



Empowered lives.
Resilient nations.

Support in the Design of an integrated Environment Information Management System for Papua New Guinea

Cross-Cutting Capacity Development Project

Recommendations towards the EMIS

PRESENTED BY:

DR JAN BURDZIEJ

Version 1.1

1 Document control

Version	Date	Author(s)	Remarks
1.0	06 Dec 2017	Jan Burdziej	Updated Inception Report including recommended road map towards building the EMIS.
1.1	25 Jan 2018	Jan Burdziej	Updated document layout, added logos

2 Abbreviations and Acronyms

Abbreviation	Description
API	Application Programming Interface
CBD	Convention on Biological Diversity
CCCD	Cross-Cutting Capacity Development project
CCDA	Climate Change and Development Authority
CEPA	Conservation Environment Protection Authority
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
CKAN	free and open source data management platform
EMIS	Environmental Management Information System
GIS	Geographic Information System
GML	Geography Markup Language
GPS	Global Positioning System
IGIS	Integrated Government Information System
JSON	JavaScript Object Notation
NBIS	National Biodiversity Information System
OEDP	Open Environmental Data Portal
OGC	Open Geospatial Consortium
PNG	Papua New Guinea
PNGFA	PNG Forest Authority
REST API	Representational State Transfer API
SIMS	Species Information Management System
SPREP	Secretariat of the Pacific Regional Environment Programme
TMS	Tile Map Service
TOR	Terms of Reference
WFS	Web Feature Service
WMS	Web Map Service

3 Contents

1	Document control.....	2
2	Abbreviations and Acronyms.....	3
3	Contents.....	4
4	Introduction	5
4.1	Executive summary	5
4.2	Assessment of data sources for biodiversity reporting	6
5	Roadmap towards the EMIS	9
5.1	Internal information management	9
5.1.1	Establishing local IT infrastructure.....	9
5.1.2	Internal document management.....	11
5.1.3	NBIS Database.....	14
5.1.4	Wildlife / CITES Database.....	17
5.1.5	GIS data management.....	18
5.1.6	Environmental Permit Database	21
5.2	External information exchange	22
5.2.1	Open Environmental Data Portal	22
6	Summary.....	30

4 Introduction

4.1 Executive summary

This report provides a summary of a current state of how environmental information is managed by CEPA and what are the challenges that need to be addressed in order to set up an integrated Environmental Management Information System (EMIS).

The recommendations presented in this report are the outcome of the number of meetings with CEPA, CCDA, PNGFA and other stakeholders, which were conducted between Nov 15 – Dec 04, 2017. The analysis was also based on the assessment of existing reports from previous projects and consultancies. However, it has to be underlined that access to these documents was limited: many projects have not been documented properly or the documentation / final reports could not have been located. This clearly shows the need for more efficient information management within CEPA as well as between other stakeholders.

The lack of internal document management system at CEPA results in the situation where substantial amount of information and knowledge already gathered throughout various projects and initiatives is not utilized or even lost.

It is important to realize that the EMIS shall not be regarded as a one single software system that can address all of the problems that CEPA is currently facing in terms of environmental information management and reporting. Instead, **the EMIS should be regarded as CEPA's long-term strategy** consisting of smaller steps and individual projects aiming for an effective environmental information management. This long-term strategy shall ensure that all individual projects and their outcomes (e.g. databases such as NBIS or hardware and software purchases) are delivered in a coordinated way, contributing to the incrementally improving efficiency of information management, thus facilitating biodiversity monitoring and reporting capabilities.

The challenges that CEPA is facing in terms of information management can be grouped into two main categories:

- **Internal** – challenges related to document and information management within CEPA, such as lack of proper IT infrastructure, network environment, document management system, spatial data management system etc. Addressing these challenges shall result in more efficient and transparent procedures and workflows, unlocking the potential of existing databases and datasets and eventually supporting CEPA to manage and report environmental information more effectively.
- **External** – challenges related to enabling information sharing with external stakeholders, exchanging data and facilitating reporting to the Convention on Biodiversity. These challenges may

be addressed by providing a common data sharing platform that can be used by various stakeholders.

While the primary objective for CEPA is to enhance their internal capacities for effective information management, the need for better biodiversity reporting requires a substantial effort towards establishing real collaboration and data sharing with external stakeholders. Therefore, the EMIS shall focus on both internal and external challenges, and they should be addressed simultaneously. CEPA will not be able to report on biodiversity effectively unless they have their own IT systems and databases in place, but on the other hand CEPA's reporting capacities will not cover the full spectrum of biodiversity information unless proper collaboration and data sharing will be established with external stakeholders.

This report captures the relevance of various databases to biodiversity reporting as well as evaluates possible ways of integrating these databases so that the information can be easily exchanged and used for informed decision support by CEPA's senior management.

As indicated in Section 4.2, nearly 85% of all information used for the PNG's National Report to CBD comes from other stakeholders, and only 15% can be directly sourced from CEPA's internal resources. This clearly indicates that enabling **efficient data sharing between various stakeholders, both local and international, is of critical importance and shall be regarded as a key priority towards the EMIS**. To stimulate data sharing it is recommended to adopt the **Open Data approach** (see Section 5.2.1). **This can be best achieved by developing an Open Environmental Data Portal**, i.e. an online platform where multiple stakeholders can effectively exchange their environmental information and data.

While addressing CEPA's internal information management challenges will be a long process, the Open Environmental Data Portal can be deployed relatively easy and with low costs. Therefore, taking into consideration the organization's constraints of this CCCD project (already delayed and subjected for a no-cost extension), **the further effort and deliverables within this assignment (i.e. Short and Medium Term Action Plan as well as the Terms of References) will focus on developing the Open Environmental Data Portal**.

4.2 Assessment of data sources for biodiversity reporting

The main aim of this section is to identify key data sources that are used for developing the PNG's National Report to the Convention on Biological Diversity (CBD). The 5th National Report to CBD (draft) was screened and all major data sources quoted in the report were identified and grouped into different categories, i.e.: Data, Information and Spatial Resources.

- **“Data”** category refers to all quantitative resources, such as statistical values, indicators, numbers, percentages etc. For example number of endemic species in PNG, number and size of protected areas, GHG emissions etc.
- **“Information”** category refers to general descriptions of different projects, references to national and international policies as well as other qualitative information.
- **“Spatial Resources”** category was assigned to the information which was presented in a spatial context, e.g. using maps.

Additionally, all data sources were assigned to the data owner, i.e. the organization which has provided this particular resource. In many cases one data source was assigned to more than one organization (e.g. results from a collaborative project managed by local and international organization). However, it has to be noted that in many cases identifying data owners was difficult due to the lack of proper references. The statistics gathered from this assessment shall be regarded as rough estimates rather than precise calculations, although they clearly show the general trends.

