INTOSAI
Working Group
on Environmental
Auditing

INTOSAI WGEA Work Plan 2017-2019

Project 2.3 (g)

Training Tool on Environmental Data: Resources and Options for Supreme Audit Institutions

Further information contact: wgea@bpk.go.id

Website: www.wgea.org

Twitter ID: @WGEASecre

FOREWORD & ACKNOWLEDGEMENT

We are happy to present the "Training Tool on Environmental Data: Resources and Options for Supreme Audit Institutions". The Training Tool is being developed under Work Plan 2017-19 of INTOSAI Working Group on Environmental Auditing (WGEA). The project subcommittee to develop this training tool comprised SAI India (leader) and SAI USA (member).

1. **Background**

Collecting and analyzing environmental data is often a critical step in conducting environmental audits. The Working Group on Environmental Auditing (WGEA)'s Sixth Survey on Environmental Auditing revealed that Supreme Audit Institutions (SAIs) are conducting an increasing number of environmental audits. The survey also revealed that most common obstacles to developing and conducting these audits include insufficient data on the environment and insufficient monitoring and reporting systems.

The research paper "Environmental Data: Resources and Options for Supreme Audit Institutions", prepared by the INTOSAI Working Group on Environmental Auditing (WGEA) in 2013, discusses ways in which SAIs use environmental data; some key sources of environmental data that are available to audit institutions at the global, regional, and other levels, as well as key considerations when using such data along with variety of tools and methods that audit institutions can use when high-quality environmental data are lacking. In its work plan for 2017-19, INTOSAI WGEA embarked upon developing a training tool on Environmental Data.

Information Technology has made inroads in all fields, including environment. Latest addition in this area is Data Analytics and big data. Such data is extremely versatile and has almost become a necessity for many different environmental audit needs. A training tool encompassing certain inputs regarding GIS, Remote Sensing, Computer Assisted Auditing Techniques, etc. was therefore considered useful to Supreme Audit Institutions (SAIs).

Making of this training toolkit involved survey and compilation of recent developments and practices regarding environmental data and data analytic tools being used in various SAIs. Survey was designed to identify, collect / write appropriate case studies on the lines of earlier INTOSAI WGEA research on Environmental Data duly updated to address contemporary needs of SAIs.

We received valuable feedback from 46 SAIs through survey. Around one third (35 per cent) of the responding SAIs have used some kind of new / non-conventional techniques, tools and software like GIS, Remote Sensing, Data Analytics, etc. to analyse the environmental data while conducting various Environmental Audits since 2011 and afterwards. The responding SAIs have mentioned instances and Audit Reports where such new techniques were used for environmental audits. The survey also revealed the need for focusing on training on use of GIS software as only 12 per cent responding SAIs mentioned competence in using GIS software such as QGIS, ArcGIS, etc. during environmental audits. We have also increased the proposed duration of training from three days to five days as preferred by two third (67 per cent) of responding SAIs. On the basis of responses to this survey, we have attempted to balance the contents of the Training Toolkit by including theory, practical examples and hands on training based on environmental data in digital form.

This training toolkit is an attempt to put together essential domain knowledge into convenient modules with emphasis on using data analytic tools for environmental audit. The way the toolkit is made, it has “built in capability” to customise it for using relatable case studies which could become available in future. The development of the toolkit itself has used experience gained by three SAIs from three continents and this effort enhances the confidence of toolkit developers in their capacity to keep it dynamic and evolving in years to come.

Technology and data analysis will be an important enabler of public sector auditing in future. INTOSAI has increasingly recognised the importance of embracing use of digital data and tools for public audit. Environment is one such sector where there would be lot of scope for exploring the availability of digital data and development of tools for audit of the same. This training toolkit is expected to facilitate quick orientation of Environmental auditors to use of data analytics. It is, however, pertinent to note that this training toolkit is not intended to impart technological proficiency to acquire expertise. It simply enables demystifying the field of environmental data and opens a window to the trainee to acquire some basic skills.

2. Objectives

The development of Training Tool has been taken up with objective of providing the trainees:

i. Acquaintance with earlier research work of WGEA titled Environmental Data: Resources and Options for Supreme Audit Institutions
ii. Insight into other contemporary practices and methods being used by various SAIs with regard to data analytics
iii. Understanding of ways in which public auditors use environmental data;
iv. Capacity to broadly identify key sources of environmental data available to SAIs and key considerations when using such data

v. Knowledge to identify tools and methods SAIs may use when high-quality environmental data are lacking

vi. An opportunity to experiment the use of Data Analytics / Computer Assisted Auditing Techniques (CAATs)

vii. An insight into exploring further possibilities for analysis of environmental data in digital form

3. Quality Assurance

While developing the training tool, the project subcommittee has referred to the Guide for Project Leaders: How to Develop INTOSAI WGEA Training Materials. The project features have been presented to the WGEA Steering Committee (SC) and WGEA Assembly. The project leader SAI India has also followed its internal quality assurance mechanism by getting the tool peer reviewed by the Centre for Data Management and Analytics (CDMA) established by SAI India. Feedback and suggestions from members of the INTOSAI WGEA SC also contribute towards quality assurance in development of this Training Tool.

4. Acknowledgement

We are thankful for the continuous guidance and support from INTOSAI WGEA Steering Committee and Secretariat at various stages of the project. We have received useful support while developing the Training Toolkit in the form of approving responses to presentations made in the course of Steering committee meetings of INTOSAI WGEA and its 18th Assembly meeting at Bandung (Indonesia). Special thanks to INTOSAI member SAIs for providing useful inputs in response to survey questionnaire for the development of this Training Toolkit.

Project team:

Supreme Audit Institution of India

On behalf of SAI India following persons participated in the project team.

Sunil Dadhe, Director General,

International Centre for Environment Audit and Sustainable Development (iCED), Jaipur

dgiced@cag.gov.in

Jahangir Inamdar, Director (T&R), iCED, Jaipur

inamdarJ@cag.gov.in / iced@cag.gov.in

Virender Kulharia, Dy. Accountant General

kulhariav@cag.gov.in

On behalf of Government Accountability Office, United States of America following persons participated in the project team.

Barbara Patterson, Assistant Director, Natural Resources and Environment

pattersonb@gao.gov

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Potential trainees, to use this training toolkit, would need knowledge of CAATs and acquaintance with common functions like data extraction, summarizing, aging, stratification, etc. This training toolkit includes interactive sessions with corresponding materials on above aspects and hands on exercises on using GIS, Remote Sensing, CAATs, etc. using illustrative environmental data. The Training Toolkit has balanced theory, practical examples and hands on training based on environmental data in digital form and it is aimed at equipping an Environmental auditor with competence to choose and use appropriate data analytic tools. The Training Toolkit is divided into following four modules:

Module 1: Understanding Earlier Research Work

This module is focused on introduction to concepts like Audit Evidence, Data, Information, etc. The participants would get exposure to the importance of Audit Evidence and Data in Environmental Audit. They would get an overview of the main ways the auditors use environmental data, the sources of such data and key considerations in using such data from the earlier research work of INTOSAI WGEA titled “Environmental Data: Resources and Options for SAIs (2013)”.

Module 2: Introduction to Contemporary Practices and Methods used by SAIs for Data Analytics

This module deals with the analysis of environmental data and various stages the auditor should pass through to come up with strong analysis of the audit evidence in terms of data. It gives participants an overview of basic data analysis concepts; procedures for planning, performing and documenting data analysis. It also introduces some key techniques for interpreting and displaying data.

Module 3: Experimenting Use of Data in Environment Audit

This module is focused on introducing basics of Remote Sensing, Geographical Information System (GIS) and GIS file formats, various open source Remote Sensing data, GLOVIS data and hands on training on Google Earth. This module has three case studies on ‘Land Use Land Cover’, ‘collection and recovery of municipal waste’, ‘organizing waste treatment in oil shale mining and processing and use of GIS in audit planning phase’. These cases would illustrate how SAIs have used environmental data in various environmental audits.

Module 4: Key Sources and Considerations in using Environmental Data

This module is focused on challenges and future directions / trends in use of data for environmental audits. It will give participants an overview of the options available when quality environmental data is not available for audits. The participants would also be made aware of the future potential of technology in using data in environmental audits.

MODULE 1. 1. 0: INTRODUCTION, AUDIT EVIDENCE AND DATA

Session title: Module 1.1.0 : Introduction, Audit Evidence and Data	**Session-at-a-glance**
Session Learning Objective: The session will give participants an overview of scope of the training, the idea about the audit evidence and how data is important as an important audit evidence. Participants will learn about (1) What is Audit Evidence, (2) What are the characteristics of Audit Evidence, (3) Forms of Audit Evidence and methods of obtaining the same, (4) Comparison of Data and Information (5) Specific examples from environmental sector. **Key Points to be covered:** - Audit Evidence is the most crucial part of audit process - Characteristics, forms and methods of collection of Audit Evidence - Application of concepts to environmental data	
Teaching Method	**Time**
Introduction	05 minutes
Brainstorm / Flip-charting	10 minutes
Lecture and slides	15 minutes
Individual exercise and group discussion	15 minutes
Lecture and slides	15 minutes
Case Study and group discussion	20 minutes
Questions and Feedback	10 minutes
	Total time: 90 minutes

Participant Notes

Overview

This module is focused on introduction to the tool and introduction to concepts like Audit Evidence, Data, Information, etc. For the purpose of understanding environmental data[1], it is pertinent to begin from the nature of audit evidence and its characteristics. This would help in keeping the participants conscious about the basic requirements in terms of an appropriate and sufficient audit evidence and its importance in the process of deriving audit conclusions.

Basic Concepts

Key Teaching Point 1 – Training Tool can be used without the expert

This training tool is based on the previous research work of INTOSAI WGEA about Environmental Data with more focus on practical aspects of the work especially about use of Remote Sensing and Geographical Information System (GIS). The tool has four modules which covers the basics about audit evidence, introduction to concepts like RS and GIS and a major part dedicated to practical aspects of it in the form of case studies. The tool concludes with the future directions in use of environmental data by SAIs through tools like Data Analytics[2] and Big Data.

The tool is prepared and presented in such a way that there is no specific need for an expert for delivery of training. The in-house experts of any SAI can go through the content and deliver the training by following the steps provided and if needed with adoption of local / national examples for better impact and understanding.

[1] The research paper “Environmental Data: Resources and Options for Supreme Audit Institutions”, prepared by the INTOSAI Working Group on Environmental Auditing (WGEA) defines Environmental Data as systematically collected qualitative or quantitative information about different components of the environment (e.g., air quality, water quality and quantity, natural resources, ecosystems, and environmental health impacts) or human activities and sectors that affect the environment (e.g. agriculture, waste, and land development).

[2] Data analysis is a broader term that refers to the process of compiling and analysing data in order to present findings to management to help inform business decision making. Data analytics is a subcomponent of data analysis that involves the use of technical tools and data analysis techniques.

Key Teaching Point 2 – What is an Audit Evidence and what is its importance?

Audit evidence

Information used by the auditor in arriving at the conclusions on which the auditor's opinion is based. Audit evidence includes both information contained in the accounting records underlying the financial statements and other information.

Audit evidence is any information used by the auditor to determine whether the information being audited is stated in accordance with established criteria. Two determinants of persuasiveness of evidence are its competence and sufficiency which are discussed in detail in later parts of the presentation.

Importance of Audit Evidence

The auditor shall obtain sufficient and appropriate audit evidence in order to establish audit findings, reach conclusions in response to the audit objective(s) and audit questions and issue recommendations when relevant and allowed by the SAI's mandate. *(ISSAI 3000/106)*

Audit evidence is necessary to support the auditor's opinion and report. It is cumulative in nature and is primarily obtained from audit procedures performed during the course of the audit. Most of the auditor's work in forming the auditor's opinion consists of obtaining and evaluating audit evidence.

Reasonable assurance is obtained when the auditor has obtained sufficient appropriate audit evidence to reduce audit risk (that is, the risk that the auditor expresses an inappropriate opinion when the financial statements are materially misstated) to an acceptably low level.

ISA 330 requires the auditor to conclude whether sufficient appropriate audit evidence has been obtained to reduce audit risk to an acceptably low level, and thereby enable the auditor to draw reasonable conclusions on which to base the auditor's opinion, is a matter of professional judgment.

Key Teaching Point 3 – What are the types of audit evidence?

Audit evidence to draw reasonable conclusions on which to base the auditor's opinion is obtained by performing:

(a) Risk assessment procedures; and

(b) Further audit procedures, which comprise:

(i) Tests of controls, when required by the ISAs or when the auditor has chosen to do so; and

(ii) Substantive procedures, including tests of details and substantive analytical procedures.

Audit evidence obtained from previous audits may, in certain circumstances, provide appropriate audit evidence where the auditor performs audit procedures to establish its continuing relevance.

When designing and performing audit procedures, the auditor shall consider the relevance and reliability of the information to be used as audit evidence.

If information to be used as audit evidence has been prepared using the work of a management's expert, the auditor shall, to the extent necessary, having regard to the significance of that expert's work for the auditor's purposes:

(a) Evaluate the competence, capabilities and objectivity of that expert;

(b) Obtain an understanding of the work of that expert; and

(c) Evaluate the appropriateness of that expert's work as audit evidence for the relevant assertion.

When using information produced by the entity, the auditor shall evaluate whether the information is sufficiently reliable for the auditor's purposes, including as necessary in the circumstances:

(a) Obtaining audit evidence about the accuracy and completeness of the information; and

(b) Evaluating whether the information is sufficiently precise and detailed for the auditor's purposes.

1. Documentary evidence is more reliable than oral evidence, but the reliability varies depending on the source and purpose of the document.
2. Testimonial evidence that is corroborated in writing is more reliable than oral evidence alone.
3. Evidence based on many interviews together is more reliable than evidence based on a single or a few interviews.
4. Testimonial evidence obtained under conditions in which people may speak freely is more reliable than evidence obtained under circumstances in which people may feel intimidated.
5. Evidence obtained from a knowledgeable, credible, and unbiased third party is more reliable than evidence obtained from the management of the audited entity or others who have a direct interest in the audited entity.
6. Evidence obtained when internal control is effective is more reliable than evidence obtained when internal control is weak or non-existent.

7. Evidence obtained through the auditor's direct observation, computation, and inspection is more reliable than evidence obtained indirectly.
8. Original documents are more reliable than copied documents.

Key Teaching Point 4 – What are different processes of collecting audit evidence?

One or more types of audit procedures are used to obtain audit evidence. When selecting audit procedures, it is reasonable to take into account the relationship between the cost of obtaining the audit evidence and the usefulness of the information obtained. Consequently, in forming the assurance engagement opinion/conclusion and/or report, auditors are generally not required to examine all the information available because they can ordinarily reach conclusions using sampling and other means of selecting items for testing.

The following audit procedures may be used as risk assessment procedures, tests of controls or substantive procedures, depending on the context in which they are applied by the auditor.

Physical Examination

Inspection

Inspection involves examining records or documents, whether internal or external, in paper form, electronic form, or other media, or a physical examination of an asset. Inspection of records and documents provides audit evidence of varying degrees of reliability, depending on their nature and source and, in the case of internal records and documents, on the effectiveness of the controls over their production. An example of inspection used as a test of controls is inspection of records for evidence of authorization.

Some documents represent direct audit evidence of the existence of an asset, for example, a document constituting a financial instrument such as a stock or bond. Inspection of such documents may not necessarily provide audit evidence about ownership or value. In addition, inspecting an executed contract may provide audit evidence relevant to the entity's application of accounting policies, such as revenue recognition.

Inspection of tangible assets may provide reliable audit evidence with respect to their existence, but not necessarily about the entity's rights and obligations or the valuation of the assets. Inspection of individual inventory items may accompany the observation of inventory counting.

Observation

Observation consists of looking at a process or procedure being performed by others, for example, the auditor's observation of inventory counting by the entity's personnel, or of the performance of control activities.

Observation provides audit evidence about the performance of a process or procedure, but is limited to the point in time at which the observation takes place, and by the fact that the act of being observed may affect how the process or procedure is performed.

Confirmation

The receipt of a written or oral response from an independent third party. Auditor has client request that the third party respond directly to the auditor.

External Confirmation

An external confirmation represents audit evidence obtained by the auditor as a direct written response to the auditor from a third party (the confirming party), in paper form, or by electronic or other medium. External confirmation procedures frequently are relevant when addressing assertions associated with certain account balances and their elements. However, external confirmations need not be restricted to account balances only.

For example, the auditor may request confirmation of the terms of agreements or transactions an entity has with third parties; the confirmation request may be designed to ask if any modifications have been made to the agreement and, if so, what the relevant details are.

External confirmation procedures also are used to obtain audit evidence about the absence of certain conditions, for example, the absence of a “side agreement” that may influence revenue recognition.

Recalculation

Recalculation consists of checking the mathematical accuracy of documents or records. Recalculation may be performed manually or electronically.

Re-performance

Re-performance involves the auditor’s independent execution of procedures or controls that were originally performed as part of the entity’s internal control.

Analytical Procedures

Analytical procedures consist of evaluations of financial information through analysis of plausible relationships among both financial and non-financial data. Analytical procedures also encompass such investigation as is necessary of identified fluctuations or relationships that are inconsistent with other relevant information or that differ from expected values by a significant amount.

Inquiry

Inquiry consists of seeking information of knowledgeable persons, both financial and non-financial, within the entity or outside the entity. Inquiry is used extensively throughout the audit in addition to other audit procedures. Inquiries may range from formal written inquiries to informal oral inquiries. Evaluating responses to inquiries is an integral part of the inquiry process.

Responses to inquiries may provide the auditor with information not previously possessed or with corroborative audit evidence. Alternatively, responses might provide information that differs significantly from other information that the auditor has obtained.

Although corroboration of evidence obtained through inquiry is often of particular importance, in the case of inquiries about management intent, the information available to support management's intent may be limited. In these cases, understanding management's past history of carrying out its stated intentions, management's stated reasons for choosing a particular course of action, and management's ability to pursue a specific course of action may provide relevant information to corroborate the evidence obtained through inquiry.

In respect of some matters, the auditor may consider it necessary to obtain written representations from management and, where appropriate, those charged with governance to confirm responses to oral inquiries.

Source: International Standard on Auditing 500: Audit Evidence

Key Teaching Point 5 – What are the characteristics of Audit Evidence?

Two determinants of persuasiveness of evidence are its competence and sufficiency which are discussed in detail in later parts of the presentation.

1. Competence – the degree to which evidence can be considered trustworthy.
   i. Relevance – must pertain to the audit objective being tested.
   ii. Independence – evidence from outside the client is a stronger form of evidence
   iii. Effectiveness of client internal controls – good internal controls can mean better information.
   iv. Auditor direct knowledge – auditor determinations are stronger that client comments.
   v. Qualifications – individual is a qualified source.
   vi. Degree of objectivity – objective evidence is stronger than subjective evidence.
   vii. Timeliness – balance sheet account evidence is better when it is collected around the date of the financial statement. Income statement evidence should sample entire period.
2. Sufficiency – amount of evidence is enough to form a reasonable opinion

   It is a measure of the quantity of evidence used to address the audit objectives and support the audit findings and conclusions.

The following rules of thumb have proven helpful in judging the appropriateness of evidence:

1. Documentary evidence is usually better than testimonial evidence.

2. Audit evidence is more reliable when the auditor obtains consistent evidence from different sources or of a different nature (e.g., testimonial evidence that is corroborated by other sources is better than testimonial evidence alone).
3. Receiving and reviewing an original document is better than receiving a photocopy.
4. Evidence from credible third parties may be better than evidence generated within the audited organization.
5. The quality of information generated by the audited organization is directly related to the strength of the organization's internal controls (the auditors should have a good understanding of internal controls as they relate to the objectives of the audit).
6. Evidence generated through the auditor's direct observation, inspection, and computation is usually better than evidence obtained indirectly.

Important factors to consider in making the audit judgments by auditors include the quality of the evidence (its relevance, reliability, and validity); the level of materiality or the significance of the observation or conclusion

Relevance

Relevance deals with the logical connection with, or bearing upon, the purpose of the audit procedure and, where appropriate, the assertion under consideration. The relevance of information to be used as audit evidence may be affected by the direction of testing.

For example, if the purpose of an audit procedure is to test for overstatement in the existence or valuation of accounts payable, testing the recorded accounts payable may be a relevant audit procedure. On the other hand, when testing for understatement in the existence or valuation of accounts payable, testing the recorded accounts payable would not be relevant, but testing such information as subsequent disbursements, unpaid invoices, suppliers' statements, and unmatched receiving reports may be relevant.

A given set of audit procedures may provide audit evidence that is relevant to certain assertions, but not others. For example, inspection of documents related to the collection of receivables after the period end may provide audit evidence regarding existence and valuation, but not necessarily cut-off.

Similarly, obtaining audit evidence regarding a particular assertion, for example, the existence of inventory, is not a substitute for obtaining audit evidence regarding another assertion, for example, the valuation of that inventory. On the other hand, audit evidence from different sources or of a different nature may often be relevant to the same assertion.

Tests of controls - are designed to evaluate the operating effectiveness of controls in preventing, or detecting and correcting, material misstatements at the assertion level. Designing tests of controls to obtain relevant audit evidence includes identifying conditions (characteristics or attributes) that indicate performance of a control, and deviation conditions which indicate departures from adequate performance. The presence or absence of those conditions can then be tested by the auditor.

Substantive procedures - are designed to detect material misstatements at the assertion level. They comprise tests of details and substantive analytical procedures. Designing substantive procedures includes identifying conditions relevant to the purpose of the test that constitute a misstatement in the relevant assertion.

Reliability

The reliability of information to be used as audit evidence, and therefore of the audit evidence itself, is influenced by its source and its nature, and the circumstances under which it is obtained, including the controls over its preparation and maintenance where relevant. Even when information to be used as audit evidence is obtained from sources external to the entity, circumstances may exist that could affect its reliability.

For example, information obtained from an independent external source may not be reliable if the source is not knowledgeable, or a management's expert may lack objectivity. While recognizing that exceptions may exist, the following generalizations about the reliability of audit evidence may be useful:

- The reliability of audit evidence is increased when it is obtained from independent sources outside the entity.
- The reliability of audit evidence that is generated internally is increased when the related controls, including those over its preparation and maintenance, imposed by the entity are effective.
- Audit evidence obtained directly by the auditor (for example, observation of the application of a control) is more reliable than audit evidence obtained indirectly or by inference (for example, inquiry about the application of a control).
- Audit evidence in documentary form, whether paper, electronic, or other medium, is more reliable than evidence obtained orally (for example, a contemporaneously written record of a meeting is more reliable than a subsequent oral representation of the matters discussed).
- Audit evidence provided by original documents is more reliable than audit evidence provided by photocopies or facsimiles, or documents that have been filmed, digitized or otherwise transformed into electronic form, the reliability of which may depend on the controls over their preparation and maintenance.

Validity

It refers to the extent to which the evidence is a meaningful or reasonable basis for measuring what is being evaluated. In other words, validity refers to the extent to which the evidence represents what it is purported to represent

Link between types of audit evidence and different methods used in performance audits

Audit evidence	Methods of data collection
Testimonial evidence	Interviews Surveys, questionnaires Focus groups Reference groups
Documentary evidence	Document review File reviews Using existing statistics Using existing databases
Physical evidence	Observation of people Inspection of objects or processes Experiments, e.g. level of computer data security
Analytical evidence	For instance: Quantitative data collection methods. DEA analysis, regression analysis. Computations, comparisons, separation of information into components, and rational arguments

Key Teaching Point 6 – Data and Information in Environmental Sector

Definition of Data:

"A representation of facts, concepts or instructions in a formalized manner suitable for communication, interpretation, or processing by humans or by automatic means."

Objective view-

This objective point of view accepts that the data are validated in the sense that they must be measured and recorded accurately. Certain logic checks as to the reasonableness of the data can be done to try to determine whether the data capturing instruments or processes have failed. Data can be shown to be true if they correspond to reality. The objective view tends to assume that all data processing will be automated.

Subjective view-

Natural language: facts given, from which others may be deduced, inferred.

Info. Processing and computer science: signs or symbols, especially for transmission in communication systems and for processing in computer systems, usually but not always representing information, agreed facts or assumed knowledge; and represented using agreed characters, codes, syntax and structure.

The subjective view is very different in that it emphasises that if data are processed using a computer, the output is still only more highly structured or reformatted data.

Intersubjective view:

Data are formalized representations of information, making it possible to process or communicate that information. The intersubjective view allows for the possibility that data may be processed either by computer or directly by a person.

The data can be classified in various ways like primary-secondary, qualitative – quantitative, internal-external etc.

Characteristics of Data

- Data are the raw materials of information
- Data is a distinct piece of information
- Data must be disorganized or unprocessed

Information

- Information is described as that form of data which is processed, organised, specific and structured, which is presented in the given setting. It assigns meaning and improves the reliability of the data, thus ensuring understandability and reduces uncertainty. When the data is transformed into information, it is free from unnecessary details or immaterial things, which has some value to the researcher.
- Information is the simplified form of data.
- When the data is processed or organized it becomes information.

