GEOSPATIAL DATA SHARING PROTOCOL

FREQUENTLY ASKED QUESTIONS

Version: 23 October 2023

Context

1. What is a Data Exchange Protocol?

"An information exchange protocol is a set of rules for exchanging data between systems or programs. It describes how the data should be structured and formatted, as well as what types of messages can be sent and received. This type of protocol is used to ensure that different computer applications or devices understand each other when communicating over a network. Generally speaking, an information exchange protocol helps to facilitate efficient communication between two entities so that they are able to send and receive the same type of messages in order to achieve their desired outcome". (Post from Tech Titanz LLC, January 2020).

2. What is the Geospatial Data Sharing (GDS) Protocol?

The GDS Protocol serves as a voluntary standard aimed at enabling a consistent method of sharing geospatial data. It is designed to simplify the transfer of geospatial information and related data throughout value chains. The protocol sets forth key requirements, emphasizing attaching data to physical goods transactions, using distinct identifiers, data structuring, and deploying spatial indexing systems.

3. Why is a Protocol needed?

Recent private sector developments, including the EU Deforestation Regulation draft, have highlighted the need for streamlined methods to transfer geospatial information and related metadata through supply chains. Given the substantial data associated with individual products, a unified, efficient method became imperative. Our consultations with various certification schemes and major trading corporations indicated significant interest in a common approach that could satisfy not only EU regulation requirements but also enhance traceability for other commodities and supply chains.

4. Who are the intended users of the Protocol?

The Protocol appeals to a broad spectrum of organizations interested in sharing geospatial data relevant to harvesting sites for agricultural and forest commodities. This includes:

- Private companies with aim to trace materials to harvesting areas.
- Certification schemes and multi-stakeholder coalitions active in agroforestry commodities.
- Governmental traceability systems.
- IT service providers.
- Other key players like academic or research institutions.

Adoption of this Protocol might stem from regulatory requirements, like the EU Deforestation Regulation (*Regulation (EU) 2023/1115 of the European Parliament and of the Council of 31 May 2023 on the making available on the Union market and the export from the Union of certain commodities and products associated with deforestation and forest degradation*), or global regulatory/policy initiatives. The Protocol, being voluntary, doesn't have mandated use. However,

we advocate for a shared commitment across supplier networks to standardize practices. Digital architects, IT developers and data managers are expected to be primary implementers and drivers.

5. How will the Protocol work for less developed companies (e.g. SMEs) in addition to large multinationals?

The Protocol was designed with the distinct needs of smaller enterprises in mind. By standardizing the sharing of geospatial data and metadata, it provides a streamlined approach for small businesses to efficiently share data, such as harvest site polygons, with various companies, thereby conserving both time and resources. Furthermore, it is anticipated that tools tailored for smallholders, like phone apps that align with the Protocol, will soon emerge. Such tools will offer an efficient means for smallholders to adhere to the Protocol's requirements, ensuring inclusivity across all business scales.

6. Will the Protocol be useful for complying with the EU Deforestation Regulation?

Absolutely. The Protocol is designed to accommodate the EUDR's requirements concerning the geolocation of harvesting plots. Although the regulation itself does not mandate a specific method for data transmission, the responsibility for selecting such a method lies with suppliers and their customers. It is crucial to note that while the Protocol addresses certain aspects of the EUDR, solely relying on it doesn't guarantee full compliance. Organizations must also ensure they meet other EUDR requirements, such as those pertaining to zero deforestation and legality.

Choosing to use the Protocol is entirely voluntary, and its application is up to individual actors within the supply chain. Importantly, the Protocol's versatility makes it apt for any scenario requiring geospatial data sharing, not just those outlined by the EUDR.

7. In which other scenarios might it be useful?

The Protocol is pivotal wherever there is a need for geolocation data within value chains. While it's certainly useful for deforestation and forest degradation assessments, its applicability extends to scenarios where organizations further down the supply chain undertake due diligence procedures. This could be to ensure sourcing from forests and farms that meet sustainability criteria. Given the complexities of global supply chains, there's a growing demand for solutions grounded in geolocation data to accurately trace the origins of raw materials. The Protocol's primary goal is to standardize this data exchange, making it universally adaptable whether it's between companies, within certification systems, or any other scenario that calls for geolocation data sharing.

8. Has the inclusion of data for calculating scope 3 emissions been considered?

Numerous stakeholders have highlighted the potential value of incorporating emission data, especially for companies further down the supply chain aiming to estimate their scope 3 emissions as part of their greenhouse gas assessments. While this data hasn't been integrated into the initial version of the Protocol, its inclusion is under consideration for future iterations.

Conditions of use of the Protocol

9. Are there licensing or fees associated with using the protocol?

No, there are no licensing or other fees for using the protocol. It is available as open-source and can be freely used by anyone who finds it useful. However, it is important to note that organizations may incur costs related to updating their software systems. These updates will enable them to receive, process, and transmit data in accordance with the protocol, facilitating the flow of information to the next entity in the supply chain.