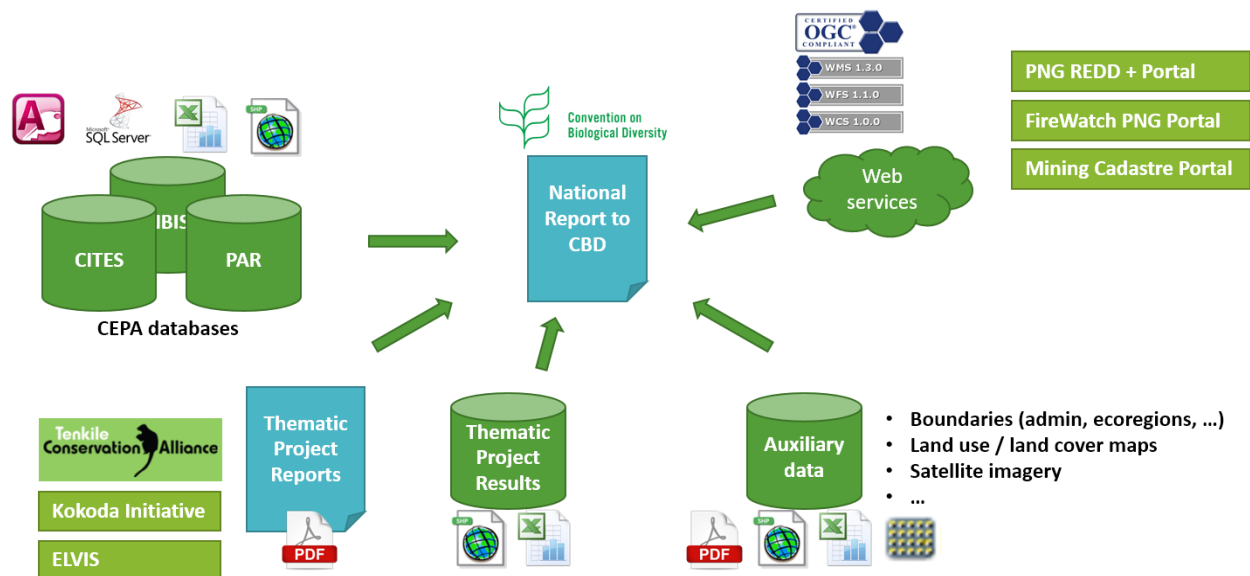


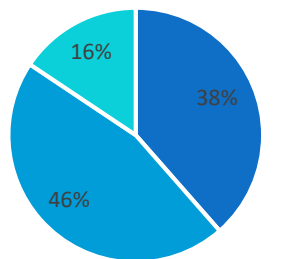
Figure 1. Overview of various data sources contributing to the PNG's National Report to CBD

The key findings from this exercise are the following:

- In total, 96 major data sources were identified within the 5NR to CBD.
- Nearly half of all resources (46%) fall into the “Information” category, i.e. they are qualitative information (e.g. description of various projects, overview of policies and actions undertaken in terms of environmental conservation etc.) rather than actual statistics or indicators.
- Still, approximately 38% of all information use quantitative values, such as indicators, absolute values, statistics etc. to describe the current status of environmental conservation,

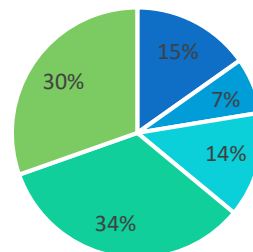
- Approximately 16% of all resources are presented using maps. However, most of the other resources (both quantitative and qualitative) can be presented in a spatial context. E.g. the location of various donor-funded projects might be presented on the map, or. Green-House Gases emissions could be supported by a map displaying emissions by provinces or districts.
- Out of all 96 resources, only approximately 15% are (or potentially might) come from CEPA's internal databases and registries. This information is related primarily to:
 - the population and distribution of selected species (NBIS database),
 - number and size of protected areas,
 - wildlife trade and export statistics,
 - crocodile survey results (Wildlife database).
- Another potentially major source of relevant environmental information within CEPA is related to Environmental Permits (water extraction and waste discharge, pesticides import, ozone depleting substances, plastic bags statistics etc.).
- The majority of information (34%) comes from various international / global organizations and databases (e.g. global biodiversity surveys, land use / land cover maps, climate information etc.).

Data Sources by Type



■ DATA ■ INFO ■ SPATIAL

Data Sources by Organization



■ CEPA ■ CCDA ■ PNGFA ■ International ■ Other local

Figure 2. Analysis of data sources from the PNG's 5th National Report to CBD

5 Roadmap towards the EMIS

5.1 Internal information management

5.1.1 Establishing local IT infrastructure

Current status

One of the main problems with data integration and information sharing within CEPA is the lack of an operational IT infrastructure, specifically lack of Local Area Network that can ensure connectivity between different computers. Another important constraint is lack of access to the Internet. CEPA is in the transition from multiple locations into the main headquarters at Savannah Heights Dynasty A building, Waigani area. It is expected that all employees and all CEPA units will move to the new location by the end of 2017, though this date is approximate and not confirmed. Currently the new office does not have a local area network established and most of the computers do not have access to the Internet. However, the infrastructure is there, e.g. CAT6 LAN cables are laid, LAN sockets are available in every room and a server room is equipped with Rack cabinet, switch etc.

Additionally, two servers, i.e. Data Storage and GIS Server were acquired and delivered for CEPA in March 2017. They are located in the server room at the Savannah Heights office. However, servers have not yet been unpacked nor installed. Below is the brief specification of the servers acquired:

- HP DL80 Gen9 12 LFF CTO Server (2 units)
- Software for Server 1: Windows server standard 2012, Windows server data center 2012, Windows server CAL 2012, SharePoint Server 2016, SharePoint Standard CAL 2016, SharePoint Enterprise CAL 2016
- Software for Server 2: Windows Server Standard 2012, Windows Server Data Center 2012, Windows Server CAL 2012, Exchange Server Standard 2016, Exchange Server Enterprise Standard CAL 2016, Exchange Enterprise CAL 2016
- Antivirus for server 1 and 2: Kaspersky AntiVirus 2016

CEPA has already engaged a local IT company and received a proposal for the design and configuring servers and local area network LAN/VLAN. However, the project was put on hold due to lack of funds, as it was estimated that approximately K300,000.00 will be required to procure all the required software, hardware as well as IT services need to set up the whole IT infrastructure for CEPA. There is no information on when the work might be commenced.

