the recommended caching solution, TIOP will utilize Redis for Apache Superset caching. Apache Superset has the following additional dependencies. Metastore SQLLite – Recommended for development, unit testing and integration test only. MySQL, PostgreySQL – Recommended for UAT and Production. To avoid the deployment complications, TIOP will utilize the same technology in all environments. TIOP will utilize MySQL in all environments. Webserver

	Gunicorn Python HTTP Server – Recommended by Apache Superset Nginx Apache TIOP team will research further on the choice of webserver.
Quality of Documentation	Apache Superset has comprehensive installation, configuration, and development documentation. Apache Superset Documentation
Ease of Integration	Apache Superset has REST APIs for all data that user can view on web dashboard. Some of the key REST APIs are: Charts Dashboards Databases Datasets Users Apache Superset REST API Apache Superset can be integrated with other web applications using iframe.
Cost of Ownership	Apache Superset and its dependencies are open source, hence no software licensing or subscription cost. There is a cost involved in hosting Superset and its dependencies. Cost of hosting depends on the hosting environment such as cloud hosting or on-premises hosting.
IP and Patent Risk	Apache Superset is distributed under Apache License. Apache License has specific clauses for patents, trademarks, and copyrights. Apache will revoke the license in case of infringement of any of the clauses. Apache License
Sustainability	Since Apache Superset is an open-source project, there is no restriction on bundling and distributing the development artifacts. Apache Superset can be hosted in containerized VMs, Cloud VMs or locally. Since this is a community developed project, there is no single point of support contact, but there is an issue repository on GitHub to post the issues and get help.
Data Segregation	Apache Superset does not store data. Users of the dashboard will use the Internet Browser from outside the organization's boundary to visualize the data.
Cost Estimate (approx.)	The following are the costs associated with hosting: VM - $9-$12 per month Containerized Service - $30-$35 per month Route 53 (DNS) - $200-$270. Per month
Time Estimate (approx.)	36 hr. To 48 hr. Per month

Elasticsearch & Kibana (ELK stack)

Product Description:

Elasticsearch is a search and analytics engine. It allows for large volume of data storage and query. Kibana is a data visualization tool with visualization and analytics capabilities. Elasticsearch & Kibana are part of ELK Stack, which contains data ingestion, query, and visualization.

Utilization within TIOP:

Elastic Search will be utilized as an EPCIS repository. Kibana will be utilized for querying and visualizing the EPCIS documents, which are stored in the Elastic Search EPCIS repository. Kibana and Elastic Search will be hosted outside of the boundary of TIOP, since EPCIS repository is a separate system outside of TIOP.

Due diligence Factor	Due diligence Output
License Compliance	The self-managed open-source ELK stack is licensed under Serverside Public License, Elastic License and Apache License. Open-source Elastic stack license The open-source software bundle comes with the licenses. There is no need for a separate license. procurement.
Onboarding Requirements	Elastic stack needs to be installed on VPC. It includes configuration of Elastic stack on VM, configuring storage, configuration of web server, configuration of DNS.
Code Maintenance	Elasticsearch's repository and Kibana's repository both have more than 75000 commits. Elastic team updates the open-source code regularly. Third party developers can also commit to the changes. Upon committing the changes to the code repository Elastic team reviews the changes before merging the changed code in the release branch. Developers must adhere to the standards Elastic team published on the GitHub repository. Elasticsearch open-source code Repository Kibana open-source code Repository
Community and Support	Since ELK stack is used worldwide, there are many forums, user discussions and blogs. Following are the major web sites for use forums, discussions, and blogs: Elastic Discussion web site Elastic Community on Slack
Security Vulnerabilities	Elastic stack software does not have known security vulnerabilities. Security vulnerabilities also depend on infrastructure security. In the self-managed environment, the team must ensure infrastructure security. In the Elastic managed environment,
External Dependencies	Self-managed by ELK stack has dependencies on cloud or on-premises infrastructure. ELK stack can also be hosted in the containerized VM (Docker). The software dependencies are included in the installation.
Quality of Documentation	Elastic has documentation, training, and certification courses. Elastic Documentation
Ease of Integration	Elastic has REST APIs for Elasticsearch and Kibana for integration. Kibana REST APIs Elasticsearch APIs
Cost of Ownership	Self-managed Elastic stack software is open source and free. For the self-managed Elastic stack, following factors incur cost: Infrastructure – On premises or Cloud Human resources for installation and configuration

