IFAD GEF Project :
Cross-cutting capacity building,
knowledge services and coordination project for the

Food Security Integrated Approach

Pilot Program

(Ethiopia, Uganda, Ghana, Burundi, Swaziland, Kenya, Senegal, Burkina, Niger, Malawi, Tanzania and Nigeria)

Dealing With Data

A Progress report submitted by UNEP (2020)

UN environment
United Nations Environment Programme


Michael Major/Crop Trust

IFAD GEF Project :
Cross-cutting capacity building,
knowledge services and coordination project for the

Food Security Integrated Approach

Pilot Program

(Ethiopia, Uganda, Ghana, Burundi, Swaziland, Kenya, Senegal, Burkina, Niger, Malawi, Tanzania and Nigeria)

Dealing With Data

A Progress report submitted by UNEP (2020)

UN environment
United Nations
Environment Programme

Dealing With Data

Training

Contents

Data Management

Why Manage Data?

Successful management of data is critical for any organization to make informed decisions from a leadership perspective.

When done right, data management will effectively help an organization drill down from vast amounts of information gathered to effectively making decisions.

For this to succeed, data must be **categorized**, **analyzed**, and **organized.**

Data Management

Importance of data management

- Data management plays a significant role in an organization's ability to generate revenue and controlling costs.
- Successfully being able to share, store, protect and retrieve the ever-increasing amount of data is a key competitive advantage for organizations today.
- Data management helps organizations to mitigate risks effectively.
- It enables organizations to make risk-based decision making.

Data Management

Costs of poor data management

- Misinterpretation of the data is a common occurrence.
- Data is often lost, and critical data is unavailable or obscured.
- Data could be completely inaccessible for meaningful use.
- It would be difficult to defend data due to poor quality/issues with incomplete data.
- Leads to significant waste of time and money repeating data gathering or correcting bad data.
- Causes missed deadlines and opportunities.
- Impacts management decision-making as a result of poor quality data.

Data Management

Benefits of data management

- Maintaining optimum data quality for use/reuse.
- Improved user confidence when dealing with stored data.
- Efficient and timely access to data.
- Improved knowledge and understanding of available data.
- Boosting confidence in decision making.

Data Management

Scope of data management

- Data Management is a broad field of study, it encompasses the processes of managing data as a resource that is valuable to an organization or business. This includes can also include development and execution of architectures, policies, practices, and procedures in order to manage the information lifecycle needs in an effective manner for the organization.
- The following elements all fall under the scope of data management:
  - ✓ Data Modeling
  - ✓ Data Warehousing
  - ✓ Data Movement
  - ✓ Database Administration
  - ✓ Data Mining

Data Management

Scope of data management

- **Data Modeling**: Creating a structure for the data collected for use, and then organizing this data in a way that is easily accessible an efficient to store for analysis.
- **Data Warehousing**: is storing data effectively so that it can be accessed and used efficiently in future.
- **Data Movement**: The ability for an organization to move data from one place to another.
- **Database Administration**: Processes, roles, and responsibilities associated with the individuals who are the administrators of their respective database environments within an organization.
- **Data Mining**: The process in which large amounts of data are sifted through to show trends, relationships, and patterns.

Data Management

Traditional data management

Any file-based approach to data management is usually an outdated practice and must be avoided. It leads to the creation of information/data silos within an organization and limits the effective use of data for critical decision making.

If you notice your organization has any of the following issues, then it is possible that a file-based data management is being used in practice:

- Data often depends on the file or program that it is stored in.
- Many file formats are used that are often incompatible with each other.
- Data sharing and availability are frequent issues when looking to do any analysis.
- Data gathered is often redundant across multiple sources.
- Data is often isolated within a department/system.
- Data isn't flexible enough to be used for different reporting purposes.
- Administration of data is too complex, resulting in process and security issues.

Data Management

Modern data management

Most organizations today use Database Management Systems (DBMS). These are specialized pieces of software that enables an organization to centralize data, manage it efficiently, and provide access to the stored data via application programs such as MySQL, MS Access, Oracle etc.

It is important to note that using a DBMS does not automatically eliminate all the problems that exist for file-based data management approaches. However, it does make the following much easier, :

- Creating a common software platform for creating and maintaining databases.
- Effective central management of stored data.
- Creating an interface between applications that use data and the databases which house the information.
- Separating logical and design views of the data.
- Simplifying administration and security for databases.