10. Will Preferred by Nature develop software to facilitate companies' use of the protocol?

While Preferred by Nature will not be crafting specific software solutions for the Protocol's implementation, given that our core expertise is not in technology but in supply chain and sustainability, we do acknowledge the significance of efficient data management for the protocol's success. Consequently, we are considering the creation of Application Programming Interfaces (APIs) or other supporting tools to aid organizations in seamlessly implementing the Protocol and optimizing their geospatial data handling.

11. Will geospatial data transmitted in accordance with the Protocol be centralised in a specific database?

No. The protocol is meant to be adjusted to a range of existing systems developed and used by the industry, certification schemes, services providers, etc. It was developed with the understanding that stakeholders will be using multiple information systems, and that supply chain data, including geospatial data, will need to pass between those various systems accurately.

12. Will there be a tool to enable the overlay of geospatial information, shared in line with this protocol, onto maps of deforestation and forest degradation?

The GSD Protocol itself does not encompass this function. Numerous existing tools can assess deforestation based on the provided geospatial data, and it's anticipated that more will emerge in the future. The Protocol remains valuable irrespective of the deforestation assessment tools stakeholders choose. Preferred by Nature advocates for service providers to adapt their current tools for seamless integration with data that complies with the Protocol.

Formats chosen

13. Why is the Protocol not using the GeoJSON format instead of the JSON format?

JSON is a general-purpose data interchange format, while GeoJSON is a specialized format for representing geographic information. GeoJSON extends JSON by adding support for geographic features and geometries. JSON is widely used in various applications for data exchange, while GeoJSON is specifically used in mapping and spatial applications. GeoJSON Specification (RFC 7946) only supports WGS 84. Spatial Indexing Systems can be used with JSON by storing the cell IDs and any associated data as JSON objects. GeoJSON is therefore not necessarily needed for this; JSON can be used to structure the spatial index data and any associated metadata or attributes. This provides more flexibility.

14. Why was spatial indexing chosen as a requirement and what is its added value?

A spatial indexing system typically partitions the space into smaller, manageable units and assigns a unique identifier to each unit. Spatial indices are data structures or techniques designed to organize and optimize the storage and retrieval of geometries in a spatial database.

Index technology for geospatial data offers advantages in terms of efficiency, scalability, uniformity, and integration with modern data systems, especially when dealing with large-scale and dynamic spatial datasets.

- It minimizes the size of data files as it converts Geospatial Data to numeric values.
- It is a good way to fix many issues associated with projection, as it passively enforces data quality issues (e.g., having a geometry that is not a closed polygon).
- It can be particularly useful to allow anonymising farmers' identity for instance.
- It makes it very easy to make polygon comparisons, like determining overlapping between polygons.
- It facilitates making operational search (inclusion, collision, overlap), which are much slower using pure GIS information
- It also works well for organisations with scarce GIS capabilities.

There are two main spatial indexing technologies currently available: S2 grid, developed by Google and relying on square geometry, and H3 grid, developed by Uber, and relying on hexagons geometry. Both work well. Both have a grid and a hashing machine system, which will save points or polygons as strings of characters.

Common geometries (e.g. KMZ, shp…) can be converted to S2 or H3 strings. APIs are available for both to make this conversion. There could be a need for some additional development to include converters to platforms and open-source tools (e.g., dedicated apps, QGIS / ArcGIS, etc.), but this is not a big issue.

AgStack is developing an open-source code base, alongside a fully automated, continuous computation engine, to create, maintain and host a global dataset of boundaries' "registry" for agricultural fields using S2 to aid in such things as food traceability, carbon tracking, crop production, and other field-level analytics. It is expected that this asset registry will assist with using this Protocol.

15. Why does the protocol offer both S2 and H3 formats? Wouldn't it be simpler if only one format was utilized?

During the development of the protocol, consensus could not be reached among the stakeholders and actors regarding a single preferred format. Each format, S2 and H3, has its distinct advantages and limitations. While Preferred by Nature recognizes the potential benefits of standardizing to one format, we also value the diversity of needs within our stakeholder community. As a result, we have decided to include both formats for now. We are planning additional consultations to determine the most suitable format for widespread adoption in the future.

16. Does the "envelope" always relate to a single "location" (point or polygon) or does it allow to communicate multiple locations at once?

Data package ("envelopes") compliant to the protocol specification may indeed contain a list of multiple locations, whether points and polygons.

17. Is it the case that the same polygons will have to appear in different "envelops" multiple times?

Yes, it is likely that the same geolocation will appear in different data package ("envelopes"), as the raw material originating from this location will be split into different products during processing

and over time. With this protocol, the transmission of geospatial information is tied to transactions of products in scope.

Other clarifications

18. Will supply chain information (i.e. the succession of trading entities) be shared through the use of the protocol?

The GSD Protocol facilitates the sharing of product-related geospatial data and metadata through a secure "envelope" system. This system may support maintaining information on who has traded products in scope, and passed on associated geospatial data, but it also allows the receiving entities to extract the geospatial data and transfer it to another entity within a new "envelope", effectively omitting confidential supply chain information. The terms of data sharing are to be determined by the trading parties involved.

19. Does the protocol account for complex derivatives that contain materials from multiple commodities or multiple ingredients made from the same commodity?