Objective view

Information has a meaning ... (it) comes from selecting data, summarizing it and presenting it in such a way that it is useful to the recipient.

Subjective view

Information usually implies data that is organized and meaningful to the person receiving It. Data is therefore raw material that is transformed into information by data processing. Information can be defined in terms of its surprise value. It tells the recipient something he Did not know.

Intersubjective view

To produce information, we have to interpret what we experience and make explicit what We know.

Characteristics of information

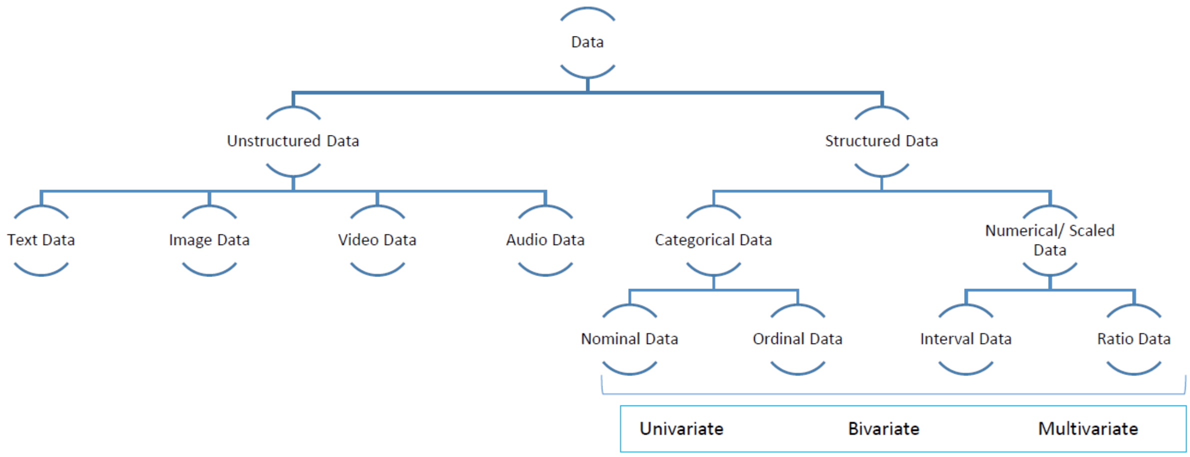
- Information is the resultant version of some data.
- Information always be processed or organized
- Information is the context in which data is taken

Examples of Data and Information

1. The history of temperature readings all over the world for the past 100 years is data. If this data is organized and analyzed to find that global temperature is rising, then that is information.

2. The number of visitors to a website by country is an example of data. Finding out that traffic from one country is increasing while that from another is decreasing is meaningful information.

3. Often data is required to back up a claim or conclusion (information) derived or deduced from it. For example, before a drug is approved by a regulator, the manufacturer must conduct clinical trials and present a lot of data to demonstrate that the drug is safe.



CLASSIFICATION OF DATA

Key Differences between Data and Information

1. Raw facts gathered about a condition, event, idea, entity or anything else which is bare and random, is called data. Information refers to facts concerning a particular event or subject, which are refined by processing.
2. Data are simple text and numbers, while information is processed and interpreted data.
3. Data is in an unorganized form, i.e. it is randomly collected facts and figures which are processed to draw conclusions. On the other hand, when the data is organised, it becomes information, which presents data in a better way and gives meaning to it.
4. Data is based on observations and records, which are stored in computers or simply remembered by a person. As against this, information is considered more reliable than data, as a proper analysis is conducted to convert data into information by the researcher or investigator.
5. The data collected by the researcher, may or may not be useful to him, as when the data is gathered, it is not known what they are about or what they represent? Conversely, information is valuable and useful to the researcher because it is presented in the given context and so readily available to the researcher for use.

6. Data is not always specific to the need of the researcher, but information is always specific to his requirements and expectations, because all the irrelevant facts and figures are eliminated, during the transformation of data into information.
7. When it comes to dependency, data does not depend on information. However, information cannot exist without data.

Summary

The auditor has to obtain sufficient and appropriate audit evidence in order to establish audit findings, reach conclusions in response to the audit objective(s) and audit questions and issue appropriate recommendations.

Throughout this module, participants will learn about (1) training tool, (2) Audit Evidence, (3) Forms of Audit Evidence, (4) Methods of collection of Audit Evidence, (5) Characteristics of Audit Evidence and (6) Difference between Data and Information. All this would be supported by examples from the environmental sector.

Instructor's Notes – Module 1: Introduction, Audit Evidence and Data

<u>Introduction</u> (05 minutes): Show slides 1 - 3 and introduce the session overview, training tool

[Notes in power point slides provide instructor's notes]

<u>Brainstorming / Flip-charting</u> (10 minutes): Show slide 4 and hold group brainstorming, record responses provided by the class on a flipchart for reference throughout the class.

[Notes in power point slides provide instructor's notes]

<u>Lecture</u> (15 minutes): Show slides 5 - 6

[Notes in power point slides provide instructor's notes]

<u>Individual Exercise and group discussion</u> (15 minutes): Conduct individual exercise with group discussion afterwards.

[Notes in power point slides provide instructor's notes]

<u>Lecture</u> (15 minutes): show slides 8 -9

[Notes in power point slides provide instructor's notes]

<u>Case Study and group discussion</u> (20 minutes): Provide a one or more case studies and ask to identify the forms of evidence, methods of collection, characteristics of evidence and the overall weakness and / or strength of audit observations from the point of view of audit evidence

[Notes in Power Point slides provide instructor's notes]

<u>Summarization, Questions and Feedback</u> (10 minutes)

[Notes in Power Point slides provide instructor's notes]

MODULE 1.2.0: DATA USAGE IN ENVIRONMENTAL AUDITS, SOURCES AND KEY CONSIDERATIONS

Session title: Module 1.2.0: **Data Usage in Environmental Audits, Sources and Key Considerations**	**Session-at-a-glance**

Session Learning Objective:

The session will introduce the participants to the earlier work of INTOSAI WGEA on environmental data. This would be summarization of what SAIs have been doing in respect of environmental data. Participants will learn (1) How SAIs use data for planning and conducting environmental audits (2) What are different sources of environmental data, (3) Characteristics of environmental data, (4) Assessment of quality of environmental data and (5) International Standards for data collection

Key Points to be covered:

- Uses of Environmental Data
- Sources of Data
- Characteristics and Quality of Data
- Assessment of quality of data

Teaching Method	**Time**
Introduction	05 minutes
Brainstorm / Flip charting	10 minutes
Lecture and slides	25 minutes
Individual exercise and group discussion	15 minutes
Lecture and slides	10 minutes
Case Study and group discussion	20 minutes
Feedback and questions	05 minutes
	Total time: 90 minutes

Participant Notes

Overview

This module is focused on data analysis. Data analysis is not a separate task from everything else involved with planning and design of an audit. The analysis needs to be fully integrated with the rest of the audit. It needs to make sense in terms of answering the audit questions, using the appropriate information the auditors have collected, and leading to the types of statements the auditors expect to make in the product.

Basic Concepts

Key Teaching Point 1 – Use data for planning and conducting environmental audits

Introduction to INTOSAI WGEA work on Environmental Data

According to the Working Group on Environmental Auditing (WGEA)'s Sixth Survey on Environmental Auditing, Supreme Audit Institutions (SAIs) are conducting more and more environmental audits. SAIs often face data challenges when conducting audits, including insufficient data on the state of the environment and insufficient monitoring and reporting systems. These data challenges can affect a SAI's ability to conduct audits and develop meaningful findings, conclusions, and recommendations.

WGEA Work Plan 2011–2013 identified environmental data as a research topic. The Office of the Auditor General of Canada (OAGC) and the United States (US) Government Accountability Office (GAO) led the research effort, with subcommittee members from Botswana, Estonia, Namibia, New Zealand, Poland, and Tanzania.

Program managers generate or use environmental data for purposes such as:

• *Develop predictive environmental models to compare program alternatives.*

• *Monitor and enforce compliance with environmental regulations.*

• *Determine the environmental effects from planned projects or from completed or abandoned projects.*

• *Inform scientific or regulatory decisions.*

Because the primary function of SAIs is government oversight, SAI often have purposed for environmental data that are related to but distinct from those of program managers. However, an auditor might use the same data to determine the extent to which the government had taken steps to ensure compliance, and so might focus on the success of enforcement measures.

Main ways that SAIs Use Environmental Data in Audits

SAIs have used environmental data to

•plan audits, including selecting topics, samples, or case studies;

•conduct audits, including assessing how well governments manage programs, environmental risks, and environmental data; and

• provide context for audit findings and recommendations.

USING ENVIRONMENTAL DATA TO PLAN AUDITS

Some SAIs have used environmental data to identify potential audit topics or to modify the audit's scope based on the data available.

e.g.

SAI Bhutan

SAI Bulgaria

USING ENVIRONMENTAL DATA TO CONDUCT AUDITS

When conducting audits, SAIs have used environmental data to:

1. to assess the progress of environmental programs toward targets and goals, including compliance with laws and treaty obligations;
2. to evaluate practices for assessing and managing environmental risks, including risks arising from environmental emergencies such as toxic waste spills; and
3. to evaluate government practices for managing environmental data, including data collection and reporting, and to develop findings on how inadequately managed data affect government programs.

Assess progress of environmental programs toward targets and goals

1. SAIs have used environmental data to measure progress toward the goals of national environmental programs. Ex- SAI China

2. SAIs have also used environmental data to assess how well governments have complied with environmental laws and treaty obligations. For example, the SAI of the United Kingdom (UK)

3. SAIs have also used data to measure compliance at the national level. Ex- SAI of India

Evaluate practices for assessing and managing environmental risks

1. SAIs have used environmental data to assess a government's knowledge of environmental risks and monitoring capabilities. For example, the SAI of Canada

2. SAIs also sample data to determine how governments manage environmental degradation. For example, the SAI of Latvia.

3. Environmental data can also help evaluate how government entities manage environmental risks, such as those linked to hazardous waste. For example, the SAI of the US

4. SAIs have also used environmental data to evaluate how well a government prepares for and responds to environmental emergencies and manages the related financial costs. For example, the SAI of Norway

Evaluate government practices for managing environmental data

Using the approach of examining the quality or quantity of the data that managers generate, and study how the data are produced and managed\, SAIs can assess the quality of information that is available to managers for program-related decisions. Ex- SAI of Bulgaria

Using Environmental Data to Provide Context for Audit Findings

SAIs have used environmental data to provide context for and highlight audit findings, such as the impacts of program weaknesses. SAIs have also used environmental data describing the relative performance of other governments to provide context for audit results.

Key Teaching Point 2 - different sources of environmental data

Sources of Environmental Data for SAIs

SAIs may draw on different environmental data sources, depending, in part, on the way they plan to use the data.

National Data Sources

National data sources provide information on the characteristics of a single country. Such data sources include the data that auditors would normally obtain from audited entities, but could also include data from other entities. For example, national statistical agencies may collect data about the industries that are subject to environmental regulations.

SAIs may find it difficult to use data from other entities if the data from various sources are defined and collected using different conceptual frameworks[3]. Some of the key entities that may have information relevant to a given environmental issue are natural resource departments, that is, those entities responsible for land management, agriculture, forestry, and fisheries. SAIs also have the following options as given below-

1. National statistical agencies may also produce environmental data. There may be advantages to using these data because statistical agencies are politically independent and place a high priority on quality control.

2. SAIs may find it useful to consider information from other levels of government, such as regions or municipalities within the country, court records, and media reports.

3. SAIs may obtain national information from third-party non-government sources, such as Industry, academic institutions, specialized consultants, or environmental non-governmental organizations (NGOs).

4. The availability of data in electronic form is particularly important for environmental data, especially when combining the data with spatial information, however SAIs may find that relevant national data sources are not available digitally.

5. By drawing on multiple data sources, SAIs can compare information to determine if data are consistent.

Regional Data Sources

Regional data sources provide environmental data for a specified geographic region or economic-political association. Such data sources may be based on geophysical or ecosystem divisions, such as countries that border a single body of water (e.g. the

[3] INTOSAI WGEA Guidelines Environment Data: Resource and Options for SAIs, Pg 25

Caribbean), and are usually aggregated from national sources. SAIs may use regional data sources when conducting transboundary co-operative audits if certain issues cross national boundaries. SAIs can also use such sources to compare the performance of a given country with that of its neighbors. SAIs may pursue the following sources for accessing the regional environmental data[4].

1. SAIs can get regional environmental data from a variety of public and private sources. Environmental data may be collected and organized based on formal or voluntary political or economic associations.

2. International organizations have also created some regional data sources. For example, the United Nations Environment Programme (UNEP)

3. Regional data sources can provide SAIs with valuable points of comparison in similar countries.

4. SAIs may also benefit from data from adjoining countries, because regional sources may provide information on components of the environment that are shared across national borders. Such components include river flows, water quality in shared water bodies, air quality, and migratory bird and animal populations. Regional data can provide a basis for regional coordinated audits

5. When specific agreements exist between countries to manage shared environmental issues, regional data sources can be used to evaluate the performance of the countries that are parties to the agreements.

Global Data Sources

Global data sources provide data covering most of the world, usually separated by country. If environmental issues touch countries in several parts of the world, SAIs may find that global data sources are the most useful as a point of comparison. Examples of such global issues include climate change, control of ozone-depleting substances, and regulation of persistent organic pollutants.

[4] INTOSAI WGEA Guidelines Environment Data: Resource and Options for SAIs, Pg 24-26

Most of the global data sources are originated from national governments, which means that any inaccuracies or inconsistencies in national data will affect these data sources. In some cases, however, the data are produced independently of national governments.

SAIs tend to use global data sources for context or background to an audit topic and assess how well audited entities complied with international agreements and comparing the audited entity with similar entities in other countries to identify good practices that might apply to the audited entities.

Spatial Data Sources

Spatial data, also called geospatial data or geographic information, describe the distribution of phenomena on the surface of the earth. Such data are usually not aggregated to a single national average value, which means that SAIs can use such data to identify specific geographic locations within a country where environmental problems are most severe. Spatial data may be reported in national, regional, or global data sources.

1. The data in these sources, which include satellite observations or data from GIS, may be collected uniformly across administrative or political boundaries. The best-known examples of such data sources are probably the maps and satellite images available from information providers such as Google.

2. SAIs may find spatial data sources particularly useful when they are examining environmental issues that have an explicit geographic aspect, such as establishing protected areas, or monitoring the distribution of air pollutants and identifying pollution sources. Spatial data sources can also be used to select samples from different regions, to identify high risk areas, and to identify patterns in the data that may not be evident without a spatial presentation.

3. SAIs may also use spatial data to analyze the regions they are considering from several different perspectives or to assess the combined effects of different environmental factors. SAIs may also choose to use spatial data sources to report audit results, to help make the findings more tangible.

4. The resolution of spatial data such as satellite images is a key consideration for SAIs when deciding how or whether to use such data sources. The resolution will determine what kinds of physical features, or changes in them, one can expect to

detect. SAIs need access to the tools and expertise to manipulate the data in a knowledgeable way in order to use Spatial Data.

Complementary Information and Tools

It helps understand the data and its sources

- Definitions
- Documentation of methods used for data set development
- Techniques used to collect data
- Tools to support use of data

Key Teaching Point 3 – Characteristics of environmental data

Key Considerations when Using Environmental Data

A common proposed framework for data quality was resulted by an international group of experts which identified the elements of a generic national quality assurance framework, including key characteristics of high-quality data, such as relevance and accuracy. SAIs have similar criteria for evaluating the quality of the data they use. International audit standards for evidence, such as INTOSAI's Performance Audit Guidelines: ISSAI 3000-3100. The characteristics of high-quality data developed for the UNSD may influence the appropriateness of the information, and hence what evidence is sufficient for audit purposes.

Characteristics of High-quality Data and Additional Data Quality Considerations for SAIs	
Characteristics of High-quality Data	Additional Data Quality Considerations for SAIs
Relevance. Relevance reflects the degree to which the information meets the needs or requirements of clients, users, stakeholders, or the audience.	SAIs may use data for purposes different from those of the program manager who generated the Data. As a result, the relevance of the data may Differ for the program manager and the SAI.
Accuracy and reliability. Accuracy reflects the degree to which the information correctly describes the phenomenon it was	Audit organizations may wish to pay particular attention to situations in which the organization generating the data manages the data and

designed to measure. Reliability concerns whether the data consistently over time measure the reality that they are designed to represent.	may have an economic or political interest in matters related to the data, such as in minimizing reported environmental damage. How environmental data are collected and reported, and by whom, may affect their reliability.
Timeliness and punctuality. Timeliness refers to how fast—after the reference date or the end of the reference period—the data are released or made available. Punctuality refers to whether data are delivered on the dates promised, advertised, or announced.	SAIs may be interested in whether the data are available quickly enough to support the management decisions that will rely on them, as in the case of disasters when a rapid response is essential. They may also find that there is a trade-off between timeliness and accuracy and eligibility.
Accessibility and clarity. The data and metadata can be found or obtained without difficulty, are presented clearly and in such a way that they can be understood, are available and accessible to all users on an impartial and equal basis in various convenient formats, and are affordable, if not offered free of charge.	Some data may not be available in electronic form, creating challenges for some SAIs. Language or technical requirements for using the data may limit how auditors use some data sources, such as complex GIS databases. Auditors may also face legal barriers to using or reporting some kinds of environmental data, such as proprietary data.
Coherence and comparability. The data are consistent internally and over time and are produced using common standards with respect to scope, definitions, classifications, and units. Users should be able to combine and make joint use of related data from different sources.	Environmental data may not be collected in the same way at all locations, which may pose problems for SAIs if more than one country is involved in a transboundary audit. The use of international standards can help address this challenge. For some global environmental concerns, such as persistent organic pollutants regulated under the Stockholm Convention, comparable measurements are essential to obtain a clear picture of the international distribution of these pollutants
Availability of metadata. Users should have access to information concerning the underlying concepts, variables and classifications used; the methodology	SAIs may be interested in metadata underlying the environmental data they are considering using, as well as the quality controls applied when environmental data were collected and

used to collect and process data; and indications of the quality of the data—in general, sufficient information to understand all of the attributes of the data, including their limitations, for informed decision making.	analyzed. This information may help them to determine how much they can rely on the data. Information about quality controls can also help SAIs assess the systems the audited entity uses to obtain and manage environmental data.

Key Teaching Point 4 – Assessment of quality of environmental data

How SAIs can Assess Data Quality?

To determine the extent of the data assessment needed and to use limited resources efficiently, SAIs need to consider the following key factors:

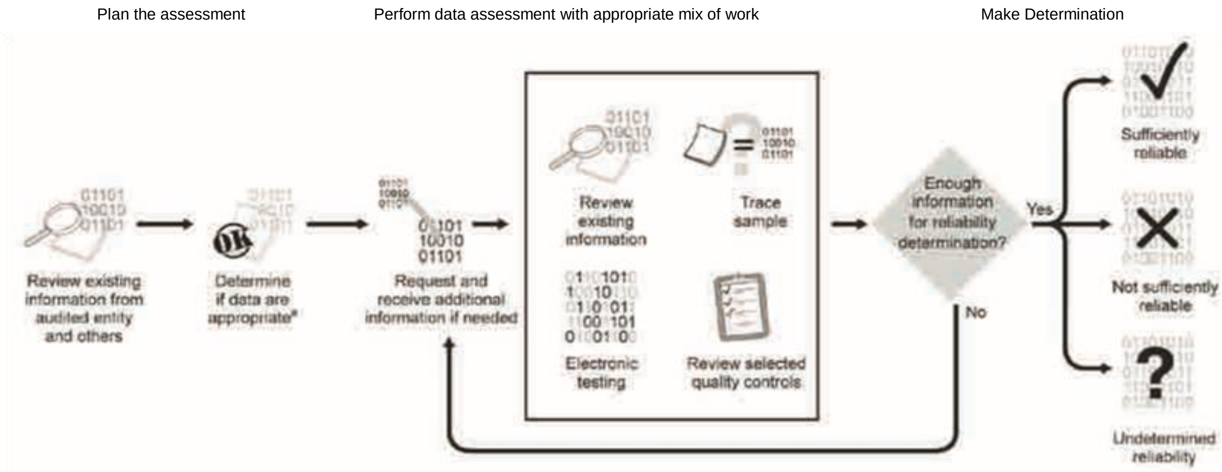
1. The data's use and expected importance in the final report. Thorough data assessment required, if the data are used solely for background context or to identify potential audit topics.
2. Data reliability results from previous examinations of the data provider's or audited entity's records and quality control practices.
3. Whether complementary evidence reinforces or contradicts the findings developed using the data.
4. Sensitivity of the topic and degree of risk associated with using the data.

SAIs can undertake a variety of activities to conduct a data assessment, such as

1. reviewing existing information about the data, such as external data assessments, reports, or studies;
2. conducting interviews with officials responsible for managing the data to determine such things as the quality of the underlying data (metadata), and how and when officials entered the data into databases or records management systems;
3. performing tests on the data to identify the extent of inconsistencies, completeness, or errors;
4. tracing data in computer systems to and from source documents; and

5. Reviewing selected internal quality controls for data or records management systems.

Framework for Data Reliability assessment process



As per INTOSAI standards when computer-processed data are significant to the findings of the audit, the SAI may want to take extra precautions to obtain sufficient and appropriate evidence that the data are accurate and reliable. These standards also state that if the reliability of an information system is the primary objective of an audit, the audit team should review the system's quality controls.

Once the data assessment is complete, SAIs will use professional judgment to determine whether the data are sufficient and appropriate for the audit's purpose within the context of the overall audit. Different audit teams evaluating the same data may make different determinations about whether to use the data in their reports, depending on the characteristics of the audit and its associated risks.

Key Teaching Point 5 – International Standards for data collection

Relevant International Data Collection Standards for Assessing Data Quality

Standards for data quality help ensure that data sources are accurate and consistent among related data sets. These standards may be developed and applied at various stages of data collection and management. Standards may also be set at many different geographic scales, from individual monitoring sites to global data compilations.

Standards are particularly important for organizations that are responsible for collating and combining data from several different sources (e.g. the FAO, the OECD, and national and international statistical agencies). Such standards are useful to SAIs, as the standards may provide some assurance of the quality of the specific database.

Frameworks and guidance, such as manuals, may not be mandatory but they list recommended steps, direction, and concepts so that nations can produce quality data. Example- Framework for the Development of Environment Statistics developed by the Environment Statistics Section of the UNSD in 1984 and revised *Guidelines for National Greenhouse Inventories in 2006 issued by IPCC.*

When SAIs are considering the environmental standards that may apply to a particular data source, one challenge is to understand the nature of the standards, in part because they may not cover all aspects of a particular data source. In such cases, SAIs may need to use other methods to determine if the data are suitable.

Another step is to examine the documentation of compliance with the standards. Many standards spell out how compliance is to be documented. This documentation may be particularly valuable for SAIs in assessing the reliability of the data source. As part of this assessment, SAIs may need to determine if the standards have evolved and take account of the standards that were in place when data were collected.

How SAIs can Use Existing Data Quality Assessments for Key Data Sources
Data quality assessments highlight the strengths and limitations of the source. They may help SAIs decide about whether and how to use a particular data source. The assessments can also help auditors determine if the entities they audit have used the data source appropriately. One common type of data quality assessment is peer review.

In other cases, comments on trends in data quality are embedded throughout its conclusions on specific research questions. The report notes, for instance, a lack of clarity in individual countries' methods for developing data on growing timber stock. This would mean that auditors should examine these data carefully when deciding whether and how to use them.

Summary

SAIs have used environmental data to identify pressing environmental issues to audit and to scope audits effectively; to evaluate the program effectiveness, environmental risk management, and environmental data collection and reporting; and to provide context for audit findings that illustrates the significance of the issues.

SAIs have access to a range of national, regional, and global sources of information on the environment, some of which are linked to international standards. Spatial information that cuts across administrative boundaries may also be useful. International statistical agencies have defined key characteristics of high-quality data; however, based in part on international audit standards, SAIs must keep in mind other issues as well.

There are a variety of resources available to SAIs to assess the data quality from key sources, such as international data collection standards, and in some cases, data quality assessments specific to a key source. SAIs must carefully consider how they wish to use environmental data and then assess whether the limitations of the data prevent their intended use.

Instructor's Notes – Module 1.1.2: **Data Usage in Environmental Audits, Sources and Key Considerations**

Introduction (05 minutes): Show slides 1 - 2 and introduce the session overview

[Notes in power point slides provide instructor's notes]

Brainstorming / Flip-charting (10 minutes): Show slide 3 and hold group brainstorming, record responses provided by the class on a flipchart for reference throughout the class.

[Notes in power point slides provide instructor's notes]

Lecture (25 minutes): Show slides 4 - 12

[Notes on power point slides provide instructor's notes]

Individual Exercise and group discussion (15 minutes): Show slides 13 and conduct individual exercise with group discussion afterwards.