Data Management

Common terminology

The following are useful common terminology and definitions:

- **Data**: collection of facts and figures processed to give information.
- **Information**: processed data.
- **Entity** (table or relation): a person, place, thing, or event for which data is collected and maintained.
- **Field** (attribute or column): a single characteristic or fact about an entity. A field is used define and store data.
- **Record** (row): a logical collection of fields that describes one instance of an entity (person, place, or object).
- **File**: a set of related records that contains data about a specific entity.
- **Database**: shared collection of logically related data that is organized so that it can be easily accessed, managed, and updated.
- **Attribute**: A characteristic of an entity; something the entity is identified by, such as "Customer Name", or "Employee Number".
- **Keys**: A field or set of fields in a record that is used to identify the record.
- **Database system**: A complete information system (database and DBMS).

Data Management

Real-world value of DBMS

The following are some common examples of the value-added by DBMS to industries in the real-world, in addition to the usual application of DBMS to process tax, payroll, and other common operational activities:

- **Banking**: DBMS is relied on heavily to carry out all functions such as daily transactions, audits, marketing, investigations, providing services, and products, etc.
- **Airlines**: DBMS is relied on to support customer reservations, customer profiles, plane schedules, and aircraft staffing availability, etc.
- **Universities**: DBMS is used to keep student records, research data, alumni information, events management, and housing/enrolment information etc.
- **Retail**: DBMS is used to store customer profiles, transactions records, products, purchases, and behavior data, etc.
- **Manufacturing**: DMBS is used to drive production volume, inventory management, order information, and supply-chain management, etc.

Data Management

Functional components of DBMS

- **Data Definition Language (DDL)**: It defines each element as it appears in the database. The DDL is the formal language programmers use to specify the content and structure of the database.
- **Data Manipulation Language (DML)**: It is a set of procedural commands that enable programmers to append, modify, update, and retrieve data. The DML uses simple verbs like sort, delete, insert, select, display, etc.
  - *A query language* - It enables the user to make queries from the database. It is a standard data manipulation language for relational database management systems.
  - *SQL Report Generators* - It enables generation of reports from a database. The programs enable reports to be presented using pictures, graphics, and maps, etc.
  - *Application Generators* - Most DBMS' include programming facilities available in 4th Generation Languages (4GLS).
  - *User Interface* - This is a shell that provides the environment for interaction of a user with the database.

Data Management

Practical advantages of DBMS and the database approach

- **Control of data redundancy**: The database approach attempts to eliminate the redundancy by integrating the file. Although the database approach does not eliminate redundancy entirely, it controls the amount of redundancy inherent in the database.
- **Data consistency**: By eliminating or controlling redundancy, the database approach reduces the risk of inconsistencies occurring. It ensures all copies of the data are kept consistent.
- **More information from the same amount of data**: With the database approach, it may be possible to derive additional information from the same data due to its ability to derive relationships with other sets or data.
- **Sharing of data**: Databases belonging to the entire organization can be shared among all authorized users.
- **Improved data integrity**: Database integrity provides the validity and consistency of stored data. Integrity is usually expressed in terms of constraints, which are consistency rules which the database is not permitted to violate.

Data Management

Practical advantages of DBMS and the database approach

- **Improved security**: The database approach can provide a level of protection to the data from unauthorized users based on how a DBMS is setup. This includes user-names and passwords to identify specific user groups which might access the data, or perform any operations on the data including retrieval, insertion, updating and deletion.
- **Enforcement of standards**: This approach makes it easier to enforce the necessary minimum standards for data formats, naming conventions, documentation standards, update procedures, and access rules.
- **Increased concurrency**: Databases can manage concurrent data access effectively. It ensures that multiple users of the dataset would not cause interference between users or result in any loss of information and integrity.
- **Improved backing and recovery services**: Modern database management systems provides settings and configurations that help minimize the amount of data lost following a system failure by using the transaction-based approach.

Data Management

Challenges faced by DBMS and the database approach

- **Complexity**: Database management systems are extremely complex pieces of software. This requires trained professionals working full-time to help the organization make the most of this asset.
- **Size**: DBMS is not a small system, and often requires significant physical resources to maintain on-site.
- **Cost**: DBMS' can easily run-up substantial costs for organizations – especially considering that one needs to employ administrators, maintain hardware, and have valid software licenses, etc.
- **Cost of transformation**: Moving to a DBMS is an equally expensive exercise, as it would involve hiring professional consultants to do discovery and implementation, project managers for execution, contractors, and vendors for licenses, management, and training etc.