	Human resources for support and maintenance
IP and Patent Risk	Self-managed Elastic stack is under an open-source license. The managing organization has access to the code and can make modifications to it. Elastic open source is licensed under open-source licenses as mentioned in the license compliance section. In the subscription model, organizations don't have access to the code.
Sustainability	The Elasticsearch open-source product is available on a public git repository, where Elastic team and community members make continuous updates. There are no known geographical restrictions.
Data Segregation	Elastic stack stores the data. Open-source (self-managed) Elastic stack is hosted in organization infrastructure.
Cost Estimate (approx.)	These are the costs associated with hosting: VM - $100 to $200 per month DNS - $50 per month
Time Estimate (approx.)	36hrs to 40 hrs

Pulse by NABP

Designed to support DSCSA traceability requirements, Pulse is the single source of truth for Trading Partners to maintain trading partner information including location (GLN), validated addresses and points of contact. Ability to retrieve information such as location validated address, point of contacts and GLN data via real time interface calls (Application Programming Interfaces / API's to retrieve information). It is a self-service application enabling vendors to create their profiles by adding data or update their existing data if any changes (i.e., changing contact personnel, etc.) Pulse provides access to Trading Partner data in pharmaceutical supply chain. Pulse has manufacturers, distributors, 3PLs, wholesalers and dispensers. Pulse has data such as names, addresses, contacts along with advance data such as GLNs, licensing data and ATP status data.

Role of Pulse on TIOP:

Pulse will not be used as a solution on the TIOP MVP due to its proprietary nature (see Analysis and Path Forward below). An open-source tool with similar capabilities, if there was one, would have been utilized for GLN validation, contact validation, and address validation. TIOP MVP will build this capability within TIOP and work with its trading partners to ensure accuracy and efficiency and where needed, validate this information with GS1 Member Organizations in the absence of a source of truth for trading partner information.

Pulse Business Strategy

Pulse by NABP is owned by a non-profit, whose mission is to protect public health and to enhance patient safety. NABP's mission and non-profit structure requires the Association to put mission over profits, which drives lower fees for the services offered. Economic viability is important and Pulse's business plan allows it to be a sustainable ongoing entity. Specifically, Pulse does not rely upon donations or grants. Domestically, Pulse's revenue model is a combination of free services to support NABP's public health mission as well as fee-based services to support the Pulse platform long term. For example, all trading partners can claim their Pulse profile, update their contact information, look up other trading partners and initiate product verification and product tracing requests at no cost. NABP is enabling these tools at no cost to remove barriers from participation—for the small dispenser community. Put more simply, the "directory/phone book" components of Pulse are free. The fee-based services in Pulse are applied if a trading partner wants to verify the license or authorized trading partner status of their fellow trading partners (dispensers,

distributors, manufacturer,3Pls.) The price domestically for these services are priced according to similarly marketed offerings in the US. As Pulse deploys beyond DSCSA, any associated services will be priced consistent with Pulse's nonprofit status and mission, as well as the market affordability per region. The existing AMS support model of Pulse will provide support for the trading partner directory and will not have additional expenses for USAID.

Pulse SWOT Analysis for TIOP

Strengths	Weakness
– Single source of truth for Trading Partners to maintain key information including location (GLN), valid business addresses and points of contact. – Comprehensive data access: Accessible interfaces for retrieving various trading partner information, including contacts and validated GLN hierarchy. – Reduced development and maintenance costs of trading partner directory: Avoids the overhead of creating and maintaining an internal data store of trading partner directory such as contact information, validated addresses and GLNs, optimizing resource allocation. – Self service capability enabling vendors to update their essential data (i.e., changing contact personnel, etc.) – Access to Pulse's user interface (GUI) for entering or creating trading partner information. – Access to Pulse's Trading Partner (TP) directory for storing information. – Access to Pulse's interfaces (API's) and user interface (GUI) for retrieving trading partner information. – Pulse has a mechanism in place to obtain data from GS1, and source authorities to periodically (weekly, bi-weekly) retrieve and update Pulse database. (Specific for GLN data) – Service Level Objectives (SLO's): 95% of interface calls have a <1 second response. 95% system availability for GUI users.	– Proprietary solution where code is owned by NABP. Although Pulse is being made available to USAID for the duration of the TIOP MVP, it is unclear of the costs of participation beyond the MVP. – Absence of Global Trade Item Number (GTIN) data – Lack of Marketing Authorization Holder (MAH) numbers and associated country codes linked to GTINs
Opportunities	**Threats**
– Access to other Trading Partner information which Pulse is already storing beyond the USAID vendors, should the pilot expand. – Inform TPD model for the global health community. – Creating international TP profiles in Pulse. NABP would add up to 30 trading partners USAID identifies (USAID to provide trading partner data) – NABP to support 1,000 queries / month of trading partner information. – USAID may have 10 named users to access / update trading partner information (GUI access)	– Sustaining MVP for USAID markets beyond TIOP MVP are not defined including cost implication. – Inability to validate international addresses. – Pulse's scalability to accommodate international vendors would mean significantly larger user base, increased data traffic, and possibly diverse geographical locations. This can strain the infrastructure, leading to slower performance, downtime, or even system failures if not properly managed. – Need to review country data agreements and align Pulse to requirements beyond the existing data protection laws, if applicable.