[Notes in power point slides provide instructor's notes]

Lecture (10 minutes): show slides 14 - 17

[Notes in power point slides provide instructor's notes]

Case Study and group discussion (20 minutes): Slide 18 – Characteristics and assessment of quality of environmental data

[Notes in power point slides provide instructor's notes]

Feedback and questions (05 minutes): Slide 19

[Notes in power point slides provide instructor's notes]

MODULE 2: PRINCIPLES OF DATA ANALYSIS

Session title: Module 2: Principles of Data Analysis	**Session-at-a-glance**

Session Learning Objective:

The session will give participants an overview of basic data analysis concepts; procedures for planning, performing, and documenting data analysis; and introduce some key techniques for interpreting and displaying data. Participants will learn how to (1) formulate researchable questions answerable with secondary data, (2) get data from an external entity, (3) assess the reliability of data, (4) develop an analysis plan, (5) perform and interpret the analysis, and (6) document the analysis.

Key Points to be covered:

- Data analysis as an iterative process
- Using secondary data
- Six basic steps for data analysis (identified above)

Teaching Method	Time
Introduction	10 minutes
Brainstorm/Flip charting	20 minutes
Lecture and slides	30 minutes
Individual exercise and group discussion	10 minutes
Lecture and slides	30 minutes
Break	20 minutes
Exercise and group discussion	30 minutes
Lecture and slides	50 minutes
	Total time: 200 minutes

Participant Notes

Overview

This module is focused on data analysis. Data analysis is not a separate task from everything else involved with planning and design of an audit. The analysis needs to be fully integrated with the rest of the audit. It needs to make sense in terms of answering the audit questions, using the appropriate information the auditors have collected, and leading to the types of statements the auditors expect to make in the product.

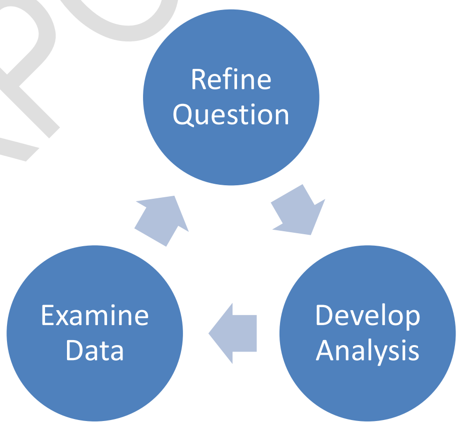
Basic Concepts

<u>Key Teaching Point 1 – Data analysis is an iterative process</u>

1. A basic principle of data analysis is that this is a process that requires careful thought and repeated reconsideration of our approach and our work. For data analysis one needs to think through the plan, assess whether we can do it, start the process, and then go back to reconsider what we originally intended.
2. One needs to recognize that one has to go through this process time and time again during the course of an analysis.

Following steps of activity are required for a preliminary data analysis-

- Refine the research question and discover what data are available
- Develop the analysis and assess the quality of the data
- Examine the data
- Re-refine the research question



For almost all audits, it is expected to go through this loop multiple times, as one learns more about the data and thinks more carefully about what one is trying to say and how it is trying to use the information.

Secondary Data

- Auditors often conduct data analysis on information obtained from others, and on information the auditors collect themselves
- "Primary data" are data auditors gather themselves (e.g., surveys, file reviews)
- "Secondary data" are data gathered by another entity and includes the data collected by the audited agency or any other agency except the data collected by the auditors themselves.

Secondary analyses involve some extra steps and complexity, and because most of the more difficult data analysis challenges involve working with other people's data. The same is also relevant to analysis of primary data.

Examples of Secondary Data

- Data used for administrative purposes (e.g., agricultural statistics, air monitoring results)
- Surveys by agencies for their own specific uses
- Data developed for use in research (e.g., population/demographic surveys)
- Hybrids- Administrative data being used for research (e.g., air monitoring results used in asthma research)

Six Basic Steps for Data Analysis

Six steps are to be followed when planning and conducting data analysis. Data analysis is of "iterative" nature, however. These aren't always simply sequential. Sometimes one learns something in a later step that forces it to go back and reconsider an earlier step.

The following steps organize the work and make sure nothing is slipping through the cracks.

1. Formalizing audit question answerable with secondary data
2. Getting data from another entity
3. Assessing reliability of data
4. Developing an analysis plan
5. Performing and interpreting the analysis
6. Documenting the analysis

Key Teaching Point 2 - Formalizing audit questions answerable with secondary data

Data Analysis and Design

- Data analysis is not a separate activity from audit design and planning work

- The quality of the auditors' analysis depends in part on the quality of the audit design. The audit analysis needs to make sense in terms of answering the objectives, using appropriate information the auditors have collected, and leading the auditors to the types of statements they expect to be able to make in the product.
- Auditors should balance exploratory analysis (that may not directly answer their engagement objectives) with the time it takes to conduct such analyses. While there is always room for some exploratory analysis, the focus should be less on 'interesting' analyses and more on 'useful' analyses.

Audit Questions: Criteria

The audit questions are crucial for determining the work the auditors are going to do. This is perhaps the key factor in engagement planning. The goal is to turn issues that are of interest into focused questions that can be addressed **objectively** and **efficiently** and meet all **auditing standards**.

The auditors are responsible for defining appropriate scope and methods, and for defining and refining the audit questions. Here are some criteria we use to assess audit questions.

- Clear and specific
- Measurable
- Fair and objective
- Policy neutral
- Feasible within time and resource constraints
- Realistic in terms of available data

Clear, specific, measurable

- Can audit question be answered empirically?
- Are there measures available that the auditor can use?
- Are the data solid enough to apply these measures?

Fair, Objective, Policy Neutral

- Is audit question objective and not based on preferences, beliefs, values, or tastes?
- Audit questions should not be subjective: Assess the objectivity of the question and rephrase the question or develop a new one if it is not objective as originally considered. For Ex- "Why isn't the Department of the Environment doing more to stop companies from polluting rivers?"

Key Teaching Point 3 – Getting data from an entity

Steps in Planning the Data Request

- A data request should come after the auditor has done their due diligence about the data including researching what is publically available; what other auditors have used and if there is public information on its reliability or relevance for your purposes.
- The Data Request Letter is the culmination of audit work about the data and should not be the first time you discuss data with the entity.
- Ideally, the team should ask questions about the availability, format and reliability of potential data sources as early as entrance.

Steps in Planning the Data Request

- Address the engagement objectives by
  - Determining what data are needed
  - The availability of the data
  - The relevance and reliability of the data
- Address the processing needs, including data formats and other specifications of the data
- Address the documentation requirements when requesting secondary data, such as data dictionaries

Requesting Data - Items to Include

- First, specifically address the engagement objectives before providing:
  - a very brief summary of the audit
  - the names of data files requested or a description
  - the time period of data requested

  any agreements with the entity about confidentiality or data security issues

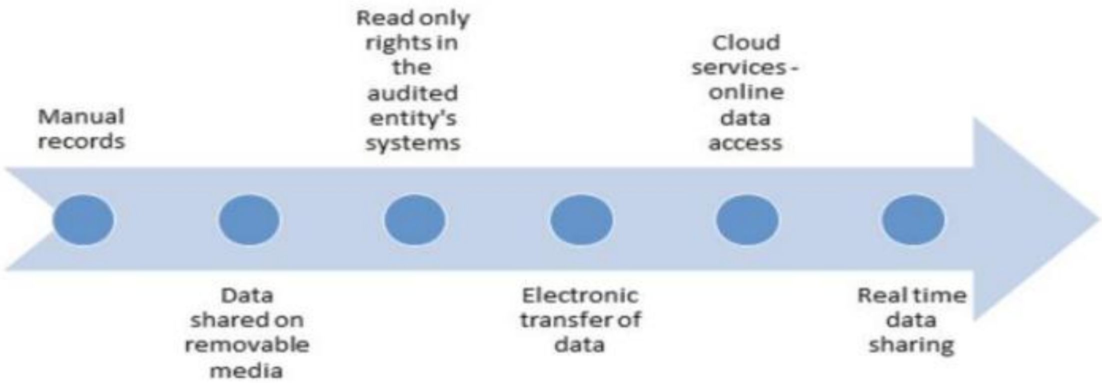
Requesting Data - Items to Include

- Auditors should also specify an agreed upon data format and method of transmission (e.g. email or through other electronic means)
- Issues to consider when agreeing upon the format of the data and mode of transmission are:
  - Sensitivity of the data
  - Size of the data
  - Who will be performing the analytical work
- The data request can be used to ask for documentation related to the data file, including:
  - data dictionary
  - record counts
  - control totals

– record layout
– file creation program

Receiving Data from an Entity

- Like other information, there needs to be a clear audit trail for the data:
  – It is important to save an original copy of the data, along with the email transmission or email confirming receipt.
  – The auditor can create a separate document to perform analysis but one 'raw' version of the data should always be saved. If the entity provides corrected versions of the same data, it can be saved as a new version.
- Auditors should not be hesitant to ask follow up questions on the data
- Data may be provided to the auditors on the entity's sites through access to the system. This can be a read-only access without any transaction rights so that the system's performance is not affected. The data may be provided through backup[7] files created in the entity's environment and shared on a removable media with the auditors. The data may also be shared electronically using electronic transfers though networks - LAN or WAN or internet or a VPN, as the case may be.
- Indicated below is a progression in the way auditors can access data from their audited entities, starting from manual records to online, real time data sharing. However, it is not essential that the progression be sequential and auditors accessing only manual records may start accessing real time data electronically without going through the intermediate steps. The access to data solely depends on the capability of the auditors, the auditing environment and the level of access established between the two.



Access to Data

- One of the ways to deal with data access is through involvement of audit from the design stage of the IT systems when it may be possible to incorporate the data requirements of audit into the system design. This would facilitate acquisition of data in the requisite format. To ensure this, field offices would need to convey the data requirements for audit to the concerned entities at the stage of important system developments thereby facilitating access to requisite data when the system is operational. These data requirements could cover information sets to be acquired, format of data, mode of transfer and periodicity of data to be made available to audit. At the same time, access to the complete system or complete data, if required for any specific audit, such as performance audits, systems audits, IT audits, special audits etc., should not be precluded by involvement of auditors at the system development stage.

Source: Guidelines on Data Analytics (2017) by SAI India.

Key Teaching Point 4 – Assessing reliability of data

Assessing Reliability: Why it is So Important?

- Auditing standards require the use of reliable data
- Credibility of the audit depends on the use and reporting of reliable information. If the data the auditor is using are not good, then the audit results will also not be good. It can result in an audit report that contains incorrect information or a wrong message.
- Garbage in = garbage out
- U.S. GAO has a guidance document on assessing the reliability of computer-processed data that can be helpful in understanding how to assess the reliability of data being considered for an audit: http://www.gao.gov/products/GAO-09-680G

Key Elements of Data Reliability

- Generally, data are considered reliable if:
  - the data entered into a computer system/application are reasonably accurate, complete, and consistent
  - the data, while stored in the computer system/application, are not inappropriately modified or deleted,
  - any computer processing performed on that data is reasonably accurate, complete, and consistent
- Reliability is a relative term (it does not have to be perfect to be reliable)
  - It must always be measured in relation to the auditors' intended use of the data
  - Data reliability is more art than science
- Auditors are only concerned with the reliability of those data elements needed to answer the audit questions
- Auditors need to assess the reliability of computer-processed data in any format, including electronic files downloaded from the Internet, and hard copy reports produced using computerized data

Signs of potentially reliable versus potentially unreliable data

- Potentially Reliable:
  - The entity can provided written data-entry protocols and well-documented data dictionaries

– There are few missing values; data conform to data dictionary

- Potentially Unreliable:
  – Typos or other signs of manual entry without controls

Entity does not use it for the purposes we intend nor does it think it reliable for these purposes

Data Reliability Assessments

- Good data reliability assessments for performance audits:
  – Make use of all existing information about the data
  – Perform at least a minimal amount of data testing
- Data reliability assessments should also:
  – Do only the work necessary to assure the auditors that the data are reliable enough for the intended purposes
  – Incorporate professional judgment
  – Involve the right people with the right skills at key decision points

Key Elements of Data Reliability

- Auditors planning to use computerized data to answer audit questions (not background) need to plan who will do different parts of the data reliability assessment. Steps include:
  – Gathering and reviewing existing information related to the reliability of the data and the system that produces it, and
  – Performing electronic tests on the data, looking for obvious problems with completeness and accuracy.
  – Additional work, such as tracing to source documents or review of internal controls, may be needed for some audits

Key Teaching Point 5 – Developing an analysis plan

What is an analysis plan?

- An analysis plan describes how the data will be analyzed and presented to answer the audit questions
- An analysis plan is developed for audits that rely on computer analyses. As there has been an increase placed on documentation standards for audits over time, analysis plans have received greater attention.

What is the purpose of an analysis plan?

- Translates the audit questions into specific tasks
- Serves as a communication tool between auditors conducting the work

- Helps the auditors “think through” what it wants to do with the data. Thinking through what each proposed analysis will allow the auditors to say often eliminates unneeded work.
- The data analysis plan can be an iterative, ‘living’ document that evolves as research questions are added/removed or refined

What does an analysis plan look like?

- May use various formats, such as
  - Table examples
  - Narrative

- The formality of the data analysis plan can be informed by the importance of the analysis to the message, the number of analysts involved, or the amount of analytical work required

Theoretically,

-- a series of emails or

-- sufficient detail incorporated into the overall audit plan or design matrix could be considered as an analysis plan, as long as labelled as such but they don’t usually contain all the elements needed in an analysis plan.

What is included in an analysis plan?

- The analysis plan is linked to the audit questions, and generally describes:

  - Key assumptions are assumptions/decisions about variables to be used.

  a. One type of key assumption would be choosing among alternative measures.
  b. Use of proxy measures - using variable that is in data to approximate something one wants to measure that is not available
  c. Criteria for development of new variables

  - Data elements to be analysed
  - Selection criteria used
  - Potential data summaries or tabulations to be presented in the product. Could be a narrative

Developing an analysis plan can be an iterative process

- Data analysis plan should address the following:

- Will proposed analysis tasks support what the auditors want to say?
- Do auditors have enough time to complete the tasks?
- Will data permit the proposed analysis or will analysis need to be modified?

– Are data sufficiently reliable for the intended use?

- Revise plan based on the above

Key Teaching Point 6 – Performing and interpreting the analysis

Data analytic approaches

The data after due preparation is analysed to derive insights using various analytic approaches. The following approaches can be used in data analytics:

a. **Descriptive analytics** tries to answer "what has happened". In descriptive analytics, raw data is summarized so that it can be understood by the user. Descriptive Analytics provides an understanding of the past transactions that occurred in the organisation. Descriptive analytics involves aggregation of individual transactions and thus provides meaning and context to the individual transactions in a larger perspective. It involves summarization of data through numerical or visual descriptions.

   Example - Almost all management reporting such as sales, marketing, operations and finance used this type of post-mortem analysis.

b. **Diagnostic analytics** is an advanced form of descriptive analytics and tries to answer the question "why did it happen" or "how did it happen". Diagnostic analytics involves an understanding of the relationship between relatable data sets and identification of specific transactions/ transaction sets along with their behaviour and underlying reasons. Drill down and statistical techniques like correlation assist in this endeavour to understand the causes of various events.

   Example- Low sales in a zone caused by one of the three salesmen.

c. **Predictive analytics**, as the name implies, tries to predict, "What will happen", "when will it happen", "where will it happen", based on past data. Various forecasting and estimation techniques can be used to predict, to a certain extent, the future outcome of an activity. Predictive analysis used many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions about future.

   Example- Sales forecasting, predicting the likelihood of insolvency of a customer/organistion etc.

d. **Prescriptive analytics** takes over from predictive analytics and allows the auditor to 'prescribe' a range of possible actions as inputs such that outputs in future can be altered to the desired solution. Prescriptive analytics can continually and automatically process new data to improve prediction accuracy and provide better decision options. It not only anticipates what will happen and when it will happen, but also why it will happen. In prescriptive analytics, multiple future scenarios can be identified based on different input interventions.

Example- To predict the optimum inventory for particular perishable product.

Data analytic techniques

Data analytic techniques are employed to leverage the above approaches. The analytical techniques that use descriptive and diagnostic approaches would help the auditor to understand the auditable entity and to identify issues therein. A predictive technique like regression will help understand the behaviour of one (or more) variables based on the changes in the other set of variables. These analytic techniques can be broadly classified as Statistical and Visual.

- **Statistical techniques** are the use of statistical measures to derive insights about the dataset.
- **Visualisation techniques** are the use of visuals, graphs and charts to derive an understanding and insight into the dataset.

A combination of various statistical and visual techniques is usually employed for data analytics.

Statistical techniques

Once the data is prepared, as a first step, descriptive statistics of the dataset can be produced to summarize the data in some way with each statistical measure describing the data set. This can be complemented by simple graphical representations such as line graphs, histograms or scatter diagrams. For example, the measures of central tendency describe the expected normal behaviour of the entity and its elements, with respect to a particular parameter or variable. The measures of spread indicates the distribution of the data points. Relationship between two or more variables can be explored or established using techniques of correlation[5] and regression[6]. Identification or segregation of important parameters can be done using regression, component analysis[7] or factor analysis[8]. Clustering[9] and classification[10] can be used to identify group(s) in the data sets based on

[5] Correlation is used to measure the strength of association between two variables and ranges between -1 to +1

[6] Regression analysis gives a numerical explanation of how variables relate, enables prediction of the dependent variable(y) given the independent variable.

[7] Principal Component Analysis aims to reduce the number of inter-correlated variables to a smaller set which explains the overall variability

[8] Factor Analysis aims to group together and summarise variables which are correlated thereby enabling data reduction

[9] . Cluster analysis is a multivariate technique used to group individuals/variables based on common characteristics. (Ref: www.statstutor.ac.uk)

one or more similarity. The results from different statistical tests need to be read together to get a final understanding of the dataset.

Data Visualization

Data Visualization serves the following two distinct purposes:

a. **Exploratory Data Analysis (EDA):** It is an approach to analysing data sets to summarize their main characteristics, often with visual methods. Primarily, EDA is undertaken for seeing what the data can tell us beyond the statistical analysis and modelling.
b. **Communication of findings / reporting**: Insights derived from data can be communicated to the users such as higher management or the readers of audit reports. Data visualisation is a powerful technique to communicate data analytic insights.

It should be noted that a single technique will not give a comprehensive understanding of the data set(s). An auditor should apply a combination of Statistical and Visual techniques to derive insights. The suitability of techniques depends upon the dataset and the purpose of the auditor.

Performing the Analysis is particularly important at the level of the individual data tables making up a data set. When conducting data analysis, it is important to use correct:

– **data terminology**

The first step is by considering how the data should be set up to answer the questions that are being asked

– **data structures**

Then data is requested at the level that analysis is envisioned. Alternatively, data is recognized to the required level

– **units of analysis**

It is important to understand at the outset of a data analysis as to at what unit or level analysis is being done. I it an important issue both analytically and organizationally.

– **types of variables**

Data sets may be reorganized to higher levels of analysis (for instance months to years, counties to states). In general, it is better to request data at the lowest level -So to the

[10] The process of arranging data into homogenous group or classes according to some common characteristics present in the data is called classification. (Ref:http://www.emathzone.com/tutorials/basic-statistics/classification-of-data.html#ixzz4r2Rlugdu)

extent possible request data that are at the lowest level of detail (duration in years versus duration in days that we summarize in years).

Key Teaching Point 7 – Documenting the analysis

Documentation – Why is Documenting the Analysis Important

Documenting audit work is a critical part of **all** audits and the responsibilities of all auditors involved in an audit.

- Documenting work is required by auditing standards
  - Auditors should prepare and maintain audit documentation. Audit documentation related to planning, conducting, and reporting on the audit should contain sufficient information to enable an experienced auditor, who has had no previous connection with the audit, to ascertain from the audit documentation the evidence that supports the auditors' significant judgments and conclusions.
- It is necessary to be able to reproduce the work. The audit work should be re-creatable by the use of tables or various figures in any of the audit products released.
- And be able to make adjustments – sometimes auditors do work that is an extension or update of previous work. It is not possible to recreate the work if there is not an analysis plan, or an adequate audit trail from that previous work.

Computer-Based Audit Documentation

Audit documentation of data analyses should include:

- Data analysis plan

- Documentation of data source to ensure the data are official records or the result of original data collection. An original copy of the data so received should be maintained. Also, for survey data that auditors gather, documentation of the source should be done.

- Data description, including the meaning for fields and evidence that data's integrity has been maintained. This may be part of the analysis plan where recoding is specified, as well as a data dictionary provided by an entity when data are requested from one.
- Description of data reliability checks

- Documentation of all changes, modifications, and deletions. Any changes to the data should be identified and recorded. These changes could be ones described in the analysis plan, with suitable evidence for making the changes.
- Evidence of data manipulation (an audit trail of what was done to the data) A complete audit trail means program logs, history file(s), and cell formulas.
- Evidence of the results of the data manipulation, this means output files, spreadsheets, and summary documents. Perhaps an explanatory flowchart to describe the logic of the analyses.
- Verification of the results of the previous steps by a knowledgeable reviewer or an independent replication of those steps

Audit documentation of data analyses may also include

- Evidence of how the analyses are expected to be used to support work plan (e.g., draft products)
- Prepare all programs, logs, and data for storage
- Embedded explanatory comments by the auditor
- Flowcharts to aid reviewers

Summary

It is as important to think through the questions we want to answer, and the ways we will assess our data, as it is to decide what particular analyses we are going to conduct. We can save ourselves an enormous amount of time by planning carefully. We don't want to waste time on data analysis work that isn't relevant to what we are trying to do on the audit.

Thinking through what each proposed analysis will allow the team to say often eliminates unneeded work. While we don't expect that everything included in the analysis plan will eventually be used in the report, we want to minimize wasted analyses, as well as make sure that the analyses we will depend on will be done correctly and on time.

We need to make sure that the data are set up to allow you to measure what we want to measure. First, consider how the data should be set up to answer the questions that are being asked. Then, request data at the level that is envisioned. Alternatively, we can sometimes reorganize the data to the required level.

Throughout this module, participants will learn how to (1) formulate audit questions, (2) get data from an entity, (3) assess the reliability of data, (4) develop an analysis plan, (5) perform and interpret the analysis, and (6) document the analysis.

Instructor's Notes – Module 2: Principles of Data Analysis

Introduction (10 minutes): Show slides 1 - 2 and introduce the session overview, learning objectives, and basic concepts.

[Notes in power point slides provide instructor's notes]

Brainstorming / Flip charting (20 minutes): Show slide 3 and hold group brainstorming, record responses provided by the class on a flipchart for reference throughout the class.

[Notes in power point slides provide instructor's notes]

Lecture (30 minutes): Show slides 4 - 13

[Notes in power point slides provide instructor's notes]

Individual Exercise and group discussion (10 minutes): Show slides 14-15 and conduct individual exercise with group discussion afterwards.

[Notes in power point slides provide instructor's notes]

Lecture (30 minutes): show slides 16 - 30

[Notes in power point slides provide instructor's notes]

Break (20 minutes)

Exercise and group discussion (30 minutes): Table exercises to identify potential data reliability concerns

[Notes in power point slides provide instructor's notes]

Lecture (50 minutes): Show slides 31 – 45

MODULE 3.1.1: INTRODUCTION TO REMOTE SENSING

Session title: Module 3.1.1: Introduction to Remote sensing	**Session-at-a-glance**

Session Learning Objective:

These sessions would introduce participants to:

- Components of Remote Sensing System
- Electromagnetic Radiation (EMR) and Remote Sensing
- Spectral Signature
- Remote Sensing Platforms
- Sensor Resolutions
- Interpretation of Satellite imageries
- Classification System

Key Points to be covered:

- Possible areas of Remote Sensing Application
- Basics of Remote Sensing
- Learning about Remote Sensing Images, their features and their interpretation.

Teaching Method	Time
Introduction	10 minutes
Brainstorm/Flipcharting	30 minutes
Lecture and slides	50 minutes
Break	30 minutes
Lecture and slides	90 minutes
	Total time: 210 minutes

Link for Training Material of Module 3.1.1 to 3.3.3

https://drive.google.com/open?id=1pLFjaJ3DUmXywhcyOVMIv8olMgeUIfvd .

Participant Notes: **Module 3.1.1: Introduction to Remote sensing**

Overview

Remote sensing can be broadly defined as the process of sensing (i.e. acquiring information) from distance. It is formally defined as the science or technique of obtaining information about objects on surface of the Earth without physically coming into direct contact with them. The process of remote sensing involves making observations and recording radiation coming from Earth features by sensors such as cameras, scanners, radiometers, etc. These sensors are mounted in platforms (such as helicopters, airplane, space shuttle and satellites) located at a particular altitude above the Earth's surface.