Recommendations

- The first action to be taken is to revise the initial proposal for setting up the local IT infrastructure. The budget of K300,000.00 seems to be high and lack of funding puts the project on hold. CEPA shall consider setting up the network environment using existing resources (especially servers procured by UNDP) to minimize costs and initiate this process as soon as possible.
- Initially, it might be sufficient to unpack existing servers, install all the available software (Windows Server, Exchange Server etc.) and configure VLANs. No additional software or hardware is needed to set up the initial local area network along with servers.
- It might be recommended to engage an independent IT network consultant to revise the current proposal for the network setup, or to design the best LAN/VLAN architecture, while taking into account business requirements as well as existing infrastructure, hardware (e.g. UNDP servers) and software (e.g. SharePoint).
 - The design shall include setting up the **web server, email server** and appropriate **security** measures (DMZ, firewalls, antivirus software etc.) to ensure the network is secure.
 - The results of the consultancy shall include the Terms of Reference for further phase, e.g. the purchase of software and hardware, training and actual establishment of the LAN.
 - Different approaches to user management, roles and permissions shall be discussed and most optimal solution shall be selected (e.g. domain-based vs workgroup-based network).
 - Existing software, such as Microsoft Exchange Server 2016 for email server shall be used when planning the network configuration.
- Due to the transformation from DEC to CEPA, a **new domain** will be required for CEPA (e.g. **cepa.gov.pg**). So far the previous domain was dec.gov.pg. However, right now most of CEPA's employees use gmail and other publicly available email accounts.
- Local WiFi network shall also be considered as part of the IT infrastructure setup so that mobile devices and temporary users can access Internet and / or local network.
- The design shall be consulted and developed in close collaboration with the internal IT staff to ensure that they understand the design and CEPA will be able to maintain it in the long term using internal human resources.
- The design shall be cross-checked and preceded by the assessment of the **potential integration with IGIS** (Integrated Government Information System), as some components may be migrated to the IGIS infrastructure (e.g. mail server). CEPA has acquired certain hardware as part of the IGIS project, though it is not clear how these resources were utilized.
- Once the design of the IT infrastructure is revised / updated, an IT contractor shall be engaged to install, set up and configure all the required hardware and software components, e.g.:
 - supply the required hardware (e.g. WiFi routers if not available) and software (if needed),

- installation of servers (including required software maintenance, updates and configuration),
- installation of Windows Server, Exchange Server and other software.
- configuring users, roles and privileges,
- establishing LAN / VLAN,
- providing training for internal IT personnel.
- **CEPA's official website** might be developed to provide basic information about the Organization, news & events, projects, regulations etc. It is recommended to design the website using one of the existing Content Management System so that updating the content may be done with internal resources.

5.1.2 Internal document management

Current status

There is no document management in CEPA at the moment. The result is that documents are difficult to find, some of them, e.g. the results of various consultancies and reports related to donor-funded projects are simply lost. Substantial amount of time is wasted on discovering various documents, many documents exist in duplicated versions and document sharing is done using external drives or email. Therefore, it is important to implement a proper document and information management, so that all documents can be easily archived, versioned, indexed and shared within the organization.

Along with the servers CEPA has acquired Microsoft SharePoint software licenses, i.e.: SharePoint Server 2016, SharePoint Standard CAL 2016, SharePoint Enterprise CAL 2016. SharePoint is a web-based, collaborative platform that integrates with Microsoft Office. It is primarily sold as a document management and storage system, but the product is highly configurable and usage varies substantially between organizations¹.

Some of the key features of SharePoint are:

- It allows groups to set up a centralized, password-protected space for document sharing.
- Documents can be stored, downloaded and edited, then uploaded for continued sharing.

SharePoint is a complex system, which offers a wide array of features. It requires experienced personnel to set up the system correctly, so that it can be used effectively across the organization.

¹ <https://en.wikipedia.org/wiki/SharePoint>

The SharePoint platform can be used by CEPA primarily for document management, so that documents can be easily shared with specific users, edited, versioned and archived. This will help reduce time spent on document discovery, reduce risks involved in using outdated documents, prevent documents from being lost etc.

The SharePoint Enterprise license introduces extra features that might be used for Business Intelligence integration, dashboards, and business data surfacing. Potentially, SharePoint can be used to connect to existing databases, such as Wildlife / CITES database, NBIS or eventually Environmental Permits database (not yet developed) in order to create live dashboards for CEPA’s management. Dashboards can be used by the management to get an instant feedback for example on number of permits issued, total revenues per permit type, total value of pending payments etc.

A dashboard is a related group of interactive scorecard and report views that are organized together in a SharePoint or Web-hosted site². A dashboard might resemble the following image:



Figure 3. Sample Dashboard

² <https://support.office.com/en-us/article/Getting-started-with-dashboards-EDC6A32C-85C5-443D-9103-56DCC7BB24A4>

“SharePoint Server enables you to create and use dashboards that provide up-to-date information in a centrally managed, easily accessed location. You can use SharePoint Server tools to create and use dashboards that are suitable for an individual, team, group, or the entire organization”³.

Despite its powerful features and potential significant business benefits, using SharePoint does have certain disadvantages and limitations. Some of them are:

- Licensing & maintenance costs,
- Lack of internal human capacities and skills to deploy and maintain SharePoint system internally,
- Risks related to system failures (e.g. due to network connectivity issues).

“A recent report found that 40% of organizations don’t consider their on-premise SharePoint implementation to be a success – with two-thirds (67 percent) blaming this on inadequate user training, 66 percent saying it’s too difficult to use, and 64 percent reporting that lack of support from senior management is the reason that their deployments have failed.”⁴

Alternative solution to SharePoint might be simply establishing a network drive with an agreed folder structure and appropriate access permissions for different users. The folder structure can reflect the structure of CEPA (e.g. subfolders for every business unit), so that each unit within CEPA might share and find their documents via a shared / network drive. This solution is simple, does not require any special skills to set up and maintain (despite basic network admin skills) and is less prone to damages caused by system / network failures. However, this solution does not allow for document versioning & discovery nor does it allow for developing customized data dashboards and harvesting data from existing databases for reporting and management purposes.

Recommendations

- Need assessment shall be conducted in order to identify if SharePoint meets the CEPA’s business needs and requirements. Additionally, costs of maintenance, license upgrades etc. need to be taken into account when deciding whether this software shall be adopted by CEPA for their data management. This need assessment might require engaging an external SharePoint consultant.

