Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Forest Damage Center
Henrik Persson and Eva Lindberg

REPORT

SLU ID: SLU.sfak.2021.1.1.1-226

11/26/2021

Forest Damage Database – a possibility for remote sensing applications

Summary

This report concerns the C-project "Establishing of digitized reference database about forest damages" within the SLU Forest Damage Center. We have investigated the conditions for establishing a database of forest damages registered in the field that would hence be suitable as references for remote sensing applications. The purpose of a database is to increase the use of remote sensing to identify and monitor forest damages faster, more accurately and at larger scale. Numerous current inventory programs have been identified as well as other useful data sources, and a number of benefits and disadvantages with these have been noticed. For research purposes, the data quality is most important, while the operational forestry would benefit from harmonization of reference data. We have concluded four viable options, which require different levels of resources. These are described in the sections "Options", "Implementation" and "Required resources". This first version of the report requires feedback from the SLU Forest Damage Center, in order to be finalized during the first quarter 2022.

Mailing address: 901 83 UMEÅ
Street address: Skogsmarksgränd 1
VAT no: SE202100281701
www.slu.se

Phone: +46 18 67 10 00 (switchboard)
firstname.surname@slu.se

Contents

Introduction

Forest damages

Damage to forests cost Swedish forest owners billions of SEK annually. Research has indicated that the damages will increase in the future, partly due to global warming. The damages in the past years, caused by e.g., storms, fires and spruce bark beetles, show that the forestry sector needs better tools to prevent, monitor and combat such threats. This requires methods to identify ongoing damages and localize areas with increased risk.

Forest remote sensing

Remote sensing is an efficient tool to inventory and map forest damages. Many platforms can carry remote sensing sensors. Drones, helicopters, and airplanes are feasible to acquire high-resolution data, e.g., laser scanning and aerial photos. Satellites, e.g., Sentinel-1/2 provide data of low to intermediate resolution spatially (10-40 m), and at a high temporal resolution. These enable fast and easy detection of substantial changes in the forest and numeral successful studies have been published. Successful solutions based on remote sensing are mostly limited by access to accurate reference data to create models for wall-to-wall predictions. Field inventories and harvester data constitute important reference data. Often, statistical

models are most feasible, which require a substantial sample size, and therefore, become expensive and prohibit detection of rare damage types. Machine-learning based methods and in particular deep learning require even larger amounts of annotated reference data. The use of harvester data could potentially provide sufficient continuous dataflows to enable development of the latter methods. Coordination of existing datasets and resources, and collaboration with other actors may be another solution to provide enough reference data.

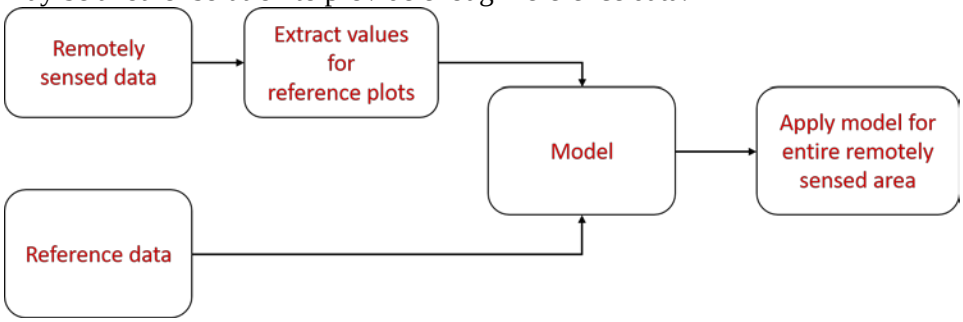


Fig 1. Schematic illustration of conventional remote sensing application.

Applications and stakeholders

Different stakeholders have different preferences and needs for a collection of reference data.

The Swedish University of Agricultural Sciences (SLU) aims to perform research and develop methods for the analysis of forest damages with support of remote sensing, which requires high quality reference data. The research is more dependent on data quality and benefits less from operational platforms. The Swedish Forest Agency (SFA) aims to provide end users with maps and decision support at large scale, which require reference data from large areas, a high degree of automation, and for AI solutions - requiring large amounts of reference data.