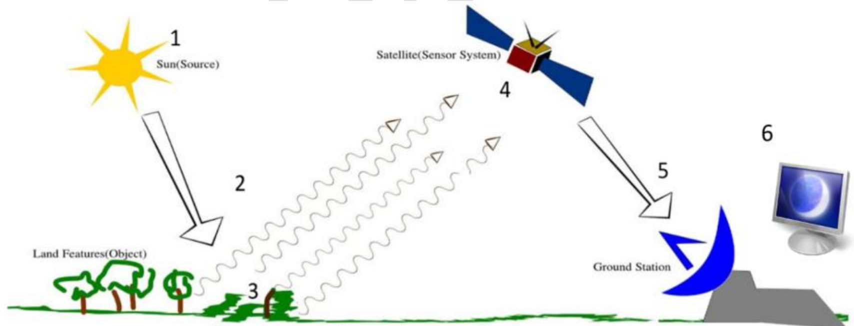
This module is focused on introducing basics of Remote Sensing. Apart from covering the Electromagnetic Radiation properties and their interaction with Earth's features, the module explains the importance of Spectral Signature of various features of the earth. The module introduces participants to various features and qualities of the Remote Sensing images. Finally, it introduces the participants to the elements of the image interpretation and how they are actually applied in digital environment to classify the images.

Basic Concepts

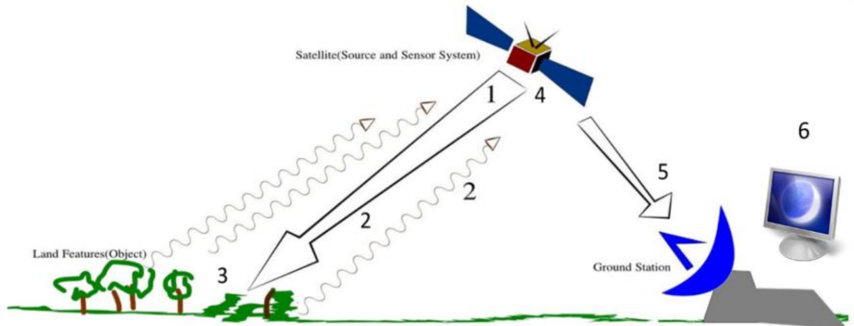
<u>Remote Sensing System</u>

There are several steps in remote sensing system as detailed below:

Source of Energy (1): The first and very important requirement for remote sensing is an energy source which provides electromagnetic energy to the Earth. It may be either from natural (e.g. solar radiation) or artificial (e.g. RADAR) sources. Basically, there are two types of sensor systems, i.e. Passive and Active Sensor System as shown below:



Remote sensing using passive sensor system



Remote sensing using active sensor system

Interaction of energy with the atmosphere (2): When energy travels from its source to the Earth surface, it comes in contact with the Earth's atmosphere where it interacts with atmospheric constituents. The energy reflected from Earth's surface is received by remote sensors. In this process the energy once again interacts with atmosphere.

Interaction with Earth surface features (3): Energy reaching the Earth surface through the atmosphere interacts with the Earth surface features. The interaction and its outcome depend on the characteristics of the features and the energy.

Recording of energy by the sensor (4): After interacting with Earth surface features the reflected and emitted energy travels to the sensor. And, the sensor records the reflected and emitted energy.

Transmission, reception, and processing of the recorded signals (5): The energy recorded by the sensor is transmitted in the form of signals to receiving and processing station on the Earth. The signals are in electronic form and are processed and converted into an image.

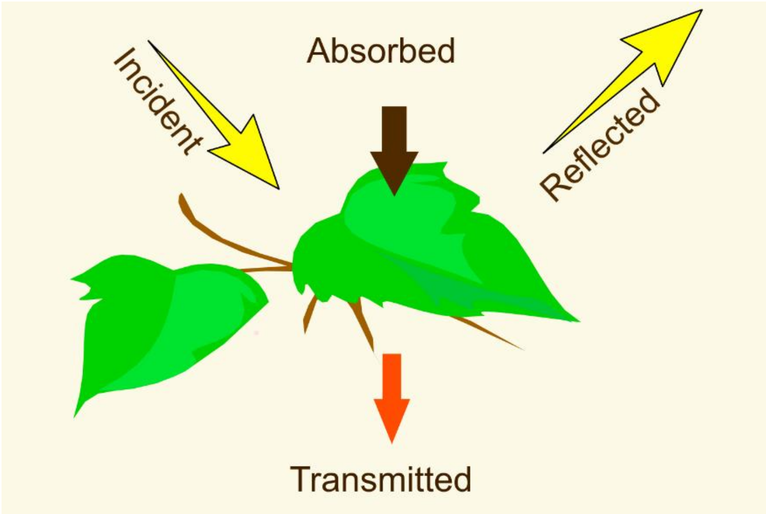
Utilization of the data (6): The processed image is interpreted and analysed to extract information about the object of interest. The above mentioned components comprise the remote sensing system and underline the importance of energy and its interaction with atmosphere and Earth features.

Remote Sensing and Electromagnetic Radiation (EMR)

Electromagnetic Radiation (EMR) is a form of energy exhibiting wave like behavior as it travels through space. EMR ranges from very high energy radiation such as gamma rays and X rays through ultraviolet light, visible light, infrared radiation and microwaves to radio waves. The range of frequencies of EMR is known as electromagnetic spectrum. This division of the electromagnetic spectrum is for practical use. Human eyes use visible light to see objects. We can feel infrared radiation as heat. We employ microwaves in microwave ovens and radio waves are used for communications. All the types of EMR are wave forms which travel at the speed of light. The radiation can be defined in terms of either their wavelength or frequency. Shorter wavelength radiations (infrared or shorter) are generally described in terms of its wavelength, whereas longer wavelength radiations (microwave, etc.) are generally described in terms of its frequency.

Sensors record energy which has interacted with Earth surface features. This energy serves as the main communication link between the sensor and the object. In remote sensing, mostly visible, infrared and microwave bands are used. By the time EMR is recorded by a sensor, it has already passed through the Earth's atmosphere twice (once while travelling from the Sun to the Earth and second time while travelling from the Earth to the sensor).

When light travels through atmosphere, a gradual reduction in its intensity occurs. This attenuation occurs mainly because of the scattering and absorption of light in atmosphere. Absorption is the process by which radiation (radiant energy) is absorbed and converted into other forms of energy such as heat or chemical energy. Absorption is wavelength-dependent.



To understand it better, let us take an example. Grass appears green because it scatters green light more effectively than red and blue light. Apparently, red and blue light incident on the grass is absorbed. The absorbed energy is converted into some other form, and it is no longer present as red or blue light.

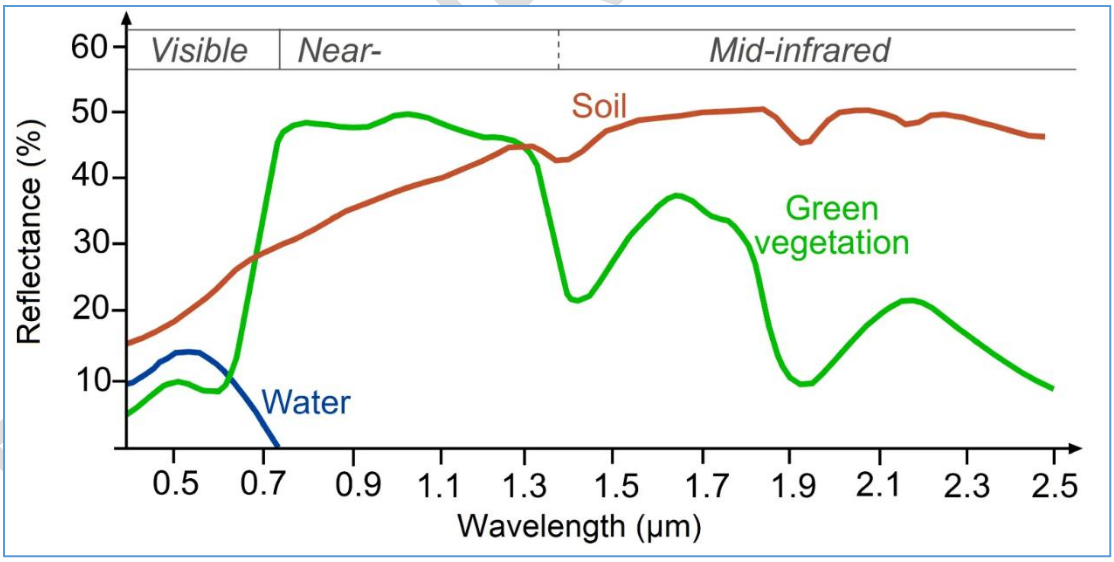
The visible and NIR spectral band from 0.3 **μm** to 3 **μm** is known as the reflective region. In this band, the Sun's radiation sensed by the sensor is reflected by the Earth's surface. The band corresponding to the atmospheric window between 8 **μm** and 14 **μm** is known as the thermal infrared band. The energy available in this band for remote sensing is due to thermal emission from the Earth's surface. Both reflection and self-emission are important in the intermediate band from 3 **μm** to 5.5 **μm**. In the microwave region (1-30 cm) of the spectrum, the sensor is normally a radar, which is an active sensor, as it provides its own source of EMR. The EMR produced by the radar is transmitted to the Earth's surface and the EMR reflected (back-scattered / radar return) from the surface is recorded and analysed. The microwave region can also be monitored with passive sensors, called microwave radiometers, which record the radiation emitted by the Earth's surface and its atmosphere in the microwave region.

SPECTRAL SIGNATURE

Spectral signatures are the combination of reflected, absorbed and transmitted or emitted EMR by objects at varying wavelengths, which can uniquely identify an object. When the amount of EMR (usually intensity of reflected radiation or reflectance in percentage) coming from the material is plotted over a range of wavelengths, the connected points produce a curve which is known as spectral signature of the material or in other words spectral response curve.

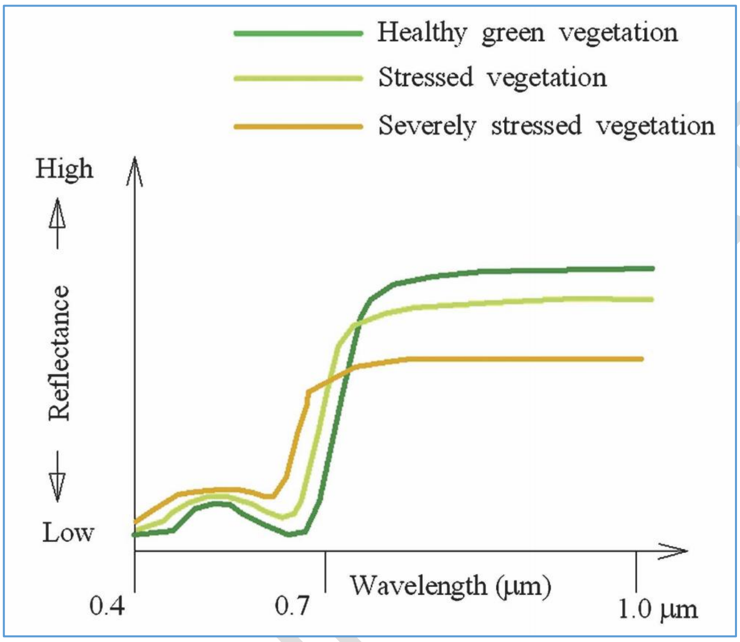
To interpret remote sensing images, it is absolutely important to start with a basic understanding of spectral signature, which means how different terrain features such as water, rock, soils, and vegetation, interact with the different wavelengths (bands) of the EMR.

As you see, vegetation appears green. We know that an object appears green when it reflects green light (light in green region within the visible range of the EM spectrum). In case of vegetation, reflection of green light is due to the presence of the chlorophyll pigment in plant leaves. Presence of the chlorophyll pigment results in unique spectral signature of vegetation that enables us to distinguish it easily from other types of land cover (non-living) features in an optical / near-infrared image. The reflectance of vegetation is low in both the blue and red regions of the EM spectrum, due to absorption of blue and red wavelengths by chlorophyll for photosynthesis. It has a peak reflectance at the green region that gives green colour to vegetation. In the near infrared (NIR) region, the reflectance is much higher than that in the visible band due to the cellular structure in the leaves. Hence, vegetation can be easily identified in the NIR region of spectrum. Typical spectral signature of green vegetation is shown below:



Reflectance of vegetation changes according to the composition, maturity and Spectral Signature health of vegetation. The amount of chlorophyll content determines the health of vegetation. Chlorophyll strongly absorbs radiation in the red and blue wavelengths but reflects green wavelengths. Leaves appear 'greenest' to us when chlorophyll content is at its maximum. In certain seasons, there is less chlorophyll in the leaves; so, there is less absorption and proportionately more reflection of the red wavelengths, making the leaves appear red or yellow (yellow is a combination of red and green wavelengths).

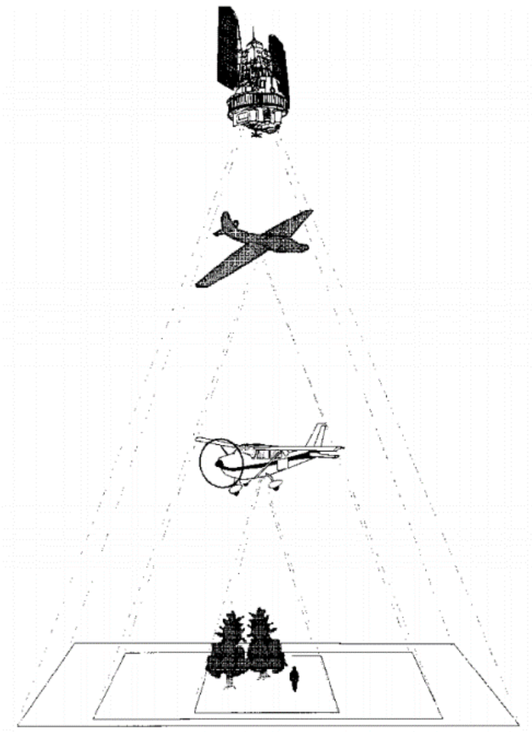
The internal structure of healthy leaves acts as excellent diffuse reflectors of near-infrared wavelengths. If our eyes were sensitive to near infrared, trees would appear extremely bright to us at these wavelengths. In fact, measuring and monitoring the NIR reflectance is one of the ways by which scientists can determine how healthy (or unhealthy) the vegetation is. Spectral signatures of healthy, stressed and severely stressed vegetation are shown below:



The shape of the reflectance spectrum varies with the type of vegetation. For example, the reflectance spectra of deciduous and coniferous tress can be distinguished based on their reflectance spectrum. Deciduous trees and coniferous trees have almost similar reflectance in the visible region, but they vary in the NIR region, where deciduous trees have higher reflectance than coniferous trees.

In case of soil in general, surfaces appear brown to the human eyes. Brown colouring is a product of green and red EMR such that ‘brown’ surfaces absorb more of blue wavelength than either green or red. Furthermore, very little energy is transmitted through soil; the majority of the incident flux is absorbed or reflected. As water is relatively strong absorber of all wavelengths, particularly those longer than the red part of the visible spectrum. Therefore, as the moisture content of the soil increases, the overall reflectance of that soil tends to decrease. Soil rich in iron oxide reflects proportionally more of the red than other visible wavelengths and therefore appears red (rust colour) to the human Spectral Signature eyes. A sandy soil, on the other hand, tends to appear bright white in imagery because visible wavelengths are more or less equally reflected; when slightly less blue wavelengths are reflected, it results in a yellow colour. Unlike soils, rock reflectance is less dependent on water content and completely independent of organic matter content, texture or structure. Rock spectral reflectance primarily depends on their mineral composition.

In case of water, the majority of the radiant flux is not reflected but is either absorbed or transmitted. At visible wavelengths of EMR, very little energy is absorbed, a small amount usually under 5% is reflected and the majority is transmitted. If you happen to be standing in water you can see your foot/toes through the water. Water absorbs strongly at NIR wavelengths, leaving little radiation to be either reflected or transmitted.



Remote Sensing Platforms and Sensor systems

Platforms are commonly called the vehicles or carriers for remote sensing devices. The three most common types of platforms are terrestrial platform, airborne platform, and space borne platform as shown in the figure.

You are familiar with the fact that today man-made or artificial satellites are widely used for a large number of purposes including military and civilian Earth observations, communication, navigation, weather forecasting and research purposes. Hence such satellites are classified into six major types namely, astronomical, communication, weather, earth observation, navigation and reconnaissance satellites based on their uses.

A sensor is a device that gathers electromagnetic radiations, converts it into a signal and presents it in a form suitable for obtaining information about the objects under investigation.

- Sensor systems can be broadly classified as passive or active systems based on the source of EMR. Passive Sensors detect the reflected or emitted EMR from natural sources. The useful wave bands are mostly in the visible and infrared region for passive remote sensing detectors. Active Sensors detect the reflected or emitted radiation from the objects which are irradiated from artificially generated energy sources, such as RADAR and LIDAR. The active sensor detectors are used in the radar and microwave regions.
- Based upon the form of the data output, the sensors are classified into photographic (analogue) and non– photographic (digital) sensors. A photographic sensor (camera) records the images of the objects at an instance of exposure. On the other hand, a non–photographic sensor `obtains the images of the objects in bit-by-bit form. These

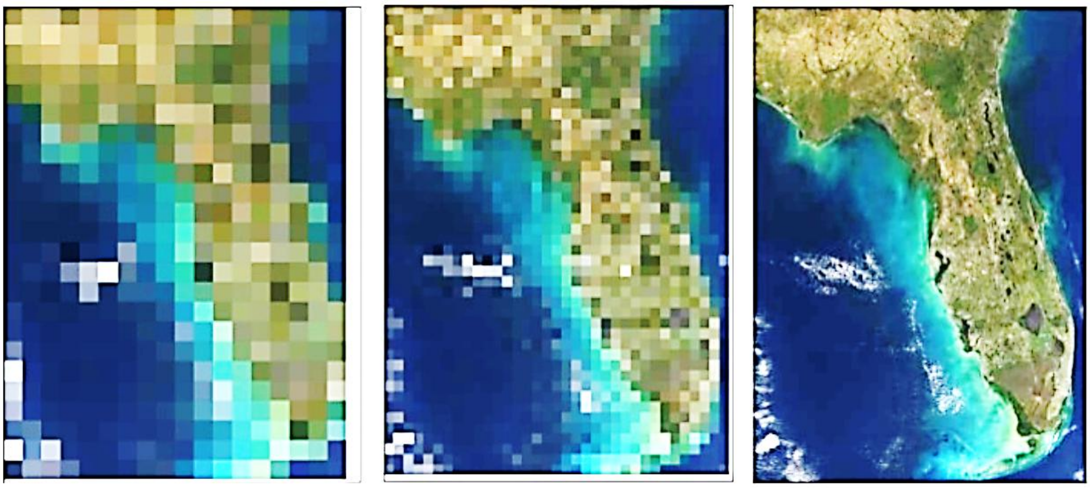
sensors are known as scanners. In satellite remote sensing, the Multi Spectral Scanners (MSS) are used as sensors. These sensors are designed to obtain images of the objects while sweeping across the field of view. A scanner is usually made up of a reception system consisting of a mirror and detectors. A scanning sensor constructs the scene by recording a series of scan lines. While doing so, the motor device oscillates the scanning mirror through the angular field of view of the sensor, which determines the length of scan lines and is called 'swath'. It is because of such reasons that the mode of collection of images by scanners is referred bit–by–bit.

Each scene is composed of cells that determine the spatial resolution of an image. The oscillation of the scanning mirror across the scene directs the received energy to the detectors, where it is converted into electrical signals. These signals are further converted into numerical values called Digital Number (DN Values) for recording on a magnetic tape.

SENSOR RESOLUTIONS

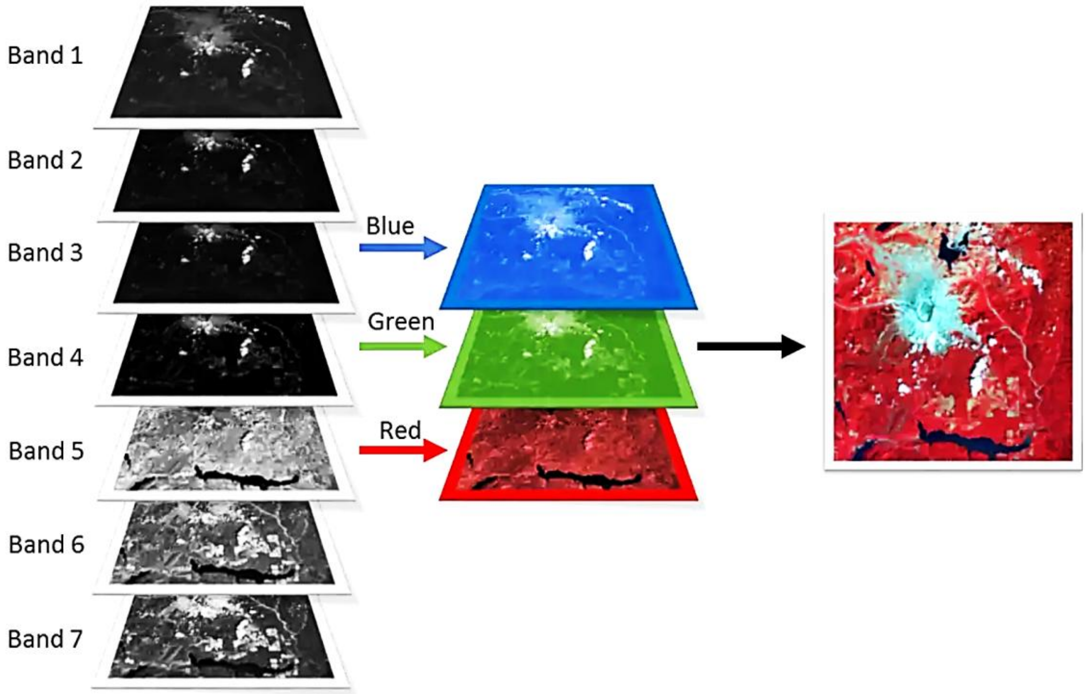
Remote sensors are characterised by spatial, spectral and radiometric resolutions that enable the extraction of useful information pertaining to different terrain conditions as detailed below:

- **Spatial Resolution:** In remote sensing, the spatial resolution of the sensors refers to the capability of the sensor to distinguish two closed spaced object surfaces as two different object surfaces. As a rule, with an increasing resolution the identification of even smaller object surfaces become possible as shown below:

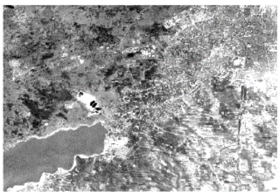


Spatial Resolution of 17 Km, 9 Km and 1Km

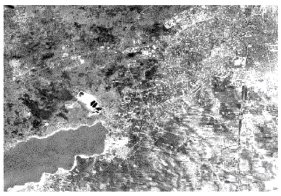
- **Spectral Resolution:** It refers to the sensing and recording power of the sensor in different bands of EMR (Electromagnetic radiation). Multispectral images are acquired by using a device that disperses the radiation received by the sensor and recording it by deploying detectors sensitive to specific spectral ranges. The images obtained in different bands show objects response differently. Figure below illustrates images acquired in different spectral regions:



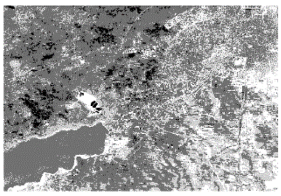
- **Radiometric Resolution:** It is the capability of the sensor to discriminate between two targets. Higher the radiometric resolution, smaller the radiance differences that can be detected between two targets.



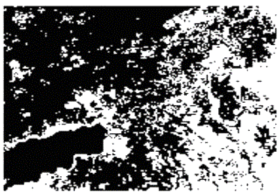
16 Values (4 bit)



8 Values (3 bit)



4 Values (2 bit)



2 Values (1 bit)

The spatial, spectral, and radiometric resolutions of some of the remote sensing satellites of the world are given below:

Satellite / Sensor	Spatial Resolution (in metres)	Number of Bands	Radiometric Range (Number of Grey Level

			Variations)
Landsat MSS (USA)	80.0 x 80.0	4	0 - 64
IRS LISS – I (India)	72.5 x 72.5	4	0 - 127
Landsat TM (USA)	30.00 x 30.00	4	0 - 255
IRS PAN (India)	5.80 x 5.80	1	0 - 127
SPOT HRV – II (France)	10.00 x 10.00	1	0 - 255

INTERPRETATION OF SATELLITE IMAGERIES

Interpretation is the process of extraction of qualitative and quantitative information of objects from satellite images. To derive useful spatial information from images is the task of image interpretation. It includes:

Detection and Identification: recognition of certain target. A simple example is to identify vegetation types, soil types, rock types and water bodies. The higher the spatial / spectral resolution of an image, the more detail we can derive from the image.

Delineation: to outline the recognized target for mapping purposes. Identification and delineation combined together are used to map certain subjects. If the whole image is to be processed by these two procedures, we call it image classification.