³ <https://support.office.com/en-us/article/Getting-started-with-dashboards-EDC6A32C-85C5-443D-9103-56DCC7BB24A4>

⁴ <https://www.huddle.com/product/collaboration/sharepoint-alternative/>

- Possible integration of SharePoint with existing SQL Server databases (e.g. NBIS or Wildlife) need to be assessed and verified.
- SharePoint works with Microsoft Office package. Therefore it is important to assess a number of Office users within the organization and draw a plan aimed at ensuring all users use the same version of the software that is compatible with SharePoint.
- If CEPA decides to implement SharePoint for their document management, next step would be to hire a SharePoint developer to implement the solution based on the requirements and the implementation plan developed in the previous step.
- To ensure the system is well maintained, internal IT staff need to be trained. This can be done through the assistance along the design and implementation phases, but also through a dedicated training. Once IT staff is trained, they can then organize internal training for end users to ensure everyone can use the platform effectively and efficiently.
- If CEPA decides not to implement SharePoint as their document management system (e.g. for the reasons of cost saving), the next recommended step would be to establish proper shared / network drive along with a proper subfolder structure and user access management. This can be achieved quickly, without any additional costs, using exiting IT and human resources.

5.1.3 NBIS Database

Current status

The final NBIS database system has been delivered to CEPA on December 1st, 2017 by Tom Stojanovic, NBIS developer. It is an upgraded version of the system previously known as the Species Information Management System (SIMS). The objective of NBIS database is to store observations of different species across PNG, so that their population and spatial extent of habitats can be monitored. Observations are grouped into Site Visits, which are further grouped into Surveys. Additionally, taxonomic information for the selected species is stored within NBIS to support reporting species for every observation.

The database has been developed using the following stack of technologies:

- Microsoft SQL Server Express database management system for data storage. Additionally, Stored Procedures are used for streamlining specific database operations (such as data update or retrieval) and ensuring database consistency.
- Microsoft Access provides a Graphical User Interface (GUI) through which a user can input new and preview existing data. It is also used for database administration (e.g. user management). Additionally, an internal scripting environment, i.e. Visual Basic for Application (VBA) was used to provide additional functionality and connectivity with the database.

The NBIS is currently configured to work in a disconnected environment, with one dedicated laptop acting as a Master NBIS database and several other laptops acting as Slave. Data can be exchanged using database backup files. However, according to the documentation, the database can be configured to work in a network environment (which was the initial recommended setup). This can be configured by the administrator without the need for source code modification.

The NBIS requires Windows 10 and Microsoft Office 2010 or above with Microsoft Access to run. This can potentially involve several risks, such as:

- Dependence on Microsoft products for every machine using NBIS – both master and slave terminals (updates, version conflicts etc.),
- Costs related to software licenses (e.g. in case more users will require access to NBIS).

According to the PNG's 5th National Report to CBD NBIS shall be not only for internal purposes, but also for sharing the information with general public and other national and international databases:

*“The current Species Information Management System (SIMS) database housed within DEC to be upgraded to a database that has **sections that are publically available and user friendly** and others, which are secure for internal use only. **This database is to be linked to others within the country and to databases outside the country** that contain information on PNG biodiversity wherever possible.”*

*“It was suggested that the Department maintain an internal Biodiversity database that is **populated from existing data and from future surveys and Environmental Impact Assessments**. This **database should be linked with other specialized biodiversity databases** in the country such as the PNG Forest Authority database and the Binatang Research Centre database through a user agreement.” (PNG 5NR – Draft)*

However, the current set up does not allow for seamless interoperability with external databases and it does not provide any public-facing interface or service that would allow specific data to be shared with general public. Despite the lack of technological arrangements, there is also lack of consensus within CEPA and NBIS administrators about which data from NBIS can actually be published online.

Recommendations

- Once the local IT infrastructure is established at CEPA, the NBIS database shall be configured to work in a network environment, with one central database and multiple client machines used for data preview and entry.

- The NBIS database shall be installed on existing servers to enable centralized maintenance and security (e.g. regular backups, setting up Redundant Array of Independent Disks – RAID to ensure better data protection, Anti-virus software protection etc.).
- NBIS shall be integrated with Wildlife database that is currently under development. In particular NBIS authority files and species taxonomy shall be used within the Wildlife database to ensure standardized naming conventions and integrity of taxonomic names used in both databases. On the other hand, Wildlife database shall be capable of supplying data directly to NBIS, especially on crocodile surveys and observations.
- If / once the SharePoint platform is deployed, it shall be integrated with the NBIS database to provide some indicators in terms of species observations (e.g. number of new observations entered last month).
- According to the PNG's 5th National Report to CBD (draft) some data from NBIS shall be made publicly available. The NBIS team along with the CEPA's management need to identify what information from NBIS can actually be published as Open Data. This might include e.g. raw observations data on non-critical species, summary statistics (e.g. number of observations for individual species, number of observations per year) or filtered observation data (e.g. without geographical coordinates or with a generalized location, referenced to the LLG or a custom sampling grid etc.). A good example of a publicly available database on species observations is Global Biodiversity Information Facility (GBIF)⁵
- Once public information is identified and extracted, it shall be published through the Open Environmental Data Portal (see Section 5.2.1). Internal procedures need to be developed in order to ensure that these datasets are updated regularly.
- The recommended method for exchanging NBIS data with external (but potentially also internal) database systems would be via web services / REST API. A good starting point would be to develop REST API for publishing specific information on the Open Environmental Data Portal. The REST API for NBIS shall be integrated with existing Stored Procedures within NBIS SQL Server to ensure data integrity and secure data access.
- The critical part of the NBIS is to ensure that there is a coordinated, systematic and continuous flow of incoming data (i.e. observations of different species across the PNG). There are several sources of information within CEPA that might supply data to NBIS on a regular basis, i.e.: Wildlife database (specifically crocodile observations) and Environmental Permits database (not yet developed) along with Environmental Impact Assessments (including species observations).

⁵ <https://www.gbif.org/>

However, even once integrated with NBIS, these databases will provide only fragmented information. Therefore, other data sources need to be evaluated. In order to encourage data input, the whole process need to be simple and easy. Dedicated mobile application for wildlife observations may be considered in the future, or some online web interface for easy data entry. Also, crowd-sourcing can also be regarded as one of the potential data sources.

5.1.4 Wildlife / CITES Database

Current status

The database is at the stage of development. So far all the data is handled in an analogue way, through the paper forms, which makes it difficult to keep track of the export permits and Wildlife trade. The Wildlife database is being developed by Tom Stojanovic, who was also responsible for developing the NBIS database. The Wildlife database is to be developed using the same architecture as NBIS, i.e. data along will be stored in the Microsoft SQL Server Express database, while the Microsoft Access will provide a Graphical User Interface for managing the data.

It can be expected that before the completion of the development of Wildlife database CEPA will have its local IT infrastructure already in place, i.e. computers will be connected to the local area network and servers will be installed and configured.