Private actors including forest companies and private forest owners are primarily interested in the applications built upon the database. They do not always have resources for performing their own analyses. However, many consultant companies and some forest companies want to address this by developing their own operational solutions. Finally, private persons are likely most interested in free, publicly available maps and solutions.

Purpose

This has formed the need of establishing a database to collect and coordinate information related to forest damages and remote sensing. Currently, forest damages are inventoried in various national programs.

The current report covers the most important aspects of establishing such a database and estimations of the pros and cons of it. Within the Forest Damage Centre, other projects are working on establishing databases with forest damages for other purposes, but the coordination of these is outside the scope of this report.

Objectives

Describe how to develop a database for damage-related reference data for remote sensing applications to predict risk and monitoring of forest damage. It can include all types of damage, e.g., insects, fungi, bacteria, viruses, browsing, drought and fires, storms, snow. This report addresses the question of how such a database could

be established and provides additional information about required funding and its impact on the outcomes.

Existing data on forest damages

SLU

SLU coordinates a number of regular forest inventories that include forest damages.
SLU Skogsskada, “Forest damage”, is a tool for the public for identifying, reporting and providing information about forest damage online. The aim of Skogsskada is to collect data on where and how forest damages occur aiming to better understand how they might be mitigated. In this database and web-based tool, you can both report and identify damages on trees caused by pests, pathogens, game and abiotic agents. Data can be retrieved in tables, in data sheets or on maps. Skogsskada does not provide systematic inventories but relies on voluntary reports.
SLU National Forest Inventory (NFI) provides the official statistics about the status and change in Sweden's forests. It is primarily carried out through inventory on temporary and permanent plots, where the positions of the permanent plots are kept secret. The inventory includes forest damages at the stand level and on single sampled test trees. At stand-level, the total degree of damage and dominating type of damage during the past 5 years are registered. At plot level, the inventory assesses damages from climate, anthropogenic impact, vertebrates, insects, fungi, and fire, damages from grazing moose in young forest as well as damages on roots, cambium, stems and crowns on sample trees. About 11,000 plots are inventoried across Sweden annually and about 10% of these contain forest damages. However, uncommon damages in particular are rarely located within the plots, and the plots are therefore often impractical as references for remote sensing applications.
Nationell Riktad Skogsskadeinventering (NRS) is an inventory concept for specific forest damages at the regional level, led by SLU. They provide “target-tailored forest damage inventories”, carried out in a flexible manner “when needed”. This enables SLU to present annual data for operational decision making at local level and linked to specific forest damage outbreaks. The plot locations are secret. NRS has e.g., inventoried spruce bark beetles, pine blister rust, and ash dieback.
Nationellt skogsdatalabb, “Forest datalab”, is a national resource coordinated by SFA and SLU, which facilitates access to forest data.
Additionally, a number of research projects collect data on forest damages. This data could be of potential interest as reference data for remote sensing.

Forest Agency

The Swedish Forest Agency (SFA) inventories various forest damages, e.g., the browsing program ÄBIN, monitoring of spruce bark beetle swarming.
Äbin is an inventory for damages caused by large herbivores primarily by moose in younger forest of about 1 to 4 m average height. It is, e.g. used, to facilitate the regulation of the damaging animal populations, by collaborating with hunters and forest owners. It also includes an adapted inventory for forest 0.1 to 1.5 m height, addressing deer on the island Gotland, where no moose are living.
Monitoring of spruce bark beetle swarming is done using bark beetle traps that are emptied once per week from the end of April until the beginning of September. This data is available on the SFAs public statistics database.

Furthermore, SFA monitors biodiversity with the purpose to inventory the biodiversity in forest with high conservation value since 2009 and monitor changes until 2021.

Data from external partners

The Swedish forest companies register forest damages during harvest (registered by the harvester operator) and through directed field inventories. The damages registered by the harvester include attacks by spruce bark beetles and harvester data constitute a resource for collecting larger amounts of data than field inventories, but are on the other hand currently only covering a few damage types. They are furthermore not planned with a statistically unbiased design, which hampers some scientific perspectives. However, they are one possible source for model-based applications, which in particular can provide sufficient amounts of data for deep-learning based models.

The company Biometria and research institute Skogforsk are key partners in the Swedish context of harvester data.

Currently, most forest companies in Sweden upload their harvester data to Biometria, which is a common organization owned by the forest companies, who coordinates harvester data in Sweden. They provide independent measurements of the harvested timber. Although Biometria is managing the data, they do not own the data.