Enumeration: to count certain phenomena from the image. This is done based on detection and identification. For example, we can count the number of various residential units.

Mensuration: to measure the area, the volume, the amount, and the length of certain target from an image. This often involves all the procedures mentioned above. Simple examples include measuring the length of a river and the acreage of a specific land-cover class. More complicated examples include an estimation of timber volume, river discharge, crop productivity, river basin radiation and evapotranspiration.

Elements of interpretation

The interpretation of satellite imagery involves the study of various basic characters of an object with reference to spectral bands which is useful in visual analysis. The basic elements are shape, size, pattern, tone, texture, shadows, location, association and resolution.

Shape: The external form, outline or configuration of the object. This includes natural features (e.g. Amazon River) or Man Made feature (e.g. Eiffel Tower).

Size: This property depends on the scale and resolution of the image / photo. Smaller feature will be easily indented in large scale image / photo.

Pattern: Spatial arrangement of an object into distinctive recurring forms: This can be easily explained through the pattern of a road and railway line. Even though both looks linear, major roads associated with steep curves and many intersection with minor road.

Shadow: Indicates the outline of an object and its length which is useful in measuring the height of an object. Taller features cast larger shadows than shorter features.

Tone: Refers to the colour or relative brightness of an object. The tonal variation is due to the reflection, emittance, transmission or absorption character of an objects. This may vary from one object to another and also changes with reference to different bands. In General smooth surface tends to have high reflectance, rougher surface less reflectance.

Infrared imagery: Healthy vegetation reflects Infrared radiation much stronger and appears very bright in the image. A simple example is the appearance of light tone by vegetation species and dark tone by water.

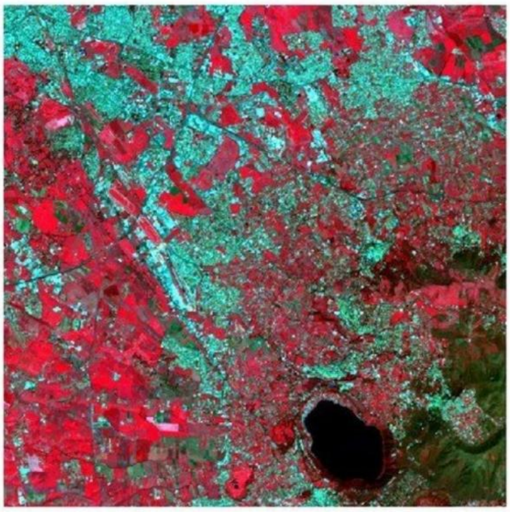
Texture: The frequency of tonal change. It creaks a visual impression of surface roughness or smoothness of objects. This property depends upon the size, shape, pattern and shadow.

Location Site: The relationship of feature to the surrounding features provides clues towards its identity. Example: certain tree species are associated with high altitude areas.

Resolution: It depends upon the photographic / imaging device namely cameras or sensors. This includes of spectral and spatial resolutions. The spectral resolution helps in identifying the feature in specific spectral bands. The high spatial resolutions imagery/photographs is useful in identifying small objects.

Association: Occurrence of features in relation to others.

Hence, careful examination has to be done to identify the features in the imagery combined with field information. The most intuitive way to extract information from satellite images is by visual image interpretation. It is based on our ability to relate patterns and colours in an image to real world features. To visually interpret digital data such as satellite images, individual spectral bands must be displayed simultaneously in the form of a colour composite. When the blue, green and red parts of the electromagnetic spectrum are superimposed, the resulting image strongly resembles what our eyes would see from the sensor's vantage point. Two terms 'true colour' and 'false colour' are used for images. True colour image means that the picture shows objects in the same colours that your eyes would normally see. False colour image means that the colours have been assigned to three different image bands which have been acquired in the wavelengths that your eyes might not normally see.



True Colour and False Colour Images

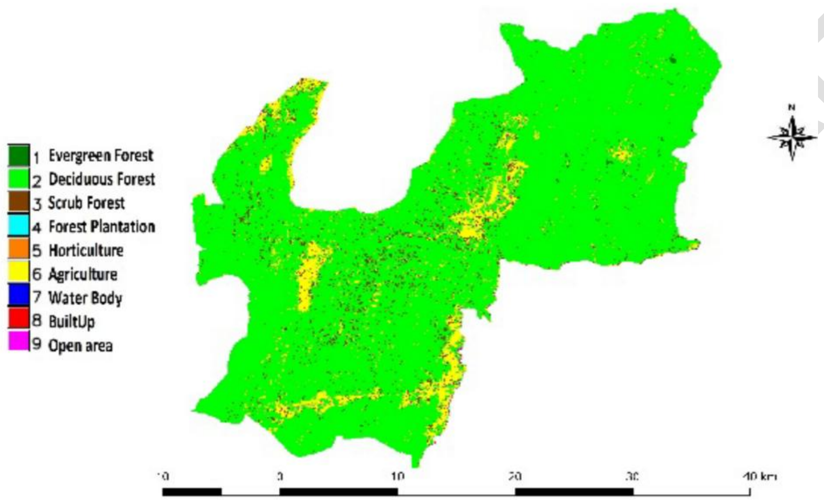
Remote sensing data in digital mode is easy to store and distribute. Computer assisted image interpretation approach mimics the visual image interpretation approach to a certain level. Manual image analysis uses most of the elements of image interpretation such as tone, colour, size, shape, texture, pattern, height, shadow, site and association whereas computer assisted image interpretation involves the use of only a few of the basic elements of image interpretation. In fact, majority of all digital image analysis appears to be dependent primarily on just the tone or colour of image feature.

CLASSIFICATION SYSTEM

Our ultimate aim of interpretation is to produce a land use/land cover map. Land cover refers to the type of features present on the surface of the land. It refers to a physical property or resources e.g., water, snow, grassland, deciduous forest, sand, sugarcane crop, etc. The term Land use relates to the human activity or economic function for a specific area e.g., urban use, industrial use, recreation area or protected area.

It is of prime importance in overcoming problems such as unplanned development, deteriorating environmental quality, loss of agricultural lands, destruction of wetlands and loss of fish and wildlife habitat. Land use data are required in the understanding environmental processes and problems. Land use changes through time can also be interpreted from land cover change maps. These maps are an important tool in the planning process. Such data is increasingly used in tax assessments, natural resource inventories, water-resource inventory, flood control, water-supply planning and waste-water treatment. This data is also required for assessment of environmental impact resulting from development and management of energy and natural resources. This data is also helpful to make national summaries of land use patterns and changes for national policy formulation.

MM Hills Wildlife Sanctuary - 1973



There have been several classification systems developed across the globe. Important one include classification by U.S. Geological Survey, Anderson and others, 1976, Indian (NRSC), etc.

In addition to satellite or aerial imagery, other data can be used to increase the accuracy of the classification. Some possible ancillary data are digital elevation models (DEMs) and their derived datasets (slope and aspect), climate data such as rainfall and temperature, and vector overlays such as roads, rivers and populated places.

All image based mapping processes require field observations. Field observations can be used to gather local knowledge beforehand to guide the interpretation. “Ground Truthing” is the term which is used for data collected ‘on site’. Ground truth data are complementary to the remote sensing data as it helps to link the image data to the ground reality.

When dealing with a new area, some of the features observed on the images will not be understood and field observations will help in interpreting these features. Field observations are also used to validate the interpretation made. Ground truth is a term used in a range of remote sensing techniques. It generally refers to the

The quality of the result of an image interpretation depends on a number of factors such as the interpreter, image data used and the guidelines provided.

PROCESSING AND CLASSIFICATION OF REMOTELY SENSED IMAGES

Classification is the process of assigning spectral classes into information classes. Spectral classes are groups of pixels that are uniform with respect to their brightness values in the different spectral channels of data. Information classes are categories of interest that an analyst attempts to identify in the image on the basis of his knowledge and experience about the area. For example, a remote sensing image contains spectral signatures of several features present on the ground in terms of pixels of different values. An interpreter or analyst identifies homogeneous groups of pixels having similar values and labels the groups as information classes such as water, agriculture, forest, etc. while generating a thematic map. When this thematic information is extracted with the help of software, it is known as digital image classification. It is important to note that there could be many spectral classes within an information class depending upon the nature of features the image represents or the purpose of the classification. In other words, different spectral classes may be grouped under one information class. There are two general approaches to image classification:

- **Supervised Classification**: It is the process of identification of classes within a remote sensing data with inputs from and as directed by the user in the form of training data, and
- **Unsupervised Classification**: It is the process of automatic identification of natural groups or structures within a remote sensing data.

Both these methods can be combined together to come up with a '**hybrid'** approach of image classification.

The image classification process consists of following three stages i.e. **training, signature evaluation and decision making.**

Training is the process of generating spectral signature of each class. For example, a forest class may be defined by minimum and maximum pixel values in different image bands, thus defining a spectral envelope for it. This simple statistical description of the spectral envelope is known as signature. Training can be carried out either by an image analyst with guidance from his experience or knowledge (i.e. supervised training) or by some statistical clustering techniques requiring little input from image analysts (i.e. unsupervised training).

Signature Evaluation is the checking of spectral signatures for their representativeness of the class they attempt to describe and also to ensure a minimum of spectral overlap between signatures of different classes.

Decision Making is the process of assigning all the image pixels into thematic classes using evaluated signatures. It is achieved using algorithms, which are known as decision rules. The decision rules set certain criteria. When signature of a candidate pixel passes the criteria set for a particular class, it is assigned to that class. Pixels failing to satisfy any criteria remain unclassified.

Accuracy assessment is the final step in the analysis of remote sensing data which help us to verify how accurate our results are. It is carried out once the interpretation/classification has been completed. Here, we are interested in assessing accuracy of thematic maps or classified images which is known as thematic or classification accuracy. The accuracy is concerned with the correspondence between class label and 'true' class. A 'true' class is defined as what is observed on the ground during field surveys. For example, a class labeled as water on a classified image/map is actually water on the ground.

All remote sensing images are susceptible to a variety of distortions. These distortions occur due to data recording procedure, shape and rotation of the Earth and environmental conditions prevailing at the time of data acquisition. Therefore, they cannot be used directly and require **image correction**. The primary aim of image correction operations is to correct distorted image data to create a more accurate representation of the original scene. Image corrections, also known as a preprocessing, is a preparatory phase that improves quality of images and serves as a basis for further image analysis.

Limitations of the Remote Sensing:

Though, Remote Sensing is a powerful tool, it has some limitations. Some of these limitations are listed below:

1. It is fairly expensive method of analysis and requires a special training to analyze the images.
2. There is scope for human error in analysis.
3. The instruments used in remote sensing may sometimes be un-calibrated which may lead to un-calibrated remote sensing data.
4. Different features being analyzed may look the same during measurement which may lead to classification error.
5. The image being analyzed may sometimes be interfered by other phenomena that are not being measured and this should also be accounted for during analysis.
6. Remote sensing technology is not a panacea that will provide all the solution and information for conducting physical, biological or scientific study.
7. Sometimes large scale engineering maps cannot be prepared from satellite data.

Summary

The RS images are created by data obtained from various sensors deployed on RS platforms such as Satellite, airplane, etc. It is important to understand the EMR properties and their interaction with Earth's features. The sensors records the EMR and Spectral Signature of various features of the earth basically enables us to identify them in Remote Sensing images. For generating maps with classification, groups of pixels that are uniform with respect to their brightness values in the different spectral channels of data are clubbed into spectral classes. Information classes are assigned to these spectral classes. Such classification is used to prepare Land use/ Land Cover maps which are used to observe changes over a period of time.

Instructor’s Notes – Module 3.1.1: Introduction to remote sensing

[Notes in Power Point slides provide instructor’s notes]

Link to Power Point slides is

https://drive.google.com/open?id=1T2jCF_x7l_e62mBm4ZVVORB_w6JsNUPq .

MODULE 3.1.2: INTRODUCTION TO GIS AND GIS FILE FORMATS

Session title: Module 3.1.2: Introduction to GIS and GIS file formats	**Session-at-a-glance**

Session Learning Objective:

These sessions would introduce participants to:

- Geographic Information Technologies
- GIS Applications
- GIS Data Models
- Data Layers
- Spatial Data formats: raster and vector data
- GIS activities
- GIS Software

Key Points to be covered:

- Possible areas of GIS Application
- Basics of GIS
- Basic working knowledge of QGIS software.

Teaching Method	**Time**
Introduction	10 minutes
Brainstorming / Flip charting	30 minutes
Lecture and slides	60 minutes
Break	20 minutes
Lecture and slides	30 minutes
Hand on training on QGIS software	70 minutes
Break	45 minutes
Hand on training on QGIS software	100 minutes
	Total time: 345 minutes

Link for Training Material for Module 3.1.1 to 3.3.3

https://drive.google.com/open?id=1pLFjaJ3DUmXywhcyOVMIv8olMgeUIfvd .

Participant Notes: Module 3.1.2: Introduction to GIS and GIS file formats

Overview

Geographical Information System (GIS) is a system for capturing, storing, checking, integrating, manipulating, analysing and displaying data, which are spatially referenced to the Earth. This is normally considered to involve a spatially referenced computer database and appropriate applications software. Geographical information is of two types i.e. spatial and non – spatial. The spatial data are characterised by their positional, linear and areal forms of appearances. The most common source of spatial data is topographical or thematic maps in hard copy (paper) or soft copy form (digital). All such maps are characterised by:

- A definite scale which provides relationship between the map and the surface it represents,
- Use of symbols and colours which define attributes of entities mapped, and
- An agreed coordinate system, which defines the location of entities on the Earth's surface.

This module is focused on introducing basics of GIS including the technological backbone i.e. RS, GPS and software Applications. This module introduces GIS Data Models, Data Layers, raster and vector data models and their features, advantages and disadvantages, etc. Finally the module introduces important software available in public domain. Thereafter, hand on training is required for developing working knowledge of QGIS Software.

Basic Concepts

GIS is a computer-based system to aid in the collection, maintenance, storage, analysis, output, and distribution of spatial data and information. Geographic information is the information about places on the earth's surface i.e. "what is where when". Time is very essential. GIS encompasses end-to-end processing of data including capture, storage, retrieval, analysis/modification, display, etc. It uses explicit location on earth's surface to relate data. The aim of GIS is to support decision making in areas which ranges from Disaster Management, land use planning, urbanisation, agriculture, etc.

GIS Technologies

The technological backbone of the GIS includes:

- **Global Positioning Systems (GPS):** a system of earth-orbiting satellites which can provide precise location on the earth's surface.

- **Remote Sensing (RS):** use of satellites or aircraft to capture information about the earth's surface.
- **GIS Software:** with capability for input, storage, manipulation/analysis and output/display of geographic (spatial) information.

GPS and RS are sources of input data. GIS software helps in storing and manipulating GPS and RS data. E.g. Exploratory (Spatial) Data Analysis – EDA/ESDA

GIS Applications

The important analysis which can be done using GIS include:

- Resources inventory (what is available at where?)
- Network Analysis (How to get to a place in the shortest amount of time?)
- Location Analysis (Where is the best place to locate a garbage dump, industry, warehouse, etc.?)
- Terrain Analysis (Which areas are most vulnerable to natural disaster such as flood? Or Where to locate a cyclone shelter?)
- Spatio-Temporal Analysis (Land use: what has changed over the last twenty years in an urban location, near a factory, garbage dump, etc. and why?)
- Calculation of areas, distances, route lengths.

Some of the examples of uses of GIS in Governance are:

1. Highway Department may store information on the state of pavement everywhere on the state highway network, maintain an inventory of all highway signs, analyze data on accidents, look for 'black spots'. Network Analysis can help route planning, vehicle tracking, traffic management.
2. Education Department may analyze whether Government schools cover the largest number of habitations, schools are within a reasonable distance, away from polluting areas, etc.
3. Mining Department may give permits for mining on a particular area which needs to be marked on the map and can be monitored using Satellite imagery.
4. Town and Country Development or Urbanization Department may prepare development plans around the existing cities based on location, terrain and Terrain analysis.
5. Forest Department may do Spatio-Temporal Analysis to find out which areas of the reserved forests are prone to encroachment. Or plan animal corridors to maintain contiguity of the forests.

The same analysis can be done in audit to assess the impact of various interventions made by Government or lack of it.

GIS Data Model: Integration of Spatial and non- spatial data

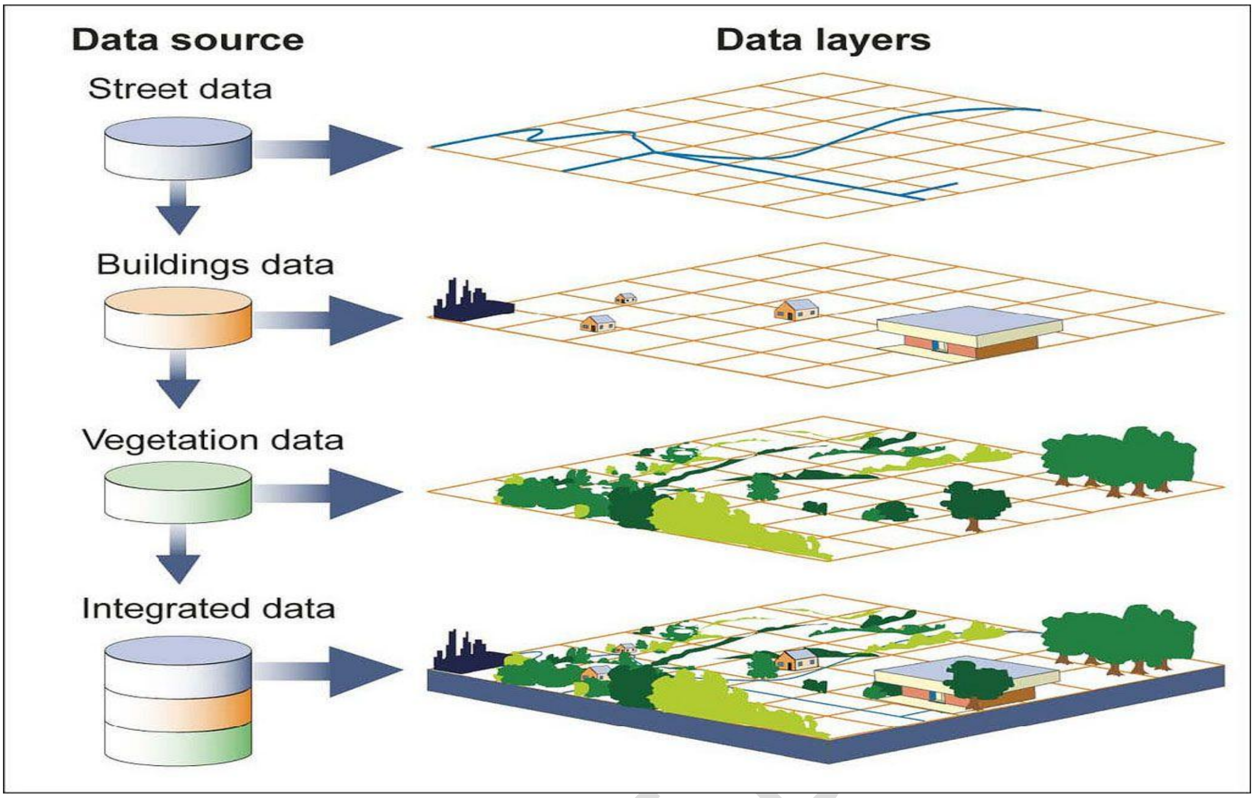
Spatial Data represent a geographical space. They are characterised by the points, lines and polygons. The point data represent positional characteristics of some of the geographical features, such as schools, hospitals, wells, tube-wells, etc. Similarly, lines are used to depict linear features, like roads, railway lines, canals, etc. Polygons are made of a number of inter-connected lines, bounding a certain area, and are used to show area features such as administrative units (countries, districts, states, blocks); land use types (cultivated area, forest lands, degraded/waste lands, pastures, etc.)

Non-spatial Data or attribute data describes the spatial data. For example, if you have a map showing positional location of your school, you can attach the information, such as the name of the school, subject stream it offers, number of students in each class, schedule of admissions, teaching and examinations, available facilities, like library, labs, equipment, etc. In other words, you will be defining the attributes of the spatial data.

For instance, the polygon representing area of a city with proper geographical coordinates is Spatial data, whereas the name, population, per capita income, etc. are non-spatial or attribute data.

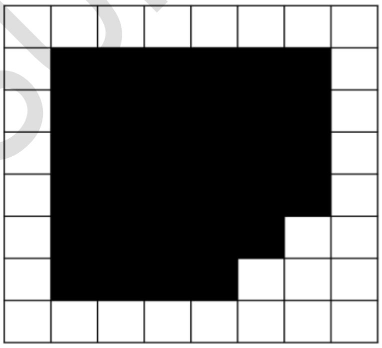
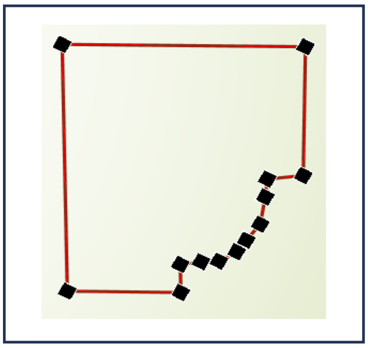
Data Layers in GIS

A layer represents geographic data, such as a particular theme of data. Examples of map layers include streams and lakes, terrain, roads, political boundaries, parcels, building footprints, utility lines, and orthophoto imagery. Each map layer is used to display and work with a specific GIS dataset. Various layers can be superimposed over each other to create various maps and do spatial analysis as described below:



Spatial Data formats: raster and vector data

Raster data represent a graphic feature as a pattern of grids of squares, whereas vector data represent the object as a set of lines drawn between specific points. Consider one of the blocks of iCED, Jaipur. Its raster and vector data formats are as below:

As seen from above, a raster file would represent this image by sub-dividing the area into a matrix of small rectangles, similar to a sheet of graph paper called cells. Each cell is assigned a position in the data file and given a value based on the attribute at that position. Its row and column coordinates may identify any individual pixel. This data representation allows the user to easily reconstruct or visualise the original image.

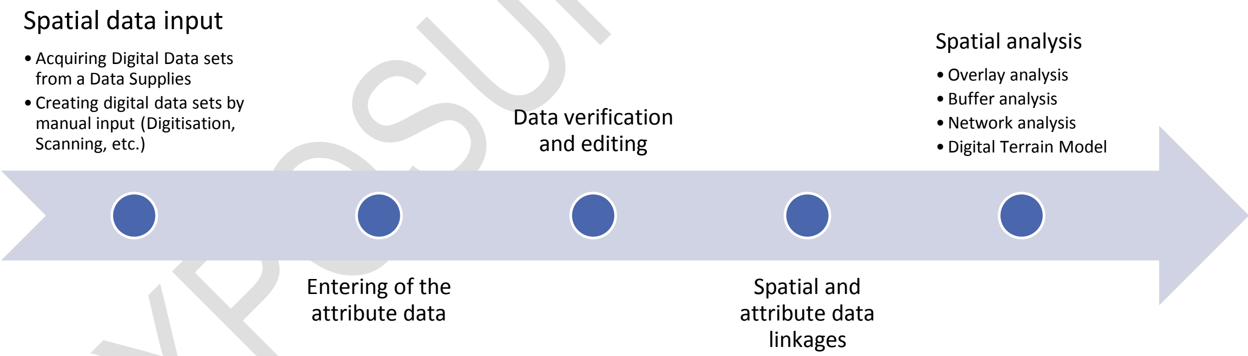
A vector representation of the same image would record the position of the line by simply recording the coordinates of its starting and ending points. Each point would be expressed as two or three numbers (depending on whether the representation was 2D or 3D, often referred to as X,Y or X,Y,Z coordinates) Joining the measured points forms the vector. A vector data model uses points stored by their real (earth) coordinates. Here lines and areas are built from sequences of points in order. Lines have a direction to the ordering of the points. Polygons can be built from points or lines. Vectors can store information about topology. Manual digitising is the best way of vector data input.