Recommendations

- Once the local IT infrastructure is established at CEPA, the Wildlife database shall be configured to work in a network environment, with one central database and multiple client machines used for data preview and entry.
- The Wildlife / CITES database shall be installed on existing servers to enable centralized maintenance and security (e.g. regular backups, setting up Redundant Array of Independent Disks – RAID to ensure better data protection, Anti-virus software protection etc.).
- When designing the Wildlife database it has to be ensured that relevant information (e.g. crocodile surveys) are automatically shared with NBIS database (see Section 5.1.3). Additionally, species taxonomy data from NBIS need to be integrated within Wildlife / CITES database to ensure consistent use of species names.
- If / once the SharePoint platform is deployed, it shall be integrated with the Wildlife database to provide relevant data for internal dashboards for CEPA's management (e.g. number of export permits issued, revenues from export permits, crocodile survey statistics etc.).

- Wildlife / CITES database has a major potential for offering online services to the interested stakeholders / customers. For example, export applications may (and eventually should) be available for the customers (citizens, researchers, companies etc.) through the on online system and interactive forms. Establishing online services will require developing appropriate technical solutions (e.g. web services / REST API for communication with the Wildlife database) but it might also require updating certain regulations and / or policies.

5.1.5 GIS data management

Current status

CEPA is managing a substantial set of spatial datasets. However, there is currently no coordinated way of managing these datasets effectively. As a result, multiple copies of same datasets are in use, some datasets cannot be easily discovered (are stored on local drives, memory stick, external drives etc.).

Existing databases, i.e. NBIS and Wildlife, have limited support for handling spatial data. The NBIS contains a list of species observations with a location identified by latitude and longitude coordinates. Similarly, the Wildlife database will use geographic coordinates to describe the location of crocodile surveys or the locations where wildlife specimens are collected. The location information, however, is not stored in a spatial data types (such as “geography” type used by SQL Server⁶), rather than it is stored in a flat non-spatial table using decimal values for geographic coordinates. This introduces certain limitations in terms of how data can be queried and analyzed, e.g. searching for data by drawing a specific extent will not be possible. Also other spatial queries are not possible with this approach, e.g. spatial intersections, buffering, identifying points in polygons, etc.

Other databases, such as Environmental Permit Database (that is still to be developed) will also require a dedicated GIS database to handle spatial information. One of the most popular GIS database / server solutions is ArcGIS Server (ArcGIS Enterprise), which is a proprietary system that can be deployed using other database systems, such as Oracle, SQL Server or PostgreSQL. This solutions has been adopted by PNG Forest Authority, as well as by the Mineral Resource Authority (MRA). However, a wide range of free and Open Source solutions exist and can be effectively implemented for hosting and serving spatial data, such as GeoServer, MapServer, PostGIS database etc.

⁶ <https://docs.microsoft.com/en-us/sql/t-sql/spatial-geography/spatial-types-geography>

ArcGIS Enterprise includes several software components that are designed to work together⁷. A foundational setup of ArcGIS Enterprise consists of a number of these components configured in a certain way; this is called a base ArcGIS Enterprise deployment. The base deployment consists of the following:

- ArcGIS Server—This ArcGIS Server must be licensed as ArcGIS GIS Server Standard or ArcGIS GIS Server Advanced and configured as the hosting server for your portal.
- Portal for ArcGIS.
- ArcGIS Data Store—Upon setup, this should be configured as a relational and tile cache data store.
- ArcGIS Web Adaptor—One Web Adaptor installation for traffic to your portal and one for traffic to your ArcGIS GIS Server.

There are several ways to configure the base deployment. See the sections below for details. For details about the required machine specifications for the ArcGIS Enterprise components:

- Single-machine deployment
- Multitiered deployment
- Highly available deployment

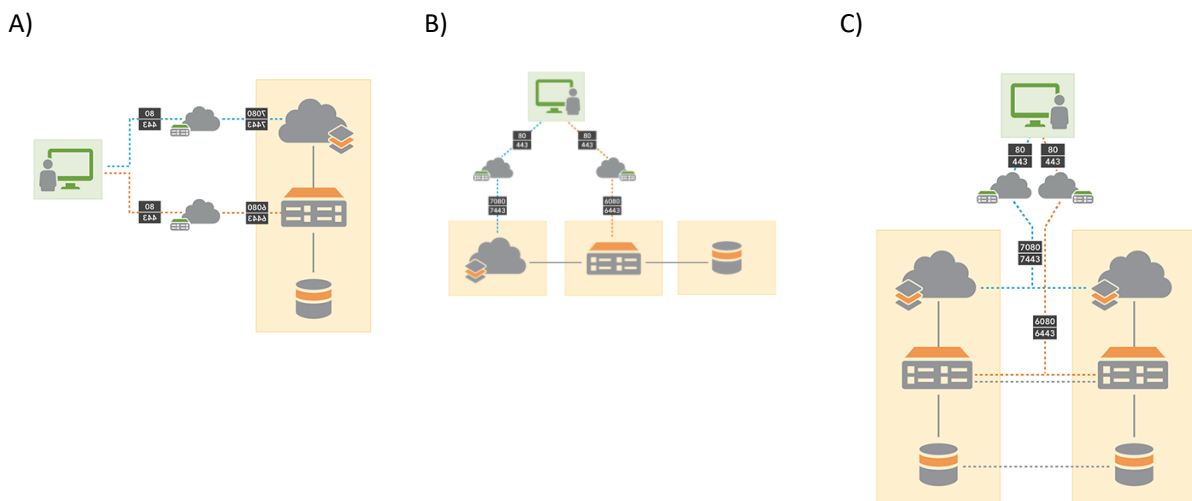


Figure 4. Different ArcGIS Enterprise deployment scenarios: A) Single-machine, B) multitiered, C) highly available

⁷ <http://server.arcgis.com/en/server/latest/get-started/windows/base-arcgis-enterprise-deployment.htm>

Single-machine deployment. The first way to configure the base ArcGIS Enterprise deployment is the minimal all-in-one configuration, where all of the components are installed on a single machine or virtual machine.

Multitiered deployment. You can also configure the base ArcGIS Enterprise deployment in a multitiered environment, where each component is installed on a separate physical or virtual machine.

Highly available deployment. There are also highly available configurations where each component is configured with redundancy to minimize downtime in scenarios where one or more machines become unavailable. High-availability deployments mirror the two configurations outlined above.

The main alternative to ESRI ArcGIS Server / Enterprise is a stack of various open source systems, i.e.:

- PostgreSQL / PostGIS – spatially enabled relational database management system
 - Data can be edited using free and open source desktop QGIS software
- GeoServer – can be used for publishing spatial datasets via web services, such as WMS, WFS, WCS as well as Tile Map Services.
 - This solution has been implemented e.g. by the REDD+ Forest Monitoring Portal
- Web services / REST API – developed using open source frameworks, such as Django / GeoDjango.
 - Web services can be used for both retrieving spatial data from PostGIS database, as well as for data update, mainly through web and / or mobile applications.
- Leaflet or Open Layers – free and open source mapping libraries that can be used for visualizing and editing spatial data through the web browser.