All harvesters use a format called StanFord2010, which has been developed by Skogforsk. The files contain fields that can be used to separate any damage types, although it has not been actively updated since 2011.

Forest companies could contribute with data given the right conditions, which likely vary with each company. They are generally reluctant to share complete datasets, but they could be more likely to share selected parts, e.g., harvester data related to damages. A motivation for the forest companies could be the large overall benefits for the Swedish forest sector. Sharing data could also be beneficial for the companies if it results in applications that they can use. We have contacted a few larger companies to get an impression of their interest.

Sveaskog

As state owned forest company, Sveaskog would be suitable for pilot studies. They have a database with harvester data, including bark beetle damaged trees, which however is not stored by Biometria. Sveaskog is willing to contribute with harvester damage data from their own forest, if it does not harm their business model. Therefore, they do not foresee sharing of data from healthy forest, or complete stands. They are willing to upload data regularly (exported automatically from their existing database solution), or only register meta data. They would, however, prefer uploading data, since it reduces their workload. They can also do minor conversions of data, but no substantial development. Sveaskog does also have various additional inventory data that may be shared. They register subjective damage information on the management units (stands) too, but this is described as % damaged area. Sveaskog is aware of the right of public access to data for all Swedish agencies (see next section), including SLU, but may be more willing to share some types of data if this can be controlled.

Holmen Skog

Holmen Skog is positive to contribute given the right conditions. Benefits of contributing include, e.g., the access to more data which requires other actors to also contribute. They see a need of transparency to progress as company and sector. They would prefer Biometria as the natural data center of information. Other partners

could develop tools and software from the data. Holmen also conducts field inventories about damages in young forests.