Analysis and Path Forward

Fundamental to the MVP is that all TIOP artifacts will be made public and/or used as a global good to support a competitive market for global and/or country-level solutions in the future. Pulse does not

offer a pathway to this requirement. It is recommended to build the trading partner directory within TIOP and not integrate it with the Pulse solution. Additionally, the lead time to working legal teams from host national governments on data access governance provision may impact TIOP MVP schedule for delivery. NABP has ambitions to scale Pulse beyond DSCSA requirement and has been engaging with other organizations in international use cases. As those are implemented, and NABP considers its future stewardship and openness; it is recommended for the global community through forums such as the Verification and Traceability Initiative to consider Pulse or a similar solution that offers up-to-date, well-maintained master data describing locations. The global health community needs to discuss how to close the gap of a well-functioning repository of GLN information that can be queried through direct party name searching. GLN information displayed in real time and is external-facing, open to all global registry users. However, to ensure that systems stay up to date, it is imperative that functionality be developed to push or regularly query updates for systems using GLNs through direct integration much like Pulse. Further, there is significant momentum across the global health community in driving the implementation of GS1 global standards for health care, seeking benefits in advancing end-to-end data visibility, improving supply chain security, gaining supply chain efficiencies, and ensuring patient safety. Leveraging GLN as a standard for identification of locations and entities is an important piece of this puzzle, particularly as advancements are made in transaction and event data exchange. However, implementation can be complex and costly, and it must be done with a focus on data quality, master data management, and data access to realize the anticipated benefits. GLN adoption is not something that can be planned in silos, but rather requires coordination and intentional consideration. With the appropriately tailored vision and strategy, system architecture and policy frameworks, and standards champions in place, GLN adoption can greatly strengthen data quality and communication within the global health ecosystem. As a part of the TIOP MVP lessons learned dissemination plan, it is recommended to have discussions with the global health community on efforts to advance advocacy and solutions for a coordinated GLN initiatives and possibly a registry.

Proposed Engagement Plan for Global Health GLN Registry

Step I: Identify Stakeholders: Led by GHSC-PSM

GHSC-PSM will develop a matrix of key global health community partners that may impact or be impacted and inform pain points, lessons, and use of a global GLN registry. Illustratively, these partners may include the following:

Africa-based GS1 Member Organizations (MO): As momentum grows toward adoption of GS1 standards for health care generally, and specifically in sub-Saharan Africa, the membership within regional GS1 MOs such as GS1 South Africa, GS1 Nigeria and GS1 Ghana is expected to continue to grow in the health care sector. These GS1 MOs to potentially offer services like GLN allocation and training services at a sustainably lower cost.

Partner Governments: As regulations are passed and national traceability repositories are deployed in an increasing number of countries, the drive towards standardized data exchange (e.g., EPCIS) necessitates the need to validate GLN for global entities and communicate to trading partners accurate and up to date GLN information about its locations and legal entities.

Traceability and Verification System (TRVST): The TRVST solution has been deployed to fully comply with the EPCIS standard to support parent/child hierarchies between locations. TRVST must have the ability to validate the company prefix against configured list of recognized company prefixes and would have the ability to validate GLNs contained in the event data when data is received in TRVST.

NABP: As the only Trading Partner Directory to support regulatory requirements, NABP can inform the building blocks for standing up a GLN registry including required governance, data access and privacy, validations, and data structures among other lessons.

International procurement agents (IPAs): IPAs require unique identifiers and accurate master data for pick-up and ship-to locations to accurately execute and have visibility into end-to-end supply chain processes. As IPAs implement electronic transaction data exchange mechanisms (e.g., EDI), the ability to leverage a GLN as the unique identifier across trading partners creates efficiency in the process.

Donors: Serve as catalyst for GLN allocation, promotion of a coordinated, non-duplicative solution to manage and share location data. Promote tools that ensure countries and IPAs access to GLN registry tools and are held accountable for the provision and maintenance of accurate information.

GS1 Global: Promote global health use cases and pain-points in discussions around GLN global strategies.

Step 2: Convene Stakeholders through a workshop to determine the following: (Led by GHSC-PSM)

- Define GLN Registry use cases for the global health community.
- Identify pain-points resulting from a lack of GLN registry for global health.
- Agree on the appropriate forum to advance the discussions.

Step 3: Assess viability of a GLN registry, develop strategic and advocacy plan: Led by TBD (appropriate forum identified under Step 2)

Step 4: Execute strategic plan or advocacy: Led by TBD (appropriate forum identified under Step