Recommendations

- Before setting up the GIS server / database solution, the local IT infrastructure need to be established. Otherwise the GIS server will not be used, which may result in unnecessary licensing costs.
- Once the IT infrastructure is installed, it is recommended to hire a GIS consultant to conduct a detailed GIS need assessment, including:
 - Inventorying existing GIS software licenses and developing long-term GIS licensing policy,
 - Benchmarking open source and proprietary solutions and mapping them with current and prospective GIS needs within CEPA,
 - Identifying and evaluating existing spatial datasets managed by CEPA,
 - Evaluating GIS capacities of existing systems (NBIS, Wildlife) as well as systems to be developed (e.g. Open Environmental Data Portal, Environmental Permit Database etc.),

- The need assessment should be followed by the development of specific Terms of Reference for the GIS Server database deployment,
- Evaluation of needs in terms of mobile spatial data collection.
- Both on-premise as well as on-line (cloud-based) solutions shall be evaluated. However, the unstable and unreliable access to the Internet need to be accounted for when evaluating cloud-based solutions.
- It has to be ensured that GIS Server is compatible with CEPA's desktop GIS software (ArcGIS / QGIS?).
- The integration with existing and planned databases such as NBIS, Wildlife or Environmental Permit databases need to be evaluated so that the selected GIS server solution can be effectively used to preview, edit, update and export spatial data.
- Future licensing costs (initial purchase and maintenance) need to be analyzed and accounted for. Proprietary GIS server solutions might incur substantial licensing costs in the long term. CEPA need to ensure that, once the GIS server solution is selected, it will be able to sustain the licensing costs in the future.

5.1.6 Environmental Permit Database

Current status

Managing environmental permits is one of the key responsibilities of CEPA. This is particularly important because the process involves managing and collecting payments related to environmental permits. CEPA has made two attempts to develop the environmental permit database system. However, none of these projects resulted in a usable solution. Currently, Environmental Permits are managed using stand-alone Excel spreadsheets, which makes it difficult to keep track of all the permits, monitor required payments or enforce various obligations related to each permit.

Within the Environmental Permit process CEPA is acquiring a substantial amount of information, including Environmental Impact Assessments. However, this information is delivered mainly through hardcopy reports. There is a tremendous amount of information that can be extracted at different levels of the permitting process. This requires a robust database system that will allow to store, analyze, process and report all the related information.

The scope of the Environmental Permit Database is similar in scope and type of information to the Mining Cadastre Portal system developed by Mineral Resource Authority (MRA). This Portal was funded through

the World Bank's Second Mining Sector Institutional Strengthening Technical Assistance Project and implemented by Spatial Dimension, the developers of FlexiCadastré⁸.

Recommendations

- Consultant will be required to conduct a detailed need assessment for the Environmental Permit Database. This need assessment shall identify and document all use cases, documents, forms, workflows and regulations that are connected with permitting process.
- As part of the need assessment, it is recommended to contact the Mineral Resource Authority and the company that has developed the Mining Cadastre system to investigate if this existing solution could be adapted for CEPA and the Environmental Permit Database. The costs of implementing this solution need to be compared with costs of developing new, custom system.
- Based on the need assessment, the software requirements specification and a Terms of Reference document shall be developed.
- It is recommended that the system is designed to work in a network environment, so that the permitting process can be delivered to the end users through the web-based interface and online services.
- Once the GIS Server database is installed, it can be used as part of the Environmental Permit Database system, to reduce costs of the development.

5.2 External information exchange

5.2.1 Open Environmental Data Portal

Current status

As indicated in Section 4.2, nearly 85% of all information used for the PNG's National Report to CBD comes from other stakeholders, and only 15% can be directly sourced from CEPA's internal resources. This clearly indicates that enabling **efficient data sharing between various stakeholders, both local and international, is of critical importance and shall be regarded as a priority task towards the EMIS.**

The current world-wide trend supporting data sharing is referred to as an Open Data Approach.

⁸ <https://portal.mra.gov.pg>

- Open data is data that anyone can access, use and share.
- Governments, businesses and individuals can use open data to bring about social, economic and environmental benefits.
- It is a global initiative implemented by many countries, such as United States, United Kingdom, Australia, Singapore, New Zealand and many others.
- The concept of Open Data supports Sustainable Development Goals:
„This kind of data is essential both to help achieve the SDGs and to measure progress in meeting them.“
„Open Data (...) helps ensure environmental sustainability through transparent data that can help reduce pollution, conserve natural resources and build resilience to climate change.“ (World Bank)

While there are many motivations for Open Data, some of the most important are:

- **Transparency.** Citizens have right to know.
 - Where are protected areas? What are the restricted activities there? Where is logging allowed?
 - Increased trust in organizations
- **Innovation.** Open data stimulates innovative data use.
 - E.g. mobile app for reporting illegal logging
 - Engaging stakeholders: local ICT entrepreneurs, R&D, academia
- **Participatory Governance.** Involving citizens in decision-making.
 - Crowdsourcing – benefiting from local knowledge
 - Stimulates the development of e-Government and online services
- **Cost savings**
 - Reduce data redundancy, less time & money wasted on data discovery, prevent lost data

5.2.1.1 UN E-Government Development Index

“The E-Government Development Index presents the state of E-Government Development of the United Nations Member States. Along with an assessment of the website development patterns in a country, the E-Government Development index incorporates the access characteristics, such as the infrastructure and educational levels, to reflect how a country is using information technologies to promote access and inclusion of its people. The EGD Index is a composite measure of three important

dimensions of e-government, namely: provision of online services, telecommunication connectivity and human capacity.”⁹

Mathematically, the E-Government Development Index (EGDI) is an average of normalized scores on the three most important dimensions of e-government, namely: scope and quality of online services (Online Service Index, OSI), status of the development of telecommunication infrastructure (Telecommunication Infrastructure Index, TII) and inherent human capital (Human Capital Index, HCI).

OSI evaluates mainly:

- open government data,
- e-participation,
- multi-channel service delivery,
- usage up-take,
- innovative partnerships through ICT.

Unfortunately, the PNG is underperforming in terms of E-Government development. By developing Open Environmental Portal PNG will be able to improve its performance in terms of E-Government.