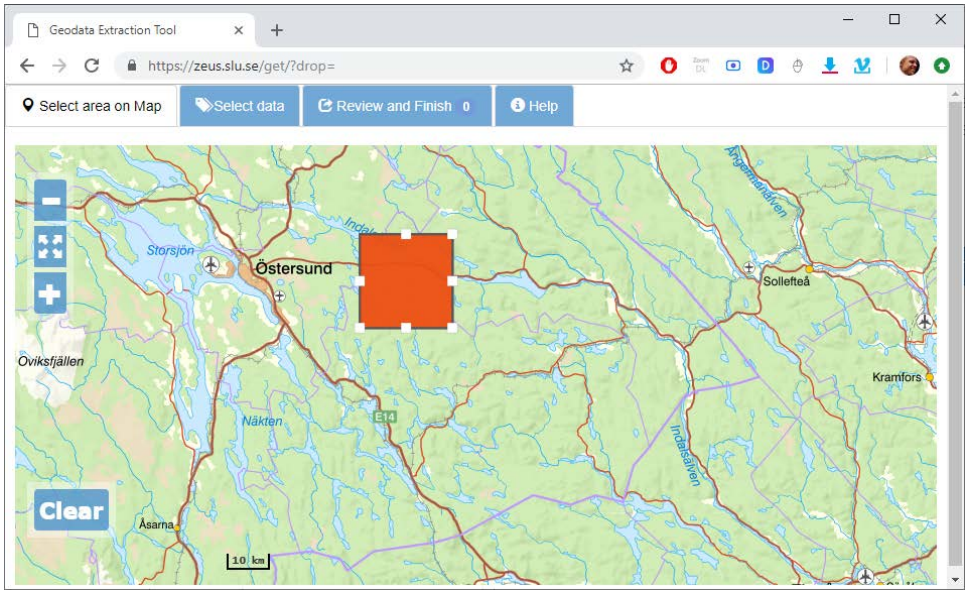
Data ownership and copyright

The development of a reference database involves various challenges and partners. If all partners of the database aim for an open-access solution, there are no problems with copyright. If some partners prefer protecting some of the data, there are certain regulations that can be used to limit this. Some partners or some subsets of the data may require permission to be used. The need for temporal confidentiality may be used for subsets of the data until the intended research has first been published.
Public agencies (SLU and SFA) are in general obliged to share information on request due to the "public access to information and secrecy", "offentlighetsprincipen" in the Swedish law. Any data handled by public partners and agencies are per definition possible to request according to the "public access to information". On request, the agency should provide the data non-discriminatory and at a reasonable cost. If someone shares data with a public agency, those data are condition to this law. The organizations that provide data could prefer to limit access to the data to certain users and for certain purposes. Private companies may therefore be reluctant to share data publicly.
If the data sharing can be part of contract research, "samverkansforskning", the possibilities for confidentiality are strong, up to 10 years. Yet, this also requires some intellectual contribution of the data provider, in addition to pure data sharing. There is also an EU legislation "database protection", which provides up to 15 years protection of the database content if someone requests the full database. This law can be applied when substantial resources have been invested in the database.
Copyright aspects have an impact on the alternatives of storing data centrally or distributed. A database which only contains metadata (content, availability, and quality of data) could be centrally managed and would overcome the copyright hinders, while a common database storing geodata would need to consider copyright issues.

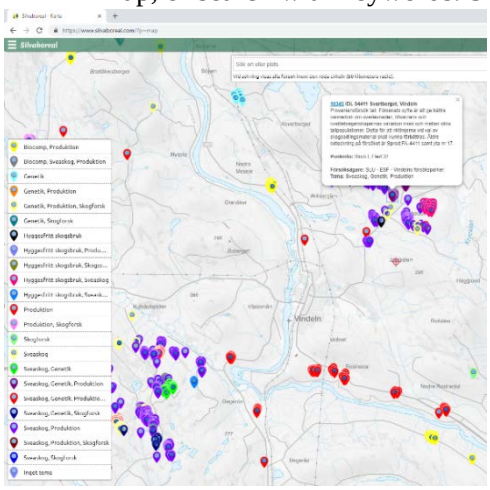
Some existing infrastructure for geodata

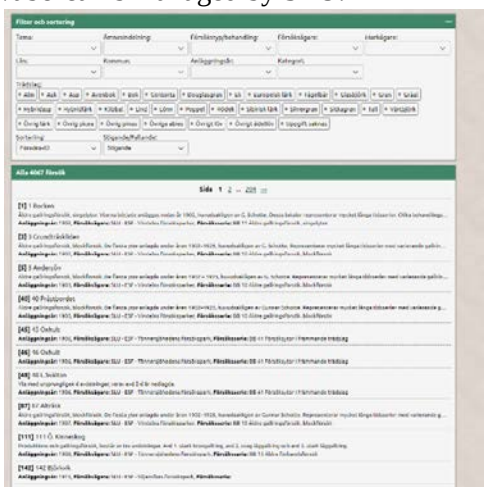
Examples on some relevant existing infrastructures that can act as role models and provide ideas for a solution of the current project.

- Geodata Extraction Tool GET (Geodata Extraction Tool) is a web service for Swedish universities providing geodata (digital maps) from the following agencies' databases: Lantmäteriet (the Swedish mapping, cadastral and land registration authority), Statistics Sweden, the Geological Survey of Sweden, and the Swedish Maritime Administration. The user can select a geographical area of interest and choose geodata to download from a list. GET is managed by SLU.