Yes, the protocol is designed to support the association of geospatial data and metadata with various raw materials, even when these materials are combined. This metadata can specify the type of commodity, facilitating the integration of different materials.

20. In the context of the EUDR, if the metadata includes harvest time and species, will it be enough to fulfil the due diligence statement requirement of the Regulation?

The protocol is aiming at facilitating the information collection requirement of the EU Regulation. However, due diligence aspects, such as assessing the legality of production and the absence of deforestation, are not in scope and need to be implemented based on the geospatial information collected.

21. Will there be training sessions or webinars to help stakeholders implement and understand the Protocol?

Yes, Preferred by Nature is committed to ensuring a smooth adoption of the GSD Protocol. We will be organizing periodic webinars, and training sessions. Additionally, a set of online resources, this FAQ, and guidelines will be available on our website to assist stakeholders throughout their implementation journey.

22. Is there a mechanism for stakeholders to provide feedback or report issues with the Protocol?

Absolutely. Feedback is essential for the continuous improvement of the Protocol. We have established a dedicated channel for stakeholders to submit feedback, report challenges, or suggest enhancements. We encourage active participation and collaboration to ensure the Protocol meets the evolving needs of all involved parties. Please contact Chloé Viala: cviala@preferredbynature.org. We look forward to hearing from you.

23. What is the planned timeline for the next updates or iterations of the Protocol?

We anticipate periodic reviews of the GSD Protocol based on stakeholder feedback and evolving industry needs. While we don't have a fixed timeline for the next version. Any significant changes or updates will be communicated well in advance to ensure stakeholders have ample time to adapt.

24. Have there been any early adopters or pilot programs for the Protocol? What was their experience?

We have expecting to have a few organizations pilot the GSD Protocol, and share their insights. We'll be featuring case studies and testimonials on our website, showcasing the real-world benefits and challenges encountered during implementation.

25. Is the Protocol applicable to cattle?

Yes, the Protocol can very well be applied to cattle. However, Protocol users must agree first on which geolocation data must be collected and shared through the value chain. Indeed, there might be several successive locations of interest associated with a single cow (rearing, breading & fattening ranches). Which information is necessary to compile and who is responsible for generating this information shall be clearly established before using the Protocol (also see question 26 below).

Out of scope elements

26. Is the Protocol covering when the initial collection of geospatial data must happen?

No, this is not covered by the Protocol, because there are too many different possible scenario. This must be agreed directly between Protocol users, depending on their supply chain details and traceability objectives.

In the context of the EUDR, the need for geospatial data applies to primary goods grown or raised on land that may have caused deforestation and this creates two classes of goods. For instance, a processed table or roasted coffee were not directly harvested as such from a plot of land on the surface of the earth. Trees and fresh coffee beans have been harvested directly from a plot of land and are the goods associated with specific intrinsic geospatial data.

Cattle will be a specific scenario where the same cows may be moved before a "harvest" event happens (as defined in the EUDR), e.g. there might be several successive points or polygons associated with a single cow (rearing, breading & fattening ranches), which is not the case for cocoa, grown on a single plot of land.

Another specific scenario may be where an entity is both harvesting and sourcing from external entities: e.g., a coffee roaster has its own coffee plots of land and is also sourcing from other farmers. In this instance, it will be necessary to both generate geospatial data and request / receive it from the suppliers.

Outside of EUDR compliance, this Protocol may also be used in such a way that geospatial data are added by successive supply chain entities (for instance: polygons for cotton farms, then locations of cotton ginners, spinners, weavers, etc.).

27. What is the connection between the GDSP and the ISEAL's Guidance on the collection of polygon location data?

As the collection or generation of geospatial information is out of scope of the GDS Protocol, there is no overlap between this and ISEAL guidance. They are on the other hand very complementary, and we do encourage stakeholders to consult and use both.

28. Given that data is passing from one actor to another via use of the protocol, are there any data security and integrity measures built into the protocol?

This first version of the protocol does not include focus on information security. It is anticipated that involved parties will establish agreements on secure data transmission methods and scrutinize the contents of data packets. Subsequent versions may concentrate on incorporating data security and integrity features.

29. Why is data validation out of scope of the Protocol and how do we address possible problems potentially coming from data collected with insufficient quality?

Although it is well understood that data quality is important and the "garbage in garbage out" scenario must be addressed, data quality and validating data were not in scope for this initial phase of the project. It was considered that additional due diligence is always needed to make sense of the geospatial data obtained. If one fails to associate the correct points or polygons (e.g., associating mills' geospatial information instead of farms' locations), then this is a due diligence issue, not a Protocol compliance issue.

Data validation may be considered in the further stages and the Protocol adjusted accordingly.

30. Is the Protocol ensuring alignment with Data Protection Laws?

Aside of the minimal metadata described in its Annex 1, the GSD Protocol is not prescriptive in how geospatial data should be collected. It does not make it mandatory to maintain information about the supplier network trading the goods in scope (e.g. farmers' identities and personal data). Protocol users must agree between themselves which additional metadata they will communicate to each other and must consider Data Protection Laws.