Country	Rank	EGDI16	Open Data Portal
United Kingdom	1	0.9193	CKAN
Australia	2	0.9143	CKAN
Republic of Korea	3	0.8915	
Singapore	4	0.8828	CKAN
Fiji	96	0.4989	
Papua New Guinea	179	0.1882	
Somalia	193	0.0270	

Figure 5. E-Government Development Index survey results(2016)

⁹ <https://publicadministration.un.org/egovkb/en-us/About/Overview/-E-Government>

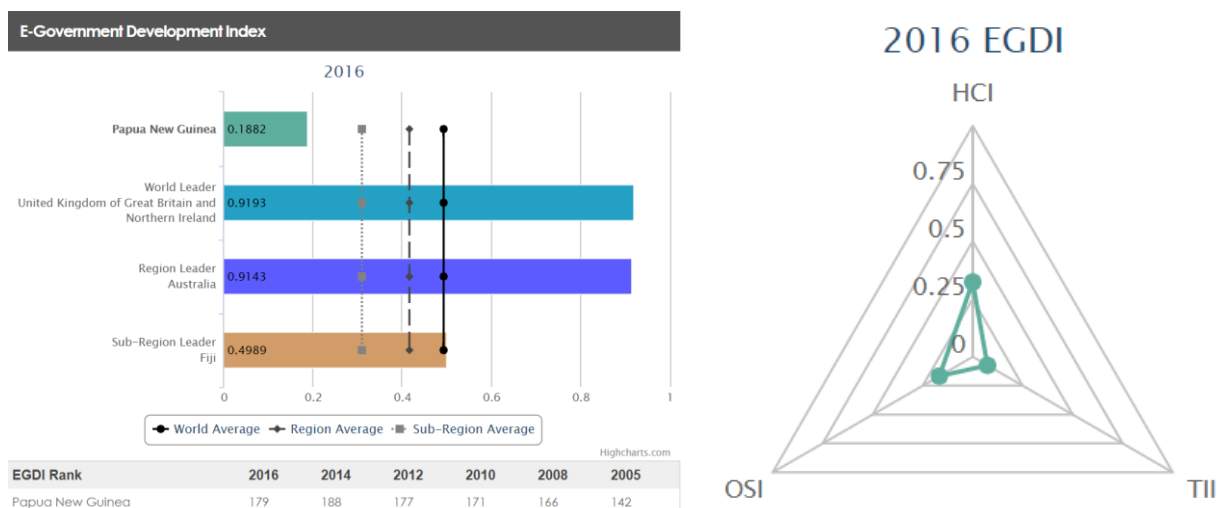


Figure 6. PNG's score in terms of E-Government

5.2.1.2 Benefits of Open Environmental Data Portal for CEPA

Through the establishment of the OEDP, the following benefits can be achieved by CEPA:

- Engage other stakeholders through the use of collaborative online platform, in particular:
 - CCDA – Climate Change and Development Authority
 - PNGFA – PNG Forest Authority
 - DAL – Department of Agriculture and Livestock
 - DLPP – Department of Land and Physical Planning
- Streamline publishing, sharing, finding and using data
- Environmental data will be discoverable from one single gateway
- Portal will support data analysis, visualization and reporting to CBD and other conventions
- Portal will provide support for both structured data (e.g. Oracle or SQL Server databases, such as NBIS or Wildlife / CITES) and unstructured data (e.g. PDF files, reports from various projects etc.)
- Open Data Portal can be deployed using free and Open Source data platform, i.e. CKAN, which was used for developing portals for over 140 countries and organizations across the world, including national portals for such countries like United Kingdom, United States, Ireland, Australia, New Zealand, Singapore, Estonia
- Using existing platform (i.e. CKAN) will significantly reduce „time to market” and decrease costs of the deployment. The initial prototype can be delivered within one month, while the more customized version might be developed incrementally, following the Agile approach (delivering individual functionalities in short-term cycles).
- The platform is easy to use. Once installed, it can be maintained by CEPA’s internal IT staff.

- CKAN has a great community support and features over 250 extensions that can be used for adding new functionalities to the Open Data Portal.
- Open Environmental Data Portal will support various international projects and initiatives, such as
 - UN E-Government and E-Participation (resulting in PNG’s better performance in global rankings),
 - UN Sustainable Development Goals
 - INFORM project coordinated by SPREP – Building National and Regional Capacity to Implement Multilateral Environment Agreements (MEA) by Strengthening Planning and State of Environment Assessment and Reporting in the Pacific (<https://www.sprep.org/Environmental-Governance-Monitoring/Planning-and-Monitoring/>)
- Establishing an Open Data Portal may be reported to the CBD as one of the PNG’s major efforts towards transparent environmental information management at the national and international level.
- By initiating the Open Data Portal CEPA will become the leader of Open Data initiative in PNG.

Country	Open Data Portal URL
Australia	https://data.gov.au/
European Union	https://data.europa.eu/euodp/en/home
New Zealand	https://www.data.govt.nz/
Republic of Korea	https://www.data.go.kr/e_main.jsp
Singapore	https://data.gov.sg/
Sweden	https://ecds.se/
United Kingdom	https://data.gov.uk/
United States	https://www.data.gov/

Figure 7. Examples of open data portals

5.2.1.3 CKAN vs DKAN

There are two major free and open source open data platforms available on the market, i.e. CKAN and DKAN, as well as some other proprietary systems, such as ArcGIS Open Data. The benefits of using free and open source platforms are obvious: reduced licensing costs, escaping from the vendor lock-in trap, enabling unrestricted customization of the source code, as well as embracing more sustainable, community-based support.

While there are pros and cons for both CKAN and DKAN, there are several major advantages of adopting CKAN as a platform for implementing Open Environmental Data Portal for CEPA / PNG:

- CKAN platform has been adopted by many other countries for developing their official national open data portals, e.g. in United Kingdom, United States, Ireland, Australia, New Zealand, Singapore or Estonia,
- CKAN has much larger community of contributors (see Table below),
- CKAN is a “standalone” platform while DKAN is based on Drupal Content Management System,
- CKAN uses Python as a main programming language, which is currently gaining more popularity over PHP, used by DKAN (Figure 8),
- Other UN initiatives have also adopted CKAN for their data management portals, e.g. Humanitarian Data eXchange platform (HDX¹⁰) developed by UN Office for the Coordination of Humanitarian Affairs (UNOCHA).

	CKAN	DKAN
Number of commits*	18,058	3,232
Number of branches	868	127
Number of releases	72	64
Number of contributors	170	44
Primary language	Python	PHP

* As of December 5, 2017, source: <https://github.com/ckan/ckan>, <https://github.com/GetDKAN/dkan>



Figure 8. Google Trends analysis for PHP (blue) and Python (Red) over last 5 years.

¹⁰ <https://data.humdata.org/>

5.2.1.4 Overview of CKAN Data Model

CKAN supports multiple Organizations. Each Organization might have its own administrators, editor and members, i.e. users with different roles and privileges. It supports both structured (e.g. CSV, XLS or database tables) and unstructured data (e.g. PDF reports or JPG images). Structured data is stored within the Data Store (PostGIS database), while the unstructured data is stored in a File Store. Once structured data is uploaded to CKAN, it can then be automatically previewed using Data Explorer / Table View, Chart View or even Map View (if latitude and longitude values are available). Additionally, structured data is automatically exposed via REST API, which allows external applications to connect and use the data in a machine-readable format.