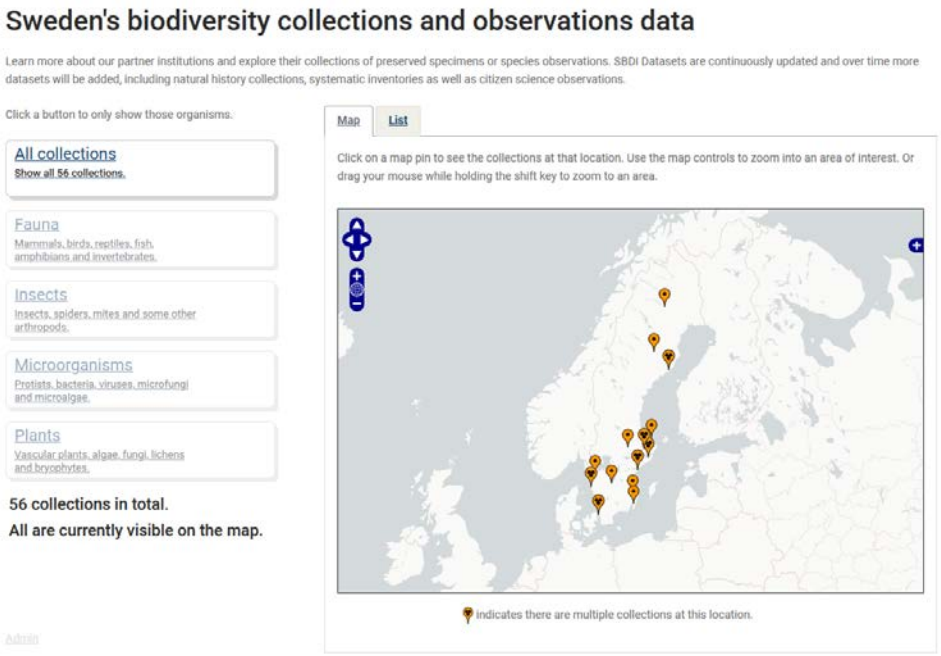


- Silvaboreal has become a national meta database for forest experiments. The user can select a geographical area of interest to view the experiments on a map, or search with keywords. Silvaboreal is managed by SLU.

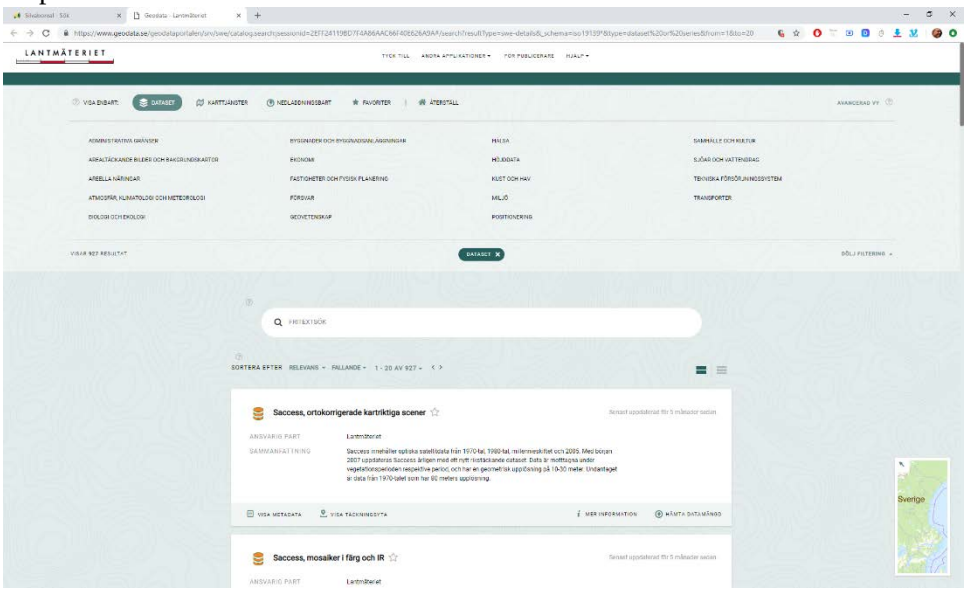




- Swedish Biodiversity Data Infrastructure (SBDI) The overall vision of SBDI is to provide Swedish researchers with unified and open access to biodiversity data, and to tools for querying, visualizing and analyzing this data. SBDI data comes from sources such as monitoring programs, citizen science portals, genetic analyses of environmental samples, natural history collections and research projects. The user can select a geographical area of interest or select collections of data from a list. SBDI is managed as a national collaboration between universities and other organizations handling biodiversity and ecosystem data.



- Geodataportalen is a portal for geodata and services in Sweden. It provides opportunities for searching, finding and look at data from different sources. The data itself is physically located at different providers. The portal facilitates access to data and services by many organizations. Geodataportalen is the Swedish connection to the European geodata portal, Inspire.



- Biometria performs independent measurements of harvest flowing between the forest and the industry. They are owned by the Swedish forest owners and support timber trade, logistic and production on the timer market.

Options

Based on the existing and potential data, and the identified needs in question, we have identified a few alternative options. They will need various levels of resources to be established and maintained, both from SLU and the contributors.
The metadata require a strict format to enable automatic searches. Requiring a strict format of the actual data would require more resources for pre-processing of the contributed data. This would easily allocate one or several full-time dedicated persons but would however be beneficial for AI solutions in need of large data amounts. A more flexible format (less requirements – anything goes) puts higher requirements on the users.
Only the fourth proposed option includes harmonizing of the data, which includes adapting the format and preparing the data for modeling purposes.