The advantages and disadvantages of the two formats is as below:

Raster Model	Vector Model
•Simple data structure	•Compact data structure
•Easy and efficient overlaying	•Efficient for network analysis
•Compatible with satellite imagery	•Efficient projection transformation
•High spatial variability is efficiently represented	•Accurate map output Disadvantages
•Simple for own programming	•Complex data structure
•Same grid cells for several attributes	•Difficult overlay operations
•Inefficient use of computer storage	•High spatial variability is inefficiently represented
•Errors in perimeter and shape	•Not compatible with satellite imagery
•Difficult network analysis	
•Inefficient projection transformations	
•Loss of information when using large cells,	
•Less accurate (although interactive) maps	

GIS activities

The various activities involved in GIS are spatial and non-spatial data inputs, verification of data, spatial and attribute linkages and finally the analysis as described below:



While importing Spatial Data, users must consider the following characteristics of the data to ensure that they are compatible with the application:

- The scale of the data
- The geo-referencing system used
- The data collection techniques and sampling strategy used
- The quality of data collected
- The data classification and interpolation methods used
- The size and shape of the individual mapping units

- The length of the record

Typical problems encountered during Data Verification are incomplete or double data or Spatial data at the wrong scale. While manipulating and analysing data, the same format should be used for all data. When different layers are to be used simultaneously, they should all be in vector or all in raster format. Usually, the conversion is from vector to raster, because the biggest part of the analysis is done in the raster domain. Vector data are transformed to raster data by overlaying a grid with a user-defined cell size

While conducting analysis, in Buffer Operation, buffer of a certain specified distance can be created along any point, line or area feature. It is useful in locating the areas/population benefitted or denied of the facilities and services, such as hospitals, medical stores, post office, asphalt roads, regional parks, etc. Similarly, it can also be used to study the impact of point sources of air, noise or water pollution on human health and the size of the population so affected. This kind of analysis is called proximity analysis. The buffer operation will generate polygon feature types irrespective of geographic features and delineates spatial proximity. For example, numbers of household living within one-kilometre buffer from a chemical industrial unit are affected by industrial waste discharged from the unit.

GIS Software

GIS Software are designed to store, retrieve, manage, display, and analyze all types of geographic and spatial data. GIS software lets you produce maps and other graphic displays of geographic information for analysis and presentation. Important GIS software available are:

Open source Desktop software

- GRASS GIS
- ILWIS
- MapWindow GIS
- **QGIS (previously known as Qgis)**
- uDig

Proprietary GIS software

- Autodesk
- Bentley Systems
- ENVI – Utilized for image analysis, exploitation, and hyperspectral analysis.
- ERDAS IMAGINE
- **Esri – Products include ArcMap, ArcGIS, ArcSDE, ArcIMS, ArcWeb services and ArcGIS Server.**

QGIS among the open source and ArcGIS among the proprietary software are very popular among GIS users. We will learn QGIS through hands on training. QGIS has no initial fee and no recurring fee as it is absolutely free. It is constantly developing as more functionalities are being added very frequently. Extensive help and documentation is available and even tutorials are freely available. It can be installed on MacOS, Windows and Linux. It has large repository of plugins capable of doing large number of tasks. Comparative analysis of a few GIS software is depicted below:

Software (Column Name)		GRASS (Geographic Resources Analysis Support System)	QGIS* (Quantum GIS)	uDig*			MapWindow GIS	ESRI ArcGIS
Functionality (Row Name)				uDig	JGrass	DivaGIS		
FOCUS		Full GUI based GIS Spatial analysis and scientific visualization	Viewing, Editing GRASS-Graphical User Interface	Viewing (OGC standards) Application Framewor	Raster Analysis based on GRASS + 3D visualization	Analyze and document potato genotypes	Providing core GIS and GUI functions, developing Decision Support Systems	Full GUI-based GIS Extensive data creation, editing and analysis
Supported Operating Systems		MacOSX	MacOSX	Windows, Linux, MacOSX	Windows, Linux, MacOSX	Windows only	Windows only	Windows only
License		Free software	Free software	Free software	Free software	Free software	Free software	Proprietary
Data Import/Export	Reading - Raster/Images	All GDAL supported formats, also voxel support	All GDAL supported formats	Yes (1.1.0 is limited to memory)	Yes (through GRASS and Geotools)	Yes (Through Geotools)	All GDAL supported formats + BGD, and others.	ECW, MrSID, JPEG,jpg, jp2, TIF, geoTIFF, PNG, GIF, img, bmp, cib,ers,img,raw,Grid,RST, MrSID,Others
	Reading - Vector	Yes (through OGR)	Yes (through OGR)	Yes	Yes	Yes	At least: SHP, DXF(P), CSV(P)	SHP,DXF,DGNmultiple geodatabases, others
	Reading tabular data	Yes (CSV + many more)	Yes	No			DBF, CSV(P), MDB(P)	
	Reading - Databases	PostGIS, PostgreSQL,	PostGIS	PostGIS, Oracle, DB2,	PostGIS, Oracle, DB2,	PostGIS, Oracle, DB2,	PostGIS(p)	ArcSDE(p),PostGIS,
	Writing - Raster/Images	All GDAL-formats	no	yes	yes	yes	All GDAL supported formats and more	
	Writing – Vector	All OGR-formats	All OGR-formats	Yes (Shape)	Yes (Shape)	Yes (Shape)	At least: SHP	SHP,DXF,DGNmultiple geodatabases, others
	Writing tabular data	CSV	no	no	no	no	copy to Excel table possible	DBF
	Writing - Databases	PostGIS (p, limited)	PostGIS	PostGIS, Oracle, DB2, ArcSDE	PostGIS, Oracle, DB2, ArcSDE	PostGIS, Oracle, DB2, ArcSDE	PostGIS(p), MDB(p)	PostGIS(p), MDB(p)
	Attribute Calculator	yes, via SQL	No	no	no	no	yes	yes
	Q/A: Quality Assurance Tools	yes	No	yes	yes	yes	no	yes
	Join tables	yes	yes (ftools plugin)	no	no	no	no	yes
Printing		yes	yes	yes	yes	yes	Yes	yes
Queries		yes	Yes	limited	yes	no	Attribute queries	Attribute queries, spatial queries
Database Queries (SQL)		yes	yes	no	no	no	no	yes
Styling / Mapping	Point styles	color, size, shape, fill pattern	color, size, shape, fill pattern	Full SLD compliance	Full SLD compliance	Full SLD compliance	Color, size, shape, image	Extensive library plus symbol creation tools
	Line styles	color, width	color, width, pattern	Full SLD compliance	Full SLD compliance	Full SLD compliance	Color, size, line style	Extensive library plus symbol creation tools
	Polygon styles	color, outline, fill	color, outline, fill, pattern	Full SLD compliance	Full SLD compliance	Full SLD compliance	Color, outline, fill patterns.	color, outline, fill patterns, transparency
	Text labeling	Yes (simple labeling)	Yes (simple labeling)	Full SLD compliance	Full SLD compliance	Full SLD compliance	Yes (simple)	
Thematic mapping	Charts	Bar, Pie charts, Graduated Symbol, Ranges	Bar, Pie charts (developed but not included yet)	no	no	no	no	Bar, Pie charts, Graduated Symbol, Ranges
	Classification	equal range, interval, std_deviation, quartiles, custom_breaks	equal range and quantiles	Quantile, equal interval, custom	Quantile, equal interval, custom	no	Equal range, continuous ramps, custom- defined	natural breaks(Jenks) equal range, interval, std_deviation, quantiles, custom_breaks
Vector based Spatial Analysis Tools	Interpolation	IDW, Splines (2D and 3D RST, B-Splines)	GRASS functions via GRASS Plugin	no	GRASS functions	no	Some available through plug-ins	Kriging and co-kriging via add-on
	Spatial Statistics	Link to R-stats	yes (R plugin)	no	GRASS functions	no	Basic raster statistics(p)	Geostatistical analyses
	Buffer	yes	GRASS functions via GRASS Plugin	Yes with Axios plugin	GRASS functions and Axios plugin	no	yes (p)	yes, single and multi-ring, merging
	Join geometries by attributes	yes	no	no	no	no	yes (p)	yes
	aggregate /join attributes spatially	yes	GRASS functions via GRASS Plugin	no	GRASS functions	no	yes (p)	yes
	Spatial Correlation Analysis	yes (via R interface)	no	no	no	no	no	yes, multiple methods
	Other	union (= dissolve), centroid calculation		Union, split, intersect, reproject	Union, split, intersect, reproject	no	Union, merge, centroid, intersection, differences (p)	Merge, dissolve, intersection, difference, convex hull, spatial join
Multi-language support	Yes/no	yes	yes	yes	yes	yes	yes	yes
Support of different Projections/ Coord. Reference Systems		yes	yes	yes	yes	yes	yes	yes
Data Exploration Tools	Histograms	yes	yes	no	no	yes	No	yes
	classification	yes	GRASS functions via GRASS Plugin	no	GRASS functions	yes	no	yes
Topology creation		yes	GRASS functions via GRASS Plugin	no	no	no	TIN creation and viewing ability	yes
	On the fly topology editing	yes	under development	no	no	no	no	yes
	Triangulation (planar)	yes	no	no	no	no	no	yes
3D-Views		yes	GRASS functions via GRASS Plugin	no	Yes, In development	no	yes	yes
Histogramming	yes	GRASS functions via GRASS Plugin	no	GRASS functions		yes	yes	
Image Analysis Tools	Geo-Referencing	yes	Geocoding of scanned maps	no	GRASS functions		yes	yes
	Ortho-rectification	yes	no	no	GRASS functions		Under development	no
	Filter operations	yes	GRASS functions via GRASS Plugin	no	GRASS functions		Under development	yes
Version		GRASS 6.4	QGIS 1.3	Version 1.1-RC12			Version 4.5 RC2 (April 2008)	Release 9.3.1

The training manual is available at

https://docs.qgis.org/2.18/en/docs/training_manual/index.html.

Apart from this, a video prepared to demonstrate simple tasks in QGIS is available at https://drive.google.com/file/d/1zj_v9pETEEq-x6FgdDvPeG0z1bUBK5GO/view?usp=sharing

ISSAI 5540 introduces Geographical Information Systems (GIS) as an audit tool and provides practical guidance on GIS. A guidance on GIS is available at

http://www.issai.org/data/files/C6/53/40/D9/D929851068C19585BA5818A8/issai-5540-e-appendix.pdf

http://www.issai.org/en_us/site-issai/issai-framework/4-auditing-guidelines.htm

Instructor's Notes – Module 3.1.2: Introduction to GIS and GIS file formats

[Notes in Power Point slides provide instructor's notes]

The Power Point Slides are available at https://drive.google.com/open?id=17i2am7KlYMWBZv8st7qPOHUYqFh2-hL3 .

For hands on training, the reference material to be used is training manual available at https://docs.qgis.org/2.18/en/docs/training_manual/index.html.

Apart from this, a video prepared to demonstrate simple tasks in QGIS is available at https://drive.google.com/file/d/1zj_v9pETEEq-x6FgdDvPeG0z1bUBK5GO/view?usp=sharing .

MODULE 3.2.0 INTRODUCTION TO OPEN SOURCE REMOTE SENSING DATA

Session title: Module 3.2.0 Introduction to Open Source remote sensing data	**Session-at-a-glance**

Session Learning Objective:

These sessions would introduce participants to various open source Remote Sensing data. The session will also provide the participants hand on training on Google earth and would also introduce them to the GLOVIS data.

Key Points to be covered:

- Various open source RS data
- Basic knowledge of GLOVIS data
- Working knowledge of Google earth

Teaching Method	Time
Introduction	10 minutes
Lecture and slides	10 minutes
Hand on training on downloading Glovis data	80 minutes
Break	20 minutes
Lecture and Slides	20 minutes
Hand on training on Google Earth	80 minutes
Break	40 minutes
Recap of hand on training	30 minutes
	Total time: 290 minutes

Link for Training Material of Module 3.1.1 to 3.3.3

https://drive.google.com/open?id=1pLFjaJ3DUmXywhcyOVMIv8olMgeUIfvd .

Participant Notes: Module 3.2.0 Introduction to Open Source remote sensing data

Overview

Recent advancements in spatial technologies have proved instrumental in management of Earth resources. The main reason for using image processing techniques and GIS is that images and maps have always been easier to understand. It brings us closer to visualising complex interactions and patterns in real-world planning.

In previous module, we studied the theoretical concepts of remote sensing, image interpretation, classification, GIS, etc. This module will deal with various tools and techniques for creating, processing and analysing geospatial data for various applications and develop basic practical skills. We will first learn use of open source geoinformatics software QGIS (Qgis) to download remote sensing and other data from internet. This would be followed by introduction Google Earth software.

Basic Concepts

Geospatial data are available from a large number of sources in a large number of formats. The data sources range from commercial suppliers to government suppliers to free websites. A good source for obtaining remote sensing data is the GLOVIS website. The website provides satellite data of different dates for almost all parts of the globe. Some of the data are free to download while the others can be downloaded on payment.

USGS Earth Explore or GLOVIS Data

Geospatial data are available from a large number of sources in a large number of formats. The data sources range from commercial suppliers to government suppliers to free websites. A good source for obtaining remote sensing data is the GLOVIS website. The website provides satellite data of different dates for almost all parts of the globe. Some of the data are free to download while the others can be downloaded on payment. To download data you will be directed to the Earth Explorer website. You have to register at this site once and then you can download data. You can even check the data available along with metadata and browse them as well before you download. To download the data you need to go to the http://glovis.usgs.gov website and search for type and date of data you require. The data is downloaded as separate zipped folder which has to be unzipped for further use.

USGS is without any doubt the best free satellite data providers. It provides access to Landsat satellite data, which is the only satellite program with 40-years of history of our Earth with consistent spectral bands. It also provides NASA's ASTER and Shuttle Radar Topography Missions global Digital Elevation Models. It provides full access to NASA's Land

Data Products and Services including Hyperion's hyperspectral data, MODIS & AVHRR land surface reflectance and disperse Radar data.

In the hands on training to be given under this module, you will know the steps to be taken to download the USGS data. Basic requirements are that participants should have a Desktop or laptop during the session with internet facility. The module has a video explaining how to download USGS data using SCP plugin in QGIS.

Google earth

Google Earth(GE) is a computer program that renders a 3D representation of Earth based on satellite imagery. The program maps the Earth by superimposing satellite images, aerial photography, and GIS data onto a 3D globe, allowing users to see cities and landscapes from various angles. Users may use the program to add their own data using Keyhole Markup Language (KML) and upload them through various sources, such as forums or blogs. Google Earth is able to show various kinds of images overlaid on the surface of the earth and is also a Web Map Service client. In version 5.0, Google introduced Historical Imagery, allowing users to view earlier imagery. Clicking the clock icon in the toolbar opens a time slider, which marks the time of available imagery from the past. This feature allows for observation of an area's changes over time.

The google earth can be accessed on the web at https://earth.google.com/web/. However, the web version does not have all the features. Google Earth Pro, the desktop software, was originally the business-oriented upgrade to Google Earth. Since January 2015, Google decided to make it free to the public. Google Earth Pro is currently the standard version of the Google Earth desktop application. The Pro version includes add-on software for movie making, advanced printing, and precise measurements, and is currently available for Windows, Mac OS X 10.8, and Linux. The application can be downloaded from https://www.google.com/earth/download/gep/agree.html and installed freely. Google Earth Outreach offers online training on using Google Earth and Google Maps for public education on issues affecting local regions or the entire globe. The tutorials can be accessed at https://www.google.com/earth/outreach/learn/. The Google earth pro is capable of:

- Adding placemarks, paths, polygons, and photos. Google Earth allows you to save your project to your computer in the .kmz file format. The tutorial for annotating Google earth is given at https://www.google.com/earth/outreach/learn/annotating-google-earth/.
- Overlaying photographs and scanned maps.
- Measure earth features such as length of road, area of a polygon, etc.
- Adding legends, logos and banners to Google Earth with Screen Overlay.
- Importing GPS and GIS data and present it on google earth.

- Viewing historical imagery of the place. There are multiple images available for all places and we can see how places have changed over time.
- Offers huge number of layers including place names, roads, types of buildings, water body outline, parks, golf courses, Panoramas, Wikipedia integration, etc.
- We can also see 3D elevation of terrain.

Google Earth with its high resolution images is very helpful as it can be used in projects like cadastral/parcel mapping, pipeline/electrical layout planning, city/town management, etc. The terrain information in GE can be used for land use studies. Historical images of GE help in change (natural and human induced) detection studies. Geoinformatics tools like GDAL tools, ESRI ArcGIS, QuantumGIS, etc. are giving functionalities to export your project data into KML. With this feature you can validate your GIS data by embedding them on GE. The Google layers plugin of Qgis allows you to download the Google images into your local computer. With this feature you can have high resolution images for executing your geoinformatics project.

In the hands on training in this module, you will learn how to use Google earth for simple tasks. Basic requirements are that participants should have a Desktop or laptop during the session with internet facility. The module has a video explaining how to accomplish simple tasks using Google Earth. The link to video is at https://drive.google.com/open?id=1kDmRlxyDfnorYSW0dLBFybZsMxE9r0I4 .

You will also learn about ESA's Sentinel Mission and Bhuvan Indian Geo-Platform.

Summary

As explained above, Geospatial data are available from a large number of sources in a large number of formats. The data sources range from commercial suppliers to government suppliers to free websites. One can choose the best source as per their requirement.

Instructor's Notes – Module 3.2.0 Introduction to Open Source remote sensing data

[Notes in Power Point slides provide instructor's notes]

The Power Point Slides are available at https://drive.google.com/open?id=174ZVFo_U0iuc3h1z6ya2-oDWy42pa6dY .

For hands on training on downloading GLOVIS/ USGS Earth Explorer data, the video at https://drive.google.com/open?id=1zj_v9pETEEq-x6FgdDvPeG0z1bUBK5GO contains a section on downloading Glovis data.

For hands on training on Google earth, the reference material from Google Earth's outreach program can be used, which is available at https://www.google.com/earth/outreach/learn/. Apart from this, a video prepared to demonstrate the usage of Google earth is available at https://drive.google.com/open?id=1kDmRlxyDfnorYSW0dLBFybZsMxE9r0I4 .

The 2013 WGEA guidance paper entitled Land Use and Land Management Practices in Environmental Perspective: is available at https://www.environmental-auditing.org/media/2937/2013_wgea_land-use_view.pdf

The 2004 WGEA guidance paper entitled Auditing Water Issues: may be referred for further study on the use of GIS. It is available at https://www.environmental-auditing.org/media/2943/2013_wgea_water_view.pdf

MODULE 3.3.1 APPLICATION OF GIS FOR ENVIRONMENTAL AUDITS CASE STUDY 1: PERFORMANCE AUDIT ON ADMINISTRATION OF NATIONAL PARKS AND WILDLIFE SANCTUARIES IN KARNATAKA, INDIA

Session title: Module 3.3.1 Application of GIS for Environmental Audits Case Study 1: Performance Audit on Administration of National Parks and Wildlife Sanctuaries in Karnataka	**Session-at-a-glance**

Session Learning Objective:

These sessions, through case study, would demonstrate to the participants on how study of "Land Use Land Cover" (LULC) change can be carried out mostly using open source data. The land management policies, population, agricultural production and urban expansion are considered as main drivers. In this regard, LULC information of any region serves as a basis for understanding bio-geophysical processes and anthropogenic pressures. LULC is an important indicator of changes happening in and around the Protected Areas which have a bearing on the conservation and protection of wildlife and their habitat.

Key Points to be covered:

- Demonstrate use of RS when other data is not reliable or available.
- Steps performed for conducting study of LULC changes.
- Deriving audit conclusions from study of LULC changes.

Teaching Method	Time
Introduction	10 minutes
Lecture and slides	10 minutes
Hand on training on downloading Glovis data	80 minutes
Break	20 minutes
Lecture and Slides	20 minutes
Hands-on training on Google Earth	80 minutes
Break	40 minutes
Recap of hands-n training	30 minutes
	Total time: 290 minutes

Link for Training Material of Module 3.1.1 to 3.3.3

https://drive.google.com/open?id=1pLFjaJ3DUmXywhcyOVMIv8olMgeUIfvd .

Participant Notes: Module 3.3.1 Application of GIS for Environmental Audits Case Study 1: Performance Audit on Administration of National Parks and Wildlife Sanctuaries in Karnataka

Overview

This case study would demonstrate how study of "Land Use Land Cover" (LULC) change can be carried out mostly using open source data. The land management policies, population, agricultural production and urban expansion are considered as main drivers of LULC changes. In this regard, LULC information of any region serves as a basis for understanding bio-geophysical processes and anthropogenic pressures. LULC is an important indicator of changes happening in and around the Protected Areas which have a bearing on the conservation and protection of wildlife and their habitat.

Background Material

The audit report is kept in this folder by the name "PA on NPWLS English.pdf". You may refer the following for the case study:

Chapter 1: Introduction

Chapter 2: Audit Approach

Chapter 3: Forest cover dynamics

Chapter 6: Encroachment / occupation of forest land and rehabilitation of villagers

GIS/ Satellite imagery offers strong collaborative evidence for assessing environmental issues such as land use changes, encroachments and damages to forest cover. GIS results are easy to understand and hard to deny.

Methodology used for studying LULC changes

One of the audit objectives of the Performance Audit was to assess whether Protection and Conservation of Wildlife, including their habitats, was adequately planned for and implemented in the administration of the Protected Areas, by examination of consolidation of boundaries, and status of encroachments and rehabilitation of persons living inside the Protected Areas.

LULC changes in selected protected areas of Karnataka with their buffer region (10 km) has been analyzed with the help of temporal remote sensing (RS)data, ancillary data (collateral data compiled from government agencies) and field investigations.

Multi resolution RS data acquired through the sensors of U.S. Geological Survey Earth Observation Satellites (EOS), Indian Remote Sensing (IRS) system at temporal scale have been used.

The RS data used in the study are Landsat MSS (1973), Landsat-5 i.e. TM (1992), Landsat-8, IRS p6L4X (2016) and online Google Earth data (http://earth.google.com) as below:

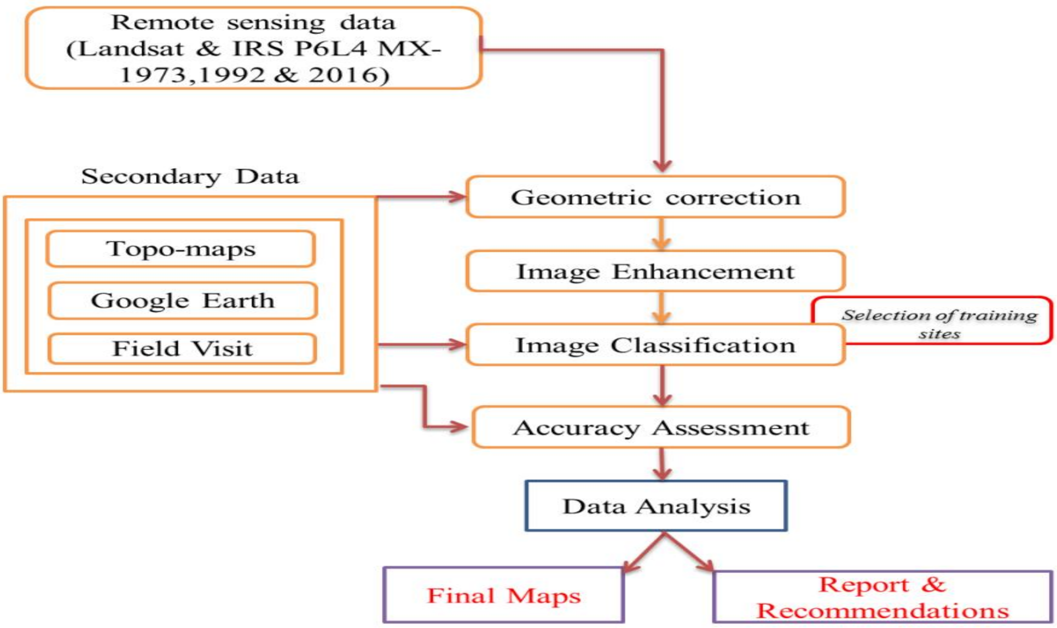
Satellite	Year	Resolution			
		Spatial	**Spectral**	**Radiometric**	**Temporal**
Landsat-1	1973	60 m	RBV(3), MSS (4)	18 days	6-bits
Landsat-5	1992	30 m	MSS (4), TM (7)	16 days	8-bits
Landsat-8	2016	30 m	OLI (11)	16 days	10-bits
IRS	2016	5 m	LISS IV (3)	5 days	10-bits

The ancillary data is used to assist the interpretation of different land use types from remote sensing data. Topographic maps provided ground control points to rectify remotely sensed data and scanned paper maps (topographic maps). Survey of India (SOI) topo sheets (1:50000 and 1:250000 scales) and vegetation map of South India developed by French Institute (1986) of scale I :250000 was digitized to identify various forest cover types and temporal analyses to find out the changes in vegetation.