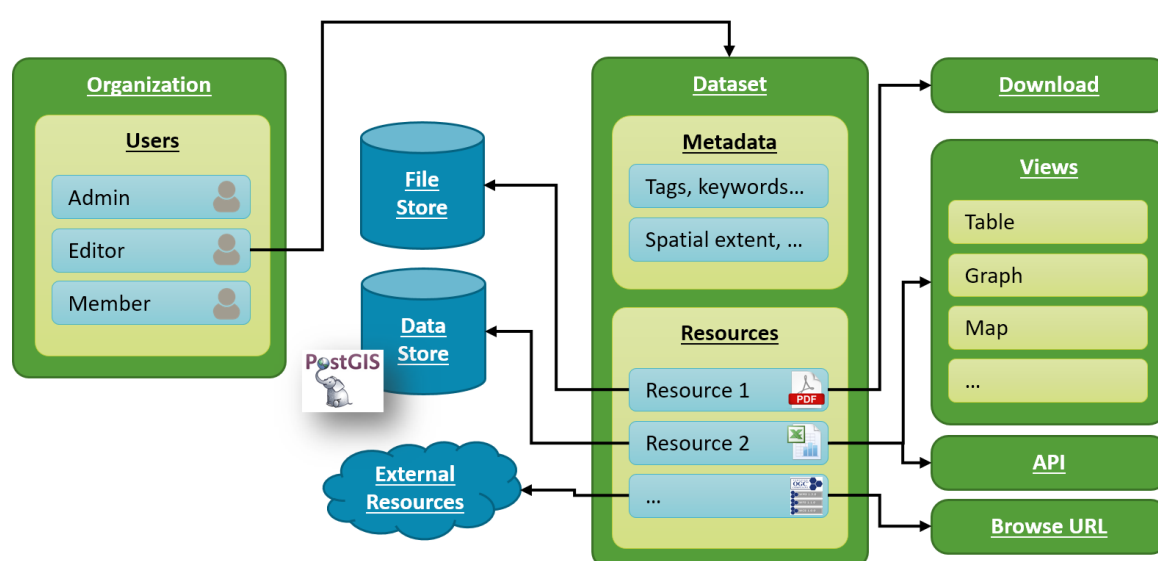


Figure 9. Overview of CKAN data model

By using existing free and open source CKAN platform, the PNG's Open Environmental Data Portal can be deployed within a relatively short time: the prototype version (with just a default functionalities and standard interface) can be deployed within just a few days, while the more advanced and customized version might be delivered within approximately 1-3 months, depending on the scope of customizations required.

Recommendations

- It is recommended to deploy an Open Environmental Data Portal as one of the key components of the overall EMIS setup. The portal will meet the key objectives of the EMIS, i.e. it will support CEPA to address the convention obligation related to reporting requirements under the UN Convention

on Biological Diversity primarily by providing a data management platform and facilitating information exchange with other stakeholders.

- The OEDP shall provide extensive support for API. All core features, such as access to datasets and resources need to be exposed through the API.
- The OEDP need to support both structured data (such as CSV files or database tables) as well as unstructured data (such as PDF reports, images etc.).
- The OEDP need to provide a support for handling spatial files (e.g. by providing built-in map preview, support for WMS services, ArcGIS REST services, shapefiles, GeoJSON etc.). The support for spatial data shall include defining spatial extent for datasets as well as map-based search functionality (i.e. identifying datasets and resources for a given area of interest).
- The OEDP shall be delivered along with the source code stored in a Git-like repository, allowing for easy customizations and further updates.
- The source code repository shall be capable of submitting tickets and bug reports so that CEPA and other users can submits any issues in a transparent and organized way.
- The architecture of the OEDP shall be capable of pulling in metadata from other data portals, allowing to create a federated network of data portals which share data between each other.
- The OEDP shall be capable of recording predefined as well as custom metadata for each dataset.
- The portal shall have an intuitive web interface, which allow publishers and curators to easily register, update and refine datasets.
- The OEDP shall provide a rich search experience, including a quick 'Google-style' keyword search as well as faceting by tags and browsing between related datasets.
- It is recommended to use CKAN to deploy the OEDP. This will ensure that the portal is compatible with other national and international Open Data portals. Using existing open source software will also help reduce costs of the implementation and ensure greater community support.
- The User Interface of the OEDP need to be customizable to reflect PNG's national colors, culture etc.
- **The detailed scope of the requirements for the OEDP will be delivered in a form of the Terms of Reference (ToR) for the implementation phase.**
- The key challenge for the OEDP will be to ensure local stakeholders engagement (this will be addressed in more details in the next report: Short and Medium Term Action Plan for EMIS).
- The development of the OEDP need to be coordinated with other projects that share some common goals, e.g.: PNG data4development (<http://www.pngdata4development.org.pg/>) or the Inform Project coordinated by the Secretariat of the Pacific Regional Environment Programme (SPREP)¹¹.

¹¹ <https://www.sprep.org/inform/home>

6 Summary

1. The EMIS shall not be regarded as one single software system that can address all of the problems that CEPA is currently facing in terms of environmental information management and reporting. Instead, the EMIS should be regarded as a long-term ICT strategy aiming for a more effective information management.
2. As such, the EMIS shall focus on both internal and external challenges related to data and information management.
3. The internal challenges span from establishing a proper IT infrastructure, rolling our document management system, integrating existing databases through developing new systems to tackle some of the still unaddressed areas, such as Environmental Permit management.
4. Based on the analysis of the PNG's National Reports to the Convention on Biodiversity, only 15% of data comes from CEPA's internal databases, while the remaining 85% need to be acquired from external stakeholders, both national and international.
5. Developing the Open Environmental Data Portal will facilitate the exchange of environmental information between various external stakeholders.
6. Additional benefits of the OPEDP for CEPA will be:
 - a. The EMIS project will meet its objectives and the portal can be easily delivered within the required deadlines,
 - b. Support for E-Government, Sustainable Development Goals,
 - c. CEPA will become a leader of the Open Data initiative in PNG.
7. The OEDP will support other national and regional initiatives, such as PNG Data-4-Development project as well as the Inform Project coordinated by the SPREP.
8. Open Environmental Data Portal will "support CEPA to address the convention obligation related to reporting requirements under the UN Convention on Biological Diversity". It will also "strengthen the availability, storage and use of environmental information".
9. The next deliverables within this Project are:
 - a. Deliverable 03 – Short and Medium Term Action Plan for implementing the OEDP
 - b. Deliverable 04 – the Terms of Reference for the development of the OEDP.