Data Management

Challenges faced by DBMS and the database approach

- **Performance**: As the database approach is to cater for many applications rather than exclusively for a particular one, some applications may not run as fast as before.
- **Impact of failure**: The database approach increases the vulnerability of the organization to a system failure due to its centralized nature. This can have far-reaching consequences if an organization is using a DBMS to deal with a large volume of customers that are from the general public – such as retail, airlines, or government services.

Data Management

Key principles of data management

- **Data Independence**: This is used to describe the separation of data or data handling from the functional processing of the data and the programs that use the data.
- **Data Integrity**: The ability to ensure that data stored cannot be repudiated as being incorrect due to unauthorized modification.
- **Data Redundancy/Duplication**: This describes the case where a data element is individually kept in several places (records, files, etc.) within the database.
- **Data Security**: The ability of a database system to preserve and protect the data which it holds.

Data Management

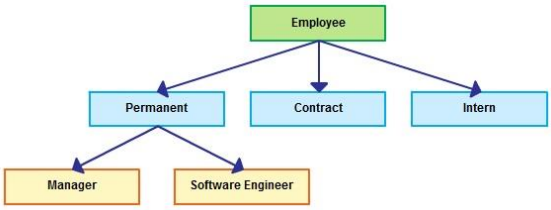
Database models used for data management

- **Conceptual Models**: These cover the conceptual nature, or theoretic nature, of data representation.
  - **One-to-One** (1:1) - In this, one record in a table is associated with one and only one record in another table.
  - **One-to-Many** (1:M) – In this, one row in a table may be linked with many rows in another table, but not vice versa.
  - **Many-to-Many** (M:M) – In this, multiple records in a table are associated with multiple records in another table.
- **Implementation Models**: These cover how the data is physically represented in the database in practice.
  - **Hierarchical** – Here, data is stored in a tree-like structure and supports one-to-many parent-child relationships.
  - **Network** – An expansion of the hierarchical model, with an owner-member relationship in which a member may have many owners
  - **Relational** – This is the most popular type of database model, where tables are related by sharing common entity characteristic(s). This model uses primary key to link records.

Data Management

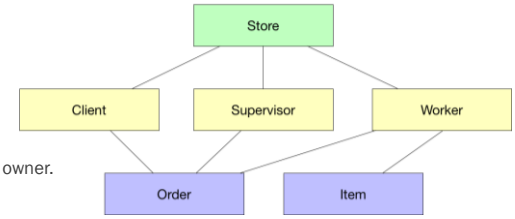
Hierarchical database models

- **Advantages**:
  - Data independence is easy to maintain.
  - Database security, such as role-based access by type of data, is easy to apply.
  - Simplest form of database concept to grasp.
- **Disadvantages**:
  - Complex implementation to keep conceptual model true.
  - Difficult to manage and lacks standards for implementation.
  - Lacks structural independence.
  - Complex for applications to handle due to programming requirements.
  - Limits implementation and use due to restrictive nature.



Data Management

Network database models

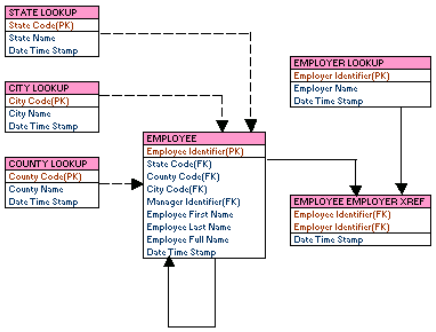


- **Advantages**:
  - Conceptually simple and easy to design.
  - Handles more relationship types than a hierarchical model.
  - Data access flexibility due to structure.
  - Promotes database integrity: a member does not exist without an owner.
  - Conforms to standards of data independence.
- **Disadvantages**:
  - System complexity is an issue, as a user-friendly database management system cannot be created using the network model, users need to be familiar with the internal structure.
  - Making structural modifications to the database requires specific application programs to be modified first before anyone can access data due to lack of structural independence.

Data Management

Relationship database models

- **Advantages**:
  - Efficient, conceptually simple and easy to design.
  - This is implemented by all the powerful database management systems.
- **Disadvantages**:
  - Substantial hardware and system software overhead.
  - Prone to poor design and implementation.
  - May promote “islands of information” problems.



Data Management

Database types

- **Single-user vs. Multi-user**: A single-user can access the database at one point of time. These types of systems are optimized for a personal desktop experience, not for multiple users of the system at the same time. All the resources are always available for the user to work. Multi-user databases can be accessed by many individuals concurrently and shares the host system's