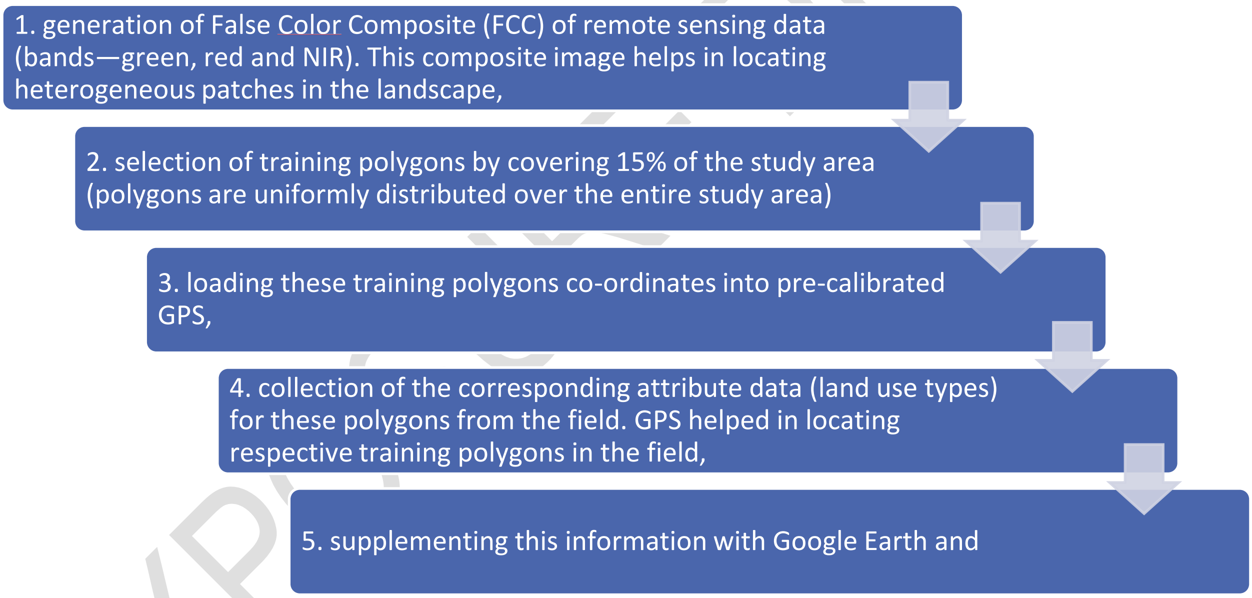
Pre-calibrated GPS (Global Positioning System - Garmin GPS unit) was used for field measurements. Ground control points were used to geometrically correct remote sensing data and verify the classified land use information.

Land cover analysis has been carried out using Normalized Difference Vegetation Index (NDVI). NDVI is the most commonly used vegetation index to distinguish healthy vegetation from others or from non- vegetated areas using red and near infrared reflectance values of RS data. NDVI provides information about the spatial and temporal distribution of vegetation communities, quality, biomass, and the extent of land degradation in various ecosystems. NDVI also known as a greenness index, value ranges between -l to +1. NDVI is sensitive to the presence, density and condition of vegetation and is correlated with absorbed photo synthetically active radiation (PAR) and vegetation primary production. Based on grey scale corresponding to a pixel, digital number dense green vegetation and non-vegetation features were separated.

The methodology adopted for the LULC changes is as below:



Steps performed were as below:



GRASS GIS (Geographical Resources Analysis Support System, http://ces.iisc.ernet.in/grass), a free and open source software with the robust support for processing both vector and raster data has been used for analyzing RS data by using available multi-temporal "ground truth" information. The same analysis can be done using SCP plugin of the QGIS.

Participants would be asked to volunteer potential audit conclusions. Their responses would be evaluated and finally they would be told the conclusions drawn in the Audit Report.

nstructor’s Notes – 3.3.1 Application of GIS for Environmental Audits Case Study 1: Performance Audit on Administration of National Parks and Wildlife Sanctuaries in Karnataka

[Notes in Power Point slides provide instructor’s notes]

The Power Point Slides are available at

https://drive.google.com/open?id=13Po5KbyTmMIC_30YxpamWbgLUj-mC7VeThe audit report referred above is available at

https://drive.google.com/open?id=10UnjsPK1bwwik8rhvVHgK2DqHSPQY9HZ .

The following chapters may be referred for the case study:

Chapter 1: Introduction

Chapter 2: Audit Approach

Chapter 3: Forest cover dynamics

Chapter 6: Encroachment / occupation of forest land and rehabilitation of villagers

At the end of discussion, participants could be told the following.

The results of the study were used to derive following audit conclusions:

- The evergreen to semi-evergreen forest area decreased in 12 of the 13 selected Protected Areas (PAs).
- The area under deciduous forests has increased in six PAs and decreased in seven PAs,
- The total area under cultivation i.e., areas under agriculture and horticulture have increased in all the PAs,
- Built up area has increased in 11 of the 13 PAs, while the open areas have increased in nine Pas
- Further comparison of encroachments as assessed through satellite based analysis and those recorded by the Divisions indicated that the encroachments recorded by Department were less.

The details are available in the report referred to initially.

Challenges encountered and lessons learnt during the PA on Management of National Parks and Wildlife Sanctuaries in Karnataka

1. **Obtaining Cadastral maps from the Department**: For obtaining the cadastral maps, two to three Departments/Organisations had to be approached. However, it was available only with Revenue Department who shared these the cadastral maps only on an assurance that these would not be shared with anybody but used for audit purpose only.

2. **Presence of different data at different levels/sources**: The Departmental statistics relating to loss due to forest fires, poaching cases, road kills, and encroachments were f under stated which had to be verified fromother reliable sources like Ministry of Road

Transport and Highways (MoRTH) and other reputed organisations. Some such instances are:

a. Divisional data of forest fires for February-March 2012 at two Tiger Reserves differed from the one observed by Principal Chief Conservator of Forests (PCCF). Moreover, satellite based assessment by Indian Space and Research Organization (ISRO) indicated that the forest area affected by fires in one tiger reserve was significantly greater than reported by the division or the PCCF Nagarahole NR.

b. The cases of poaching of tigers as maintained by the Divisions and as furnished by the PCCF (Wildlife) differed from each other. Moreover , number of tiger poaching cases reported under official database of National Tiger Conservation Authority was higher than the numbers reported by the Divisions/Department.

c.In respect of monitoring of road kills, Audit used reports from the press which had recorded details of these in some PAs. During the Exit Conference, the Department accepted the observation and agreed to sensitise the staff and monitor all road-kills.

d. In case of a Hydroelectric power project, satellite images indicated that the project was already commissioned although the department stated that approval for the project was awaited.

e. The details of encroachments maintained by the Division were also very conservative as compared to satellite based assessment data.

3. **External Expertise**: Satellite based assessment is a highly technical/professional work., Training in-house professionals for using their expertise in works like superimposing different layers. The work of quantification of changes could be done through external sources as in-house analysis could be challenged by the auditee organisations. However,one or two officers can be involved in the process of analysis of satellite imageries done by the external agencies to develop in-house expertise.

Links to sources of reading material for further suggested study:

1. NPTEL: Introduction to Remote Sensing available at https://nptel.ac.in/courses/121107009/
2. Udemy courses available at https://www.udemy.com/land-use-land-cover-classification-gis-erdas-arcgis-envi/ , https://www.udemy.com/topic/qgis/ , and https://www.udemy.com/topic/gis/,
3. COURSERA: Geographic Information Systems (GIS) Specialization Map Your World With GIS. Explore the tools, concepts, and terminology of spatial analysis and modeling Courses available at https://www.coursera.org/specializations/gis#courses
4. MountHolyoke https://www.mtholyoke.edu/courses/tmillett/course/geog205/files/remote_sensing.pdf

MODULE 3.3.2 APPLICATION OF GIS FOR ENVIRONMENTAL AUDITS CASE STUDY 2: AUDIT ON “ACTIVITY OF THE STATE AND LOCAL GOVERNMENTS IN THE COLLECTION AND RECOVERY OF MUNICIPAL WASTE” AND “ACTIVITIES OF THE STATE IN ORGANIZING WASTE TREATMENT IN OIL SHALE MINING AND PROCESSING”: NAO ESTONIA

Session title: Module 3.3.2 Application of GIS for Environmental Audits Case Study 2: Audit on “Activity of the state and local governments in the collection and recovery of municipal waste” and “Activities of the state in organizing waste treatment in oil shale mining and processing” : NAO Estonia	**Session-at-a-glance**

Session Learning Objective:

These sessions have two case studies; Part 1 deals with audit of “Activity of the state and local governments in the collection and recovery of municipal waste” and Part 2 deals with “Activities of the state in organizing waste treatment in oil shale mining and processing”. Both audits were conducted by NAO Estonia.

Part 1 deals with municipal waste. Governments generally manage to collect municipal waste. However, the collection of waste alone is not enough to reduce the burden on the environment. The last decade has brought the realisation that instead of dumping waste in landfills, it has to be recycled. The NAO conducted buffer analysis using QGIS to find out the population inside and outside of service areas of the packaging containers.

Part 2 deals with the closing of the oil shale waste landfill. The NAO used Lidar and GIS technology to arrive at the conclusion that Oil shale landfill was used to deposit waste in order to “recycle it later” which was not correct. It also concluded that most probably the closing works did not actually take place and audit informed the Environmental Inspectorate about that.

Key Points to be covered:

- Demonstrate use of RS when other data is not reliable or available.
- Steps performed for conducting buffer analysis.
- Use of Lidar technology is assessing landfill works.

Teaching Method	Time
Part 1	
Introduction	10 minutes
Brainstorming	15 minutes
Lecture and slides	30 minutes
Break	20 minutes
Part 2	
Introduction	
Brainstorming	10 minutes
Lecture and slides	15 minutes
	30 minutes
	Total time: 130 minutes

Participant Notes: Module 3.3.2 Application of GIS for Environmental Audits Case Study 2: Audit on “Activity of the state and local governments in the collection and recovery of municipal waste” and “Activities of the state in organizing waste treatment in oil shale mining and processing”: NAO Estonia

Part 1: “Activity of the state and local governments in the collection and recovery of municipal waste”

Overview

This case study would demonstrate the use of buffer analysis using QGIS to find out the population inside and outside of service areas of the packaging containers.

Audit: “Activity of the state and local governments in the collection and recovery of municipal waste” (2016)

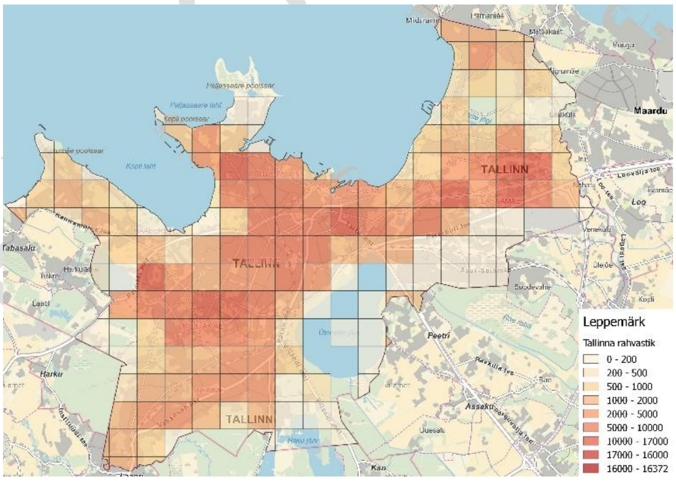
Task/question: Is the location of packaging containers convenient to use for citizens (how far are containers located from households)? (in capital region of Tallinn)

Method used: QGIS

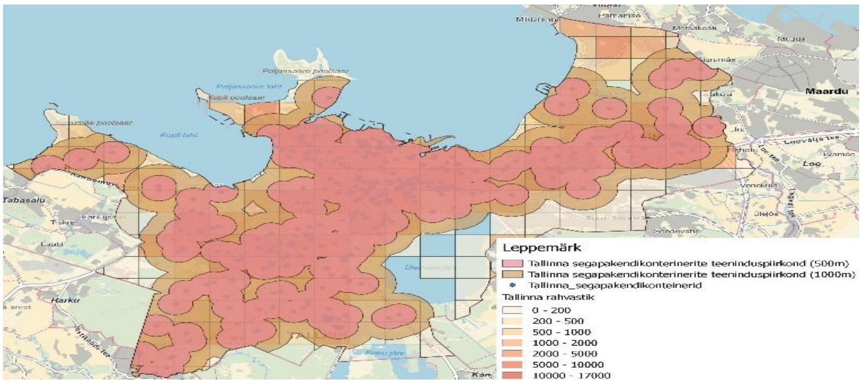
Source of data: Estonian Land Board (base map), Statistics Estonia (data on population and its location), www.kuhuviia.ee (data on location of packaging containers)

Steps taken to analyse the data:

1. importing the data on population to QGIS



2. importing the data on packaging containers to QGIS
3. creating buffers (service areas) around packaging containers in QGIS (500 m and 1000 m)



4. determining areas inside and outside of service areas the packaging containers in QGIS;
5. Calculating the population and outside of service areas the packaging containers using QGIS and Excel.

Audit conclusions (related to analyse): Publicly available packaging containers are located too far from some households and these are not convenient for those citizens to use

Report (in English): [“Activity of the state and local governments in the collection and recovery of municipal waste” (2016)]

Lessons learnt:

It is good to have a colleague with good technical skills in some of the GIS programmes

- It is important to have good quality location data (coordinates or addresses) on data what you want to import into GIS (e.g location of packaging containers)
- GIS analysis shouldn’t be used only for visualizing the information, but also for quantitative analysis (e.g how many people live inside and outside the buffer zones)

Part 2: “Activities of the state in organizing waste treatment in oil shale mining and processing”

Overview

This case study deals with the closing of the oil shale waste landfill. The NAO used Lidar and GIS technology to arrive at the conclusion that Oil shale landfill was used to deposit waste in order to “recycle it later” which was not correct. It also concluded that most probably the closing works did not actually take place and audit informed the Environmental Inspectorate about that.

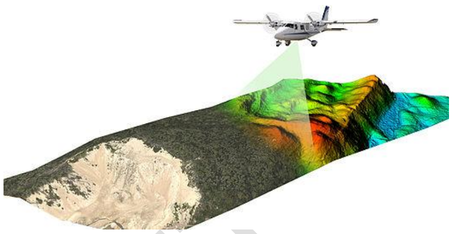
Audit: “Activities of the state in organizing waste treatment in oil shale mining and processing”

Task/question: Was the closing of the oil shale waste landfill carried out in accordance to the official documents (i.e landfill closing decision, technical project, building permit) and waste recycling legislation and within the time plan (2009-2012)?

Method used: Lidar[11], GIS analysis of aerial photographs of 2009 and 2013. Height information is collected with airborne laser scanner Riegl VQ-1561i. The equipment is mounted on the aircraft Cessna Grand Caravan 208B.

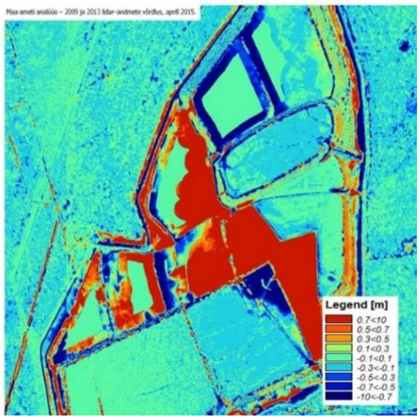
Data and Software

For image processing digital photogrammetric workstations are used. The used software was ArcGIS. (https://www.maaamet.ee/en/objectives-activities/photogrammetry)



Source of data: Estonian Land Board (Ariel photographs)

Steps taken to analyse the data



The analysis was carried out with the help of Estonian Land Board.

1. Comparison of the relief data of the oil shale waste landfill of the year 2009 with the relief data of the oil shale waste landfill of the year 2013.
2. The landscape changes were determined and this information was compared with the official documentation in order to determine whether the landfill was closed properly and according to rules, technical requirements and legislation.

Audit conclusions (related to analyse): Oil shale landfill was used to deposit waste in order to "recycle it later" which was not correct. We also determined that most probably the closing works did not actually take place and we informed the Environmental Inspectorate about that.

Lessons learnt:

- It is good to have a good cooperation with the institutions responsible on managing county's spatial data. They also can help with the analysis.
- Aerial photograph is a good source of controlling the real situation in the nature. NB! Have a look on the date/year when the aerial works were done- old photos might not reflect the latest situation.
- Land (or water level) elevation analysis can be used in audits related to mining, floods, landscaping, big construction works etc.

[11] Lidar is a surveying method that measures distance to a target by illuminating the target with pulsed laser light and measuring the reflected pulses with a sensor. Differences in laser return times and wavelengths can then be used to make digital 3-D representations of the target.

Instructor’s Notes – Module 3.3.2 Application of GIS for Environmental Audits Case Study 2: Audit on “Activity of the state and local governments in the collection and recovery of municipal waste” and “Activities of the state in organizing waste treatment in oil shale mining and processing”: NAO Estonia

[Notes in Power Point slides provide instructor’s notes]
The Power Point Slides are available at
https://drive.google.com/open?id=1mr2rXdi_JIqGIbLF6JjAzOP4hYKyJSmc .
The audit report referred above is available at
https://drive.google.com/open?id=10UnjsPK1bwwik8rhvVHgK2DqHSPQY9HZ .

MODULE 3.3.3 APPLICATION OF GIS FOR ENVIRONMENTAL AUDITS
CASE STUDY 3: AUDIT ON "ACTIVITIES OF THE STATE UPON PROTECTING GROUNDWATER": ESTONIA

<table>
<tr><td>Session title: Module 3.3.3 Application of GIS for Environmental Audits Case Study 3: Audit on “Activities of the state upon protecting groundwater” : Estonia</td><td>Session-at-a-glance</td></tr>
<tr><td colspan="2">Session Learning Objective:
The National Audit Office audited whether the activities that the state is using to prevent polluting or wasting groundwater ensure safe drinking water for the people and the preservation of rivers, lakes and bogs dependent on groundwater.
Participants would learn how GIS was used to audit planning phase and the results were used to make a selection of municipalities for the audit.

Key Points to be covered:
- Demonstrate use of GIS in selection of sample.</td></tr>
<tr><td>Teaching Method</td><td>Time</td></tr>
<tr><td>Part 1
Introduction
Brainstorming
Lecture and slides</td><td>
10 minutes
15 minutes
30 minutes
Total time: 55 minutes</td></tr>
<tr><td colspan="2">Link for Training Material of Module 3.1.1 to 3.3.3</td></tr>
<tr><td colspan="2">https://drive.google.com/open?id=1pLFjaJ3DUmXywhcyOVMIv8olMgeUIfvd .</td></tr>
</table>

Participant Notes: Module 3.3.3 Application of GIS for Environmental Audits Case Study 3: Audit on "Activities of the state upon protecting groundwater": Estonia

Overview

This case study would demonstrate how GIS was used to audit planning phase and the results were used to make a selection of municipalities for the audit.

Audit: "Activities of the state upon protecting groundwater" (2018)

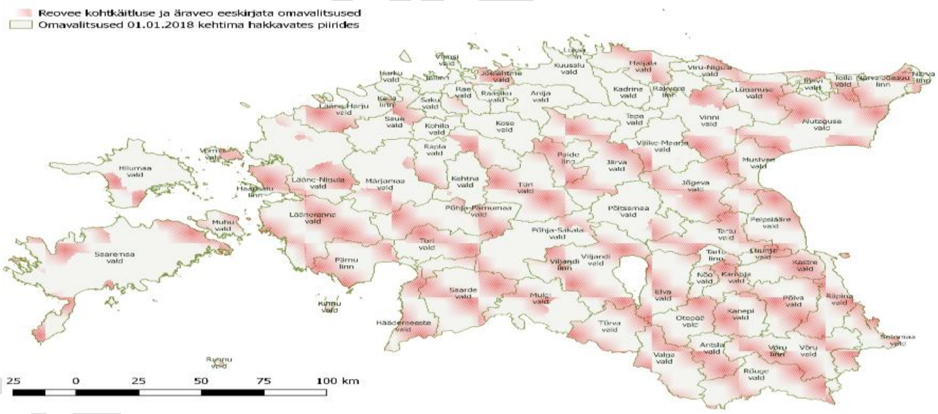
Task/question: What are the municipalities where the urban wastewater collection system situates in the area where groundwater is highly vulnerable to pollution and groundwater bodies are in bad or in endangered status? (NB! The analysis was made in audit planning phase and the results were used to make a selection of municipalities for the audit).

Method used: QGIS

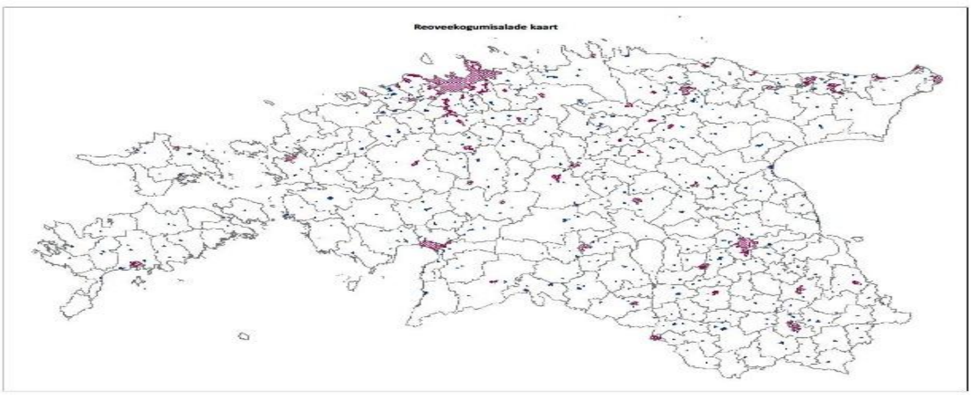
Source of data: Estonian Land Board, Estonian Environmental Agency

Steps taken to analyse the data:

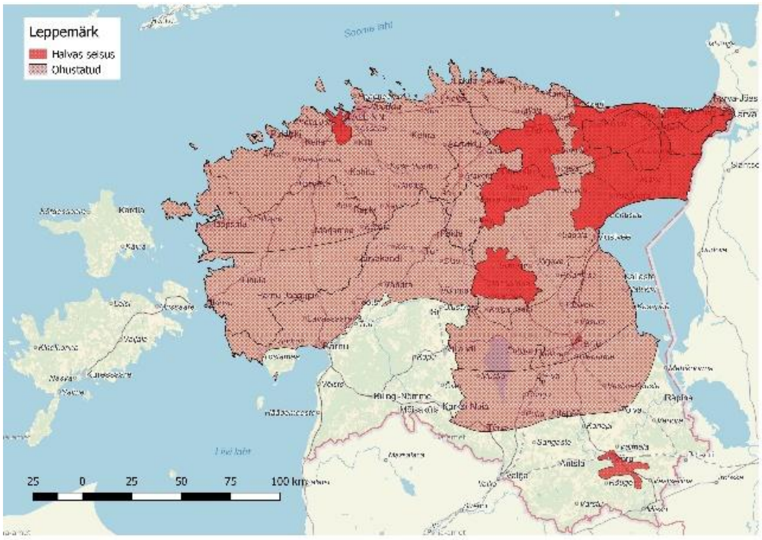
1. importing the data to QGIS on:
    a) municipalities



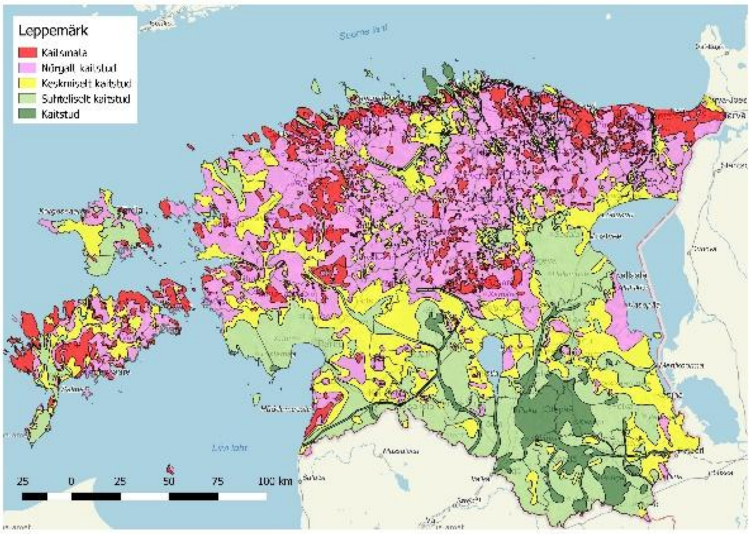
    b) location of urban waste water collection system,



c) groundwater bodies (which are in bad or in endangered status)



d) Vulnerability of groundwater (red – extremely high vulnerability, pink – high, yellow – medium, light green – low, dark green – protected)



2. Identifying the municipalities where urban waste water collection system (b) situates in the area where ground water is vulnerable (c, red or pink colour) and in bad or in endangered status (d)

Using the results (related to analyse): NAOE took sample of 20 municipalities for conducting more detailed audit activities

Report (in English): only summary in English

Lessons learnt:

- GIS can be used in planning phase of the audit (e.g to select a sample)
- It is important to define very clearly the objective/question of your analysis and to set the criteria (e.g what and where are the municipalities where groundwater is polluted (exceeding the water quality limits))?

Instructor's Notes – Module 3.3.3 Application of GIS for Environmental Audits Case Study 3: Audit on "Activities of the state upon protecting groundwater": Estonia

[Notes in Power Point slides provide instructor's notes]

The Power Point Slides are available at

https://drive.google.com/open?id=1zIZscbqMN_8MgeEmrP8LKODqBhEurM6X .

The audit report referred above is available at

https://drive.google.com/open?id=10UnjsPK1bwwik8rhvVHgK2DqHSPQY9HZ .

MODULE 4: DEALING WITH NON-AVAILABILITY OF ENVIRONMENTAL DATA AND FUTURE DIRECTIONS

Session title: Module 4: Dealing with non-availability of Environmental Data and future directions	**Session-at-a-glance**
The session will give participants an overview of the options available when quality environmental data is not available for audits. The participants would be made aware about the future potential of technology in using data in environmental audits. **Key Points to be covered:** - Options when data is not available - future directions in environmental data - Use of data analytics and big data	
Teaching Method	**Time**
Introduction	05 minutes
Brainstorm / Flip-charting	05 minutes
Lecture and slides	15 minutes
Individual exercise and group discussion	20 minutes
Lecture and slides	20 minutes
Case study and group discussion	20 minutes
Feedback and questions	05 minutes
	Total time: 90 minutes

Participant Notes

Overview

This module is focused on challenges and future directions / trends in use of data for environmental audits. The major challenge is lack of data or incomplete data with the public authorities. This problem is more prevalent in case of environmental sectors. This module is expected to introduce the participants the alternative course of action in execution of environmental audits. Further, participants are also expected to be exposed to the newer trends in respect of environmental data due to increasing use of technology in acquiring and using data by public sector as well newer tools and techniques being explored by auditors. The session is expected to trigger further learning by the participants on their own and preparing themselves to face the challenges of environmental audits.

Basic Concepts

Key Teaching Point 1 – Dealing with non-availability of data

Many SAIs in both developing and developed countries have reported challenges in planning and conducting audits when they lack high-quality environmental data. High quality environmental data should possess characteristics such as Relevance, Accuracy and reliability, Timeliness and punctuality, Accessibility and clarity, Coherence and comparability and Availability of metadata.

Absence of any of above mentioned characteristics will pose a challenge in using environmental data for audit. SAIs need to find other options to overcome the problem of non-availability of data or availability of low quality data.

Key Teaching Point 2 – Options when related data are available

Many SAIs in both developing and developed countries may face challenges in planning and conducting audits when they lack high-quality environmental data. When high-quality environmental data are not available, options still exist for SAIs to plan and conduct audits. For example, SAIs may be able to use related data to estimate unavailable data or develop their own data through a survey.

SAIs may also decide that the absence of high-quality data will be a central audit message. SAIs should consider certain factors in deciding which options to pursue, including the quality of alternative data sources, the costs of generating or obtaining data, and the expected use of the data in the audit. When the desired data are not available from the audited entity or other sources, alternative data may be used. For example, SAIs can estimate unknown environmental data from other known information and can develop or use existing models that combine alternative environmental data. When using related data, SAIs should consider their sufficiency and appropriateness for the audit's purpose.

THE USE OF RELATED DATA TO ESTIMATE UNAVAILABLE DATA

When SAIs lack high-quality environmental data, they may use alternative types of data that relate to, or help estimate, unavailable data. For example, the SAI of Morocco noted that similar data from government agencies other than the audited entity can act as benchmarks to estimate unavailable data. Another way of estimating unavailable data is to compare several data sources. SAIs may need to consider certain issues when using data estimates, such as how reliable and credible the estimates are. Estimates can also affect how a SAI reports the data in its audit work. For example, SAIs may need to qualify the estimates by framing them in terms of upper or lower bounds.

THE USE OF MODELS TO COMBINE RELATED DATA AND IDENTIFY TRENDS

SAIs may use models to combine related environmental data to evaluate how effective programs are. Models tend to be more complex than estimates and can be used to represent complex relationships among factors, as well as to integrate data to evaluate environmental programs. For example, aerial photographs could become the basis for developing a spatial model that illustrates changes in land use over time. Cameras mounted on an aircraft or other flying object could provide an instant visual inventory of a portion of the earth's surface and can be used to create detailed maps. SAIs can use computer models to compare data from several sources to determine whether the data maintained by the audited entity are reliable.

SAIs can also evaluate a government entity’s modelling efforts that draw upon environmental data, or the impact of models on program effectiveness. For example, the SAI of the US assessed whether the government incorporated sufficient information such as actual monitoring data on pollutant levels in its model to determine water quality, and evaluated how appropriate the model’s assumptions were. Before using models, SAIs should consider several issues, including their own capacity to use and evaluate models, the cost-effectiveness of using models, and the availability of the models compared to other alternatives.

<u>Key Teaching Point 3 – Options when related data are not available</u>

Several options are available to SAIs when no high-quality or related environmental data are available. For example, the lack of data could become a key message in the audit if the SAI finds that the government lacks sufficient data to manage a program effectively. SAIs can also consult individual experts or scientific panels, and use their opinions to form the basis for findings. Alternatively, SAIs can develop their own data to meet the audit’s needs through questionnaires, surveys, or physical observations from site visits.

THE USE OF THE ABSENCE OF DATA AS THE AUDIT’S CENTRAL MESSAGE

SAIs can make the lack of high-quality environmental data an audit finding by, for example, reporting on the data's incompleteness and poor reliability. For example, the SAI of Uganda reported that the government lacked records for waste management to show that it properly disposed of medical waste. SAIs can also focus the audit's message on the impact of a lack of data on how the program is managed. For example, the SAI of Estonia reported that because data were unavailable, governments could not provide evidence that they had complied with laws for preserving nature habitats.94 Data unavailability can also affect government entities' ability to make decisions. For example, the SAI of Canada found that a government entity did not have enough data on oil spill risks to predict oil spills and effectively plan emergency responses.

THE USE OF OPINIONS FROM EXPERTS OR OTHER RELEVANT PARTIES AS THE BASIS FOR FINDINGS

Another option when high-quality data are lacking is to consult with experts or other relevant parties to get opinions to support audit findings. SAIs may use a variety of methods to collect these views such as interviews, focus groups, structured questionnaires or surveys, or they may contract with a consultant to gather relevant opinions. For example, the SAI of Estonia used expert opinions about the conditions of certain habitats to identify and describe inconsistencies or other gaps in government data relevant to their field. SAIs can also combine expert opinion with other data. For example, the SAI of Bhutan obtained expert opinions and drew on relevant literature to assess a national forest inventory.

SAIs can also use the opinions of other relevant parties, such as stakeholders or program recipients, to develop audit findings. For example, a SAI can use a survey to collect views from stakeholders to assess how effective programs are over a period of time. For example, the SAI of Switzerland worked with the country's customs agency to survey companies to determine the impact of air emission tax policies between 2000 and 2006. Depending on the SAI's needs and resources, SAIs can also use consultants to gather opinions for them. For example, the SAI of the UK used a consultant to survey farmers' opinions on certain agricultural schemes to determine if they were beneficial and the process was easy to understand.

According to INTOSAI auditing standards, SAIs should consider the competence, capabilities, and objectivity of those whose expertise is required to obtain needed information. When deciding whether to obtain opinions from experts or third parties, SAIs should consider whether the experts and other parties represent a range of viewpoints and certain types of expertise are required to speak on the issue. Also, SAIs should consider whether a single expert opinion can be the basis for a finding or whether a panel of experts is more appropriate for the specific topic.

THE DEVELOPMENT OF ALTERNATIVE DATA TO MEET THE AUDIT'S NEEDS

When faced with a lack of high-quality environmental data, SAIs can develop their own data to support audit findings—such as compliance with national laws—through site visits or photographs. For example, the SAI of Paraguay inspected a sample of sites affected by tanneries to determine whether the tanneries in the selected area were complying with environmental laws. SAIs can also develop test data to demonstrate how feasible it is to collect such data when SAIs are thinking about recommending that the government collect new data.

When deciding whether to develop new data, SAIs should consider how cost-effective the work will be, what impact the data will have on the audit work, and whether the SAI has the capacity or expertise to develop the data. For example, if a SAI collects water quality samples, it may need to have the samples collected by a specialist and analyzed in a certified laboratory to ensure that standardized methods and quality controls were applied. A SAI may also need a specialist's assistance to properly interpret the data. SAIs should also consider whether they can develop data covering all the relevant areas, whether they would need to use a sample that would be statistically representative, or whether their data, and possibly conclusions, would be more limited. If they ask for computer-processed data as part of a survey, SAIs should consider including questions on data reliability in their surveys.

Key Teaching Point 4 – Future directions in environmental data

Program managers are expanding their use of data from GIS and from other new tools, such as social networking, to manage their programs. These tools and other trends create opportunities for SAIs themselves to use the tools and the resulting data. At the same time, these tools and trends create challenges arising from the need to audit different kinds of environmental data. Managers of environmental programs are relying more and more on GIS, combined with satellite-based observations, to measure results and manage their environmental programs. As an example, the Ethiopian Wolf Conservation Programme prepared maps of a recent rabies outbreak based on several different layers of information, including points (e.g., locations where affected wolves were found), lines (e.g., waterways), regions (e.g. province or country), and notes (e.g. scale, values). These maps were then used to track wolf deaths from rabies contracted from dogs—a major threat to the endangered species—against previous data on pack location and viable habitats. The result was more efficient vaccine targeting.

Such tools also place demands on program managers. To use GIS and satellite technology in this way, they need to maintain sophisticated on-the-ground data collection capabilities to validate and calibrate remotely sensed data, and to devise back up plans if problems arise. They also need the expertise to use the technology appropriately and need good quality controls to ensure the integrity and accuracy of the data used as inputs.

These spatial data tools may be useful to SAIs in their audits to evaluate environmental programs and issues. The 2010 WGEA guidance paper entitled Auditing Forests: Guidance for Supreme Audit Institutions describes how SAIs can use GIS technology in auditing forest management, for instance to determine the extent of deforestation, illegal logging activities, and illegal land use, or to identify the locations of forest fires.

The 2013 WGEA guidance paper entitled Land Use and Land Management Practices in Environmental Perspective: describes how Computer-based Geographical Information System (GIS) can map and analyze geographical spatial data and integrate what appears on a map with, for instance, data from geological analysis and general database operations. The 2004 WGEA guidance paper entitled Auditing Water Issues: An Examination of SAIs' Experiences and the Methodological Tools They Have Successfully Used: how Geographic Information Systems (GIS) has been used by SAIs, as a tool to identify changes in water resources or high-risk areas prone to water related natural disasters, such as erosion, flooding, and drought.

With program managers drawing on such information more and more as an integral part of their monitoring systems, SAIs will need to be able to evaluate the quality of the data that result. For example, SAIs may be concerned about quality controls, such as training and the use of expert judgments. In addition, microchip and wireless technologies are offering program managers new options for data collection. For example, fisheries managers are using tracking devices on fish to map how different regions of the ocean are used in situations when human observers are not feasible. Such technologies entail challenges such as ensuring that the sensors are operating accurately. These technologies will also place additional demands on SAIs to understand the strengths and limitations of such data sources. We have not identified any use of these technologies by SAIs.

Finally, with tight and, in some cases, shrinking government budgets, program managers will be under more pressure to demonstrate results. This situation may lead them to place greater weight on environmental data and the indicators derived from the data to show that their programs are working as intended. To reduce costs, they may also shift away from on-the-ground data collection efforts to methods that estimate or model the results. Such changes may affect how performance is measured and potentially how both program managers and SAIs evaluate programs. Another result of shrinking budgets may be that program managers may increase their emphasis on partnerships and on finding ways to coordinate efficiently with other parties. Different levels of government and other players may discuss who will do what, and who will pay for what.

Program managers may also shift to using data that others generate and maintain. One example is the Global Biodiversity Information Facility, an international organization governments created to encourage free and open access to biodiversity data. However, it may be difficult for SAIs to obtain access to environmental data and to get enough information about the data to ensure that they can be used appropriately and in keeping with good auditing practices.

Key Teaching Point 5 – Data Analytics and Future of environmental data

Increased availability of environmental data and increased computing power is changing the way environmental data can be analysed, investigated and reported. Increased data availability needs to be matched with improved methods of analysis and reporting. Data is only turned into information if it is able to be understood and provides a basis for decision making. Increased data availability requires efficient data analysis tools. Many programs exist to aid this process.

Technology plays a significant role in modern day governance for enhancing delivery of public goods and services. The diverse technology systems are continuously producing volumes of data in disparate forms, throwing up immense opportunities for data analytics. As a responsive Supreme Audit Institution, we have to be institutionally agile to keep pace with such developments and embrace the evolving opportunities in data analytics.

Data is available to audit today, in different forms and from different sources. Data analytics provides the potential to analyse these data sets and obtain insights to assist in the audit processes by identifying patterns, trends, descriptions, exceptions, inconsistencies and relationships in data sets and their variables. The insights so drawn would assist in setting the direction of the audits, by primarily identifying areas of interest or risk and in identifying exceptions.

Data analytics begins with identification and collection of various data sources for a particular audit. The analysis of data through various data analytic techniques will yield insights on the working of the audited entity. The risk areas or areas of interest identified through such an exercise will assist in identifying audit objectives and developing an Audit Design Matrix. Data Analytics will also assist in identifying the sample of audit units where substantive checks will be conducted.

The various analyses can then be built into a re-executable Data Analytic Model. This will ensure that results of data analysis can be used repetitively with periodic updating of data. Establishing a mechanism for receiving data periodically will be crucial for such an approach. The scope of the model once built can be expanded by incorporating the feedback from substantive checks and bringing in additional data sources.

Source: Guidelines on Data Analytics (2017) by SAI India.

Citizen environmentalists / Citizen Science

Citizen science is the involvement of the public in scientific research – whether community-driven research or global investigations. The Citizen Science Association unites expertise from educators, scientists, data managers, and others to power citizen science. Join us, and help speed innovation by sharing insights across disciplines.

Citizen science offers a unique opportunity for the public and environmental organisations to connect about environmental science and environmental protection. Citizen science is the involvement of the public in scientific research – whether community-driven research or global investigations. Citizen science mobilizes the public to participate in the scientific process to address problems. This can include identifying research questions, collecting and analyzing data, making new discoveries, and developing technologies and applications.

A citizen science program can engage communities to collect data and advocate for environmental concerns. Citizen science programs include the participation of nonscientists in the process of gathering data according to specific scientific protocols and in the process of using and interpreting that data.

Low cost mini sensors

Development of low cost mini sensors is opening a new window of opportunity in use of data for environmental monitoring. Critical situations where data cannot be gathered with conventional approached can be cost efficiently and quickly managed with mini sensors. The evolution of low cost sensors has resulted in a number of instruments becoming commercially available.

Customizable systems

Wireless Sensor Networks (WSNs) has proved itself as an emerging technology that can be used to solve many real-life problems. WSN's consists of a large number of sensor nodes that are smart enough to not only monitor the changes happening in the environment in real time but also intelligent enough to gather the information related to it. These sensor nodes are very cost effective if they are being compared with those traditional one's, they are small and come up with limited power, computational and processing resources. These sensor nodes also possess the capability to communicate with each other in order to transfer the collected data to the sink node which is the main central location responsible for wholesome management of all other sensor nodes in WSN. To make the communication possible, a network is formed by distributing the sensor nodes across the environment. Once the network is formed, all the sensor nodes liaise with each other to carry out computing and sensing activities. Its cost effectiveness, better output giving capability, easy implementation and troubleshooting and no mess of cables, are the reasons behind its wide popularity and adoption. Today's wireless sensor networks are bi-directional and are intelligent enough to control the sensor activities. WSNs are used most effectively in monitoring and sensing many physical phenomenon happening in the environment such as change in temperature, humidity, intensity, pressure and even in the field of health care, WSN has gained popularity.

Key Teaching Point 6 – Big Data and Future of environmental data

"Big data" is commonly defined as data that are too large, created too quickly, or structured in such a manner as to be difficult to collect and process using traditional data management systems. Big data sets and analytics increasingly are being used by government agencies, non-governmental organizations, and private firms to forward environmental protection. Improving energy efficiency, promoting environmental justice, tracking climate change, and monitoring water quality are just a few of the objectives being furthered by the use of big data. There are five characteristics of Big Data:

Volume: Current storage makes it possible to acquire enormous volumes in a short space of time. Big Data gathers data every second. We no longer measure this data in gigabytes, not even in terabytes, but the next stages up (by the petabyte, exabyte, zettabyte and yottabyte).

Variety: Present technologies make it possible not just to acquire enormous sets of data, but also allows for variety. For any analysis, variety is as important as volume. Much of the world's stored data is unstructured but Big Data allows for structuring and unstructured information.

Velocity: To be relevant, Big Data must be able to cope with the speed at which data is generated in order to store it and retain the most up-to-date and relevant information.

Veracity: This seeks to determine a data set's accuracy and integrity, not just of the data but also the sources that generate it. If there is no trust in the data source, the data itself is virtually useless.

Value: More concerned with the results that users extract from the Big Data, if we cannot make sense of the data then it has no value. The exercise in capturing and storing the data will have been completely useless and the exercise a waste of time, money and resources.

There are various examples of the ways big data sets and analytics are being used to achieve environmental and sustainability goals. Various governmental and non-governmental entities are using big data sets and analytics for environmental protection. In addition, several government-wide initiatives have laid the groundwork for increased use of big data by establishing government policies with respect to information collection, storage, and availability that broaden public access to government and government–supported data. Big data and data analytics are being used for environmental monitoring.

Examples of use of environmental big data

The Environmental Protection Agency and Public Health

One of the biggest areas in the US for unifying big data with environmental science is public and environmental health. There are improvements in the monitoring and mitigation of toxicological issues of industrial chemicals released into the atmosphere. Monitoring has always used the tried and tested methods such as localized environmental sampling, but now such data can be processed through computational methods, the result is more accurate, more up-to-date, faster produced, with more analytical information to allow experts to make an informed decision. Big Data allows for high throughput (more resources, a longer period of time), combined data sets (bringing together multiple, otherwise seemingly disparate data sets) and meta-analysis (studies that are the compilation of existing studies to create a more thorough and hopefully accurate picture), and deeper analysis of the results produced from these studies.

For Geographic Data

Maps are one of the most useful tools in environmental sciences. From simple cartography for naval navigation, geographic surveying, to modern uses for Geographic Information Systems (databases of data sets from which we can produce digestible maps and create visually striking imagery for an intended audience), GIS thrives on Big Data. Much of GIS strength lies in its ability to consolidate, utilize and present statistical data. The more data you have from a geographic area, the better the quality of the output and the more informed the decision making is likely to be. Its biggest contribution (so far) seems to be in spatial analytics, and that's good news for GIS technicians and for those people charged with making decisions based on the outputs of their data.

One example is in disaster and emergency relief. Satellite data and aerial imagery have already informed GIS in disaster management, with Hurricane Katrina being one of the first and best-known choices in using the technology. In future, Big Data will further enhance its efficacy.

Climate Change and Planetary Monitoring

In 2013, the UK government announced large-scale investment in Big Data infrastructure for science, particularly in the environmental sector. Of particular note to global research was a commitment to maintaining funding for a program called CEMS (Climate and Environmental Monitoring from Space). This allowed for the creation of larger databases to cope with the upcoming Big Data revolution and to allow research partner organizations to work with more data and produce more results. With a specific focus on climate change and planetary monitoring, CEMS storage removed the need to download enormous data sets while reducing the cost of access. It provides the tools as well as the data, allowing for greater efficiency, sharing in the academic community, and providing resources once beyond the reach of many institutes due to budgetary restrictions alone. Along with Cloud data, this is now the standard globally for some of the world's top research institutes. There are immense implications for the uses of Big Data for climate modeling. As early as 2010, NASA was utilizing Big Data capture and storage for creating climate models to make the most accurate climate projection models yet.

Advantages of using big data:

Collecting, Sorting, Analyzing, Presenting Quickly

Big Data's major advantage is in the capacity to collect masses of data and analyze it quickly; it's a realistic cost and resource saving tool in areas often drastically underfunded and having to cut costs. The storage capacity now exists to collect and collate; the computing power is also affordable to process and manipulate in any way necessary.

Error Mitigation

How to handle errors in data, reporting, rogue data and anomalous results has been one of the biggest problems facing any science. When sample sizes are too small, anomalous data can be given more importance than it deserves. But studies are often limited by sample size alone due to resource factors. The larger a data set, the more likely a rogue piece of information will fall in significance and not damage the overall result. Coupled with the cost and resource saving, environmental studies can, in theory, become larger and more thorough, producing more accurate results.

Better Environmental Management

This applies to urban management as our cities continue to undergo rapid and vast changes in line with changing technology and demands of residents. In one study, the Norwegian capital of Oslo was able to reduce its energy consumption through the application of Big Data Analytics when examining its energy resources.

Better Decision Making

By sheer weight of numbers, Big Data and the analytical tools used in its processing is able to process and analyze more past data than ever before. Previously, this too was limited by resources but with its increased access and availability, it is expected to permit easier presentation and reporting, delivering more confident results and therefore, better to aid decision makers and policy development professionals. Scientists and government can work together more efficiently in future, not just to react to the environmental problems of today, but work with greater foresight today to make better decisions for tomorrow.

Challenges for Big Data for Environmental Sciences

Like any other emerging technology, there are problems and limitations in use of big data analytics for environmental audits. The following are few limitations in use of big data analytics:

Technical Limitations

Due to the complexity of so-called Big Data, the method presents a number of other challenges to those who seek to acquire and use it. Big Data's increase is and so far, has been, exponential in growth. To keep up, hardware in all of the areas above will need to keep up, if not exceed the necessary capacities. There is also a possibility of human error - wrongly entered data, poor processing due to mistakes, and interpretation of that data.

Ethical Limitations

Ethical issues include problems such as cultural sensitivities as in archaeology and anthropology. Some critics are concerned that in reducing populations to Big Data information, we reduce their humanity, their individuality. However, with the improvements in disaster response time, applications in climate science, and in the enormous data processes when examining archaeological/anthropological information, it's likely that these human sciences and humanities concerned with the environment will benefit in the long-term.

Lack of Widespread “Open Access”

Research institutes and businesses are often incredibly protective of their research data, especially where mass profitability is involved. Yet there has been a move in recent decades to call for subscription-free public access to scientific data. Known as Open Access, not enough strides have been made in this area, in some disciplines, that Big Data Analytics is not presently experiencing its full potential and much data is restricted, meaning that - although studies can call on more data and do more with it - there is still a large amount of data that could prove useful in environmental science, held privately with limited or no access. Although fear of handing over information to competitors is part of the issue, other problems include lack of resources to do so or a lack of awareness of how useful Open Access can be.

Summary

Many SAIs in both developing and developed countries have reported challenges in planning and conducting audits when they lack high-quality environmental data. SAIs have been resorting to various options like relying on related data, use of models, expert opinions, development of alternative data etc. As a last resort SAIs can also make the ‘lack of high-quality environmental data’ an audit finding by, for example, reporting on the data’s incompleteness and poor reliability. SAIs can also focus the audit’s message on the impact of a lack of data on how the program is managed.

The program managers are expanding their use of data from GIS and from other new tools, such as social networking, to manage their programs. These tools and other trends create opportunities for SAIs themselves to use the tools and the resulting data. These efforts are also augmented now by use of data analytics and newer concepts like Big Data.

Data analytics is the application of data science which is an emerging area of work concerned with the collection, preparation, analysis, visualization, management, and preservation of large collections of information; approaches to gain insights from data. Data analytics provides the potential to analyse these data sets and obtain insights to assist in the audit processes by identifying patterns, trends, descriptions, exceptions, inconsistencies and relationships in data sets and their variables. The insights so drawn would assist in setting the direction of the audits, by primarily identifying areas of interest or risk and in identifying exceptions.

The Auditors have not reported wide use of these tools but the trends in greater use of technology for data management would increase their use substantially. Big Data is extremely versatile, a necessity for the many different environmental needs. It can be used to monitor an area as vast and expansive as the Amazon Rainforest, or it can monitor a small city's water supply. Big Data brings important traits for enhancing environmental protection viz. it allows entities to gather more data than ever before and that too at increased speed and ease of obtaining data.

Instructor's Notes – **Module 4: Dealing with non-availability of Environmental Data and future directions**

Introduction (05 minutes): Show slides 1 - 2 and introduce the session overview and learning objectives.

[Notes in power point slides provide instructor's notes]

Brainstorming / Flip-charting (05 minutes): Show slide 3 and hold group brainstorming about the challenges related to data faced during environmental and / or other audits, record responses provided by the class on a flipchart for reference throughout the class.

[Notes in power point slides provide instructor's notes]

Lecture (15 minutes): Show slides 4 - 6

[Notes in power point slides provide instructor's notes]

Individual Exercise and group discussion (20 minutes): Show slide 7 and conduct individual exercise with group discussion afterwards.

[Notes in power point slides provide instructor's notes]

Lecture (20 minutes): show slides 08 - 10

[Notes in power point slides provide instructor's notes

Case study and group discussion (20 minutes): Show slide 11 and distribute case study about use of data analytics and / or big data in environmental audit and / or other sector conducted by SAI or which have potential of using these tools

[Notes in power point slides provide instructor's notes]

Feedback and questions (05 minutes): Show slide 12