Option 1

The first option would start small and grow organically with needs and funding. It would originate in existing projects and data from SLU and SFA, and limit access to the database to these actors (although data may be contributed also from others). The database would include both data and metadata and the contributors can choose if they provide only metadata or both.

Option 2

The second option would take a larger, more sector-wide approach, where metadata from various actors are collected and any actor can access the metadata. The metadata would contain references to the datasets stored by the data owners, e.g., SLU and forest companies, and the user of such a database would need to contact the respective data owner to get access to the data.

Option 3

The third option, which is more extensive compared to the first two options, would allow any actor to contribute and access the database. The database would include both data and metadata and the contributors can choose if they provide only metadata or both.

Option 4

The fourth, most comprehensive option would include everything in option 3, but also include harmonizing and pre-processing the data. This would facilitate machine-learning method development that would be beneficial for many stakeholders such as the SFA and private companies that develop operational solutions. However, this would require dedicated resources that manually handle this option. It would also be feasible to include method development together with Skogforsk and Biometria to make best use of the harvester data.

Implementation

The implementation of the various options would share some elements, while some differ and yet require more resources when shared to the public.
All options require storage of metadata and an interface to add and search the metadata. Existing databases with this functionality include, e.g., Geodataportalen from the Swedish Land Survey, and Silvaboreal from SLU.

The options 1, 3 and 4 are similar and require storage of the data, which significantly increases the need for hardware resources. Existing solutions with this type of functionality, trying to coordinate various types of data at the national level, include for example the Swedish Biodiversity Data Infrastructure and Geodata Extraction Tool from SLU.
The options 2, 3 and 4 which address the public require more advanced user management.
The access to exchange data with the database can be developed as an application programming interface (API) and is dependent on the database format. There is a need of standards, and in the next section we propose such a solution.
Additionally, a graphical user interface (GUI) for the users is needed and should use the API.
The practical implementation of all options will require hardware resources in terms of a server with suitable database software and a substantial effort for the implementation.

Database format

The database format needs we have identified can be summarized as follows.
Meta data

- Two levels of meta data required. The following general level should be required.
    - Type of damage,
    - type of data (text, binary),
    - geographic information (point in specified coordinate system),
    - resolution,
    - time and date for data collection,
    - field instructions as PDF,
    - status before/after,
    - published documentation,
    - includes non-damaged references (T/F),
    - copyright (possibly open after certain date),
    - acknowledgments,
    - quality (error frequency or similar),
    - unique key.
- A second project specific level with any type of meta data can be attached as binary data

Observations/data

- Type of damage,
- degree of damage,
- geographic information,
- time,
- unique key

Required resources

The following resources are estimates of the required resources for establishing the database. If an entirely new platform is to be developed, it will need more resources, while if an existing platform (e.g., SilvaBoreal) could be cloned and

adapted to this projects need, a substantial amount of time and resources would be saved.

Option 1

Hardware resources

- 1x server with lots of storage

Know-how

- User interface (> 6 months work)
- Database (> 6 months work)
- System administration of hardware (> 3 months work)

Option 2

Hardware resources

- None. The service can be shared with existing hardware.

Know-how

- User interface (> 6 months work)
- Database (> 6 months work)

Option 3

Hardware resources

- 1x server with lots of storage space (more than option 1)

Know-how

- User interface (> 12 months work)
- Database (> 12 months work)
- System administration of hardware (> 6 months work)
- API formalized to be used for more than just the user interface (> 6 months work)

Option 4

Hardware resources

- 1x server with lots of storage space (more than option 3)

Know-how

- User interface (> 12 months work)
- Back-end interface for pre-processing and harmonizing of data (> 3 months work)
- Database (> 12 months work)
- System administration of hardware (> 6 months work)
- API formalized to be used for more than just the user interface (> 6 months work)
- Method and data harmonizing engineers (> 12 months work)

Long-term resources

We foresee the following need for dedicated personnel:

- Project management (dedicated)
- Technical developer (dedicated)

- At least two method and data harmonizing engineers (dedicated for option 4)

Conclusions

This report presents four different options that could respond to the expressed criteria. The options require different degrees of data harmonization. More data harmonization and more data stored in the database facilitates analyses and automation but require more resources. These priorities need to be clarified before the implementation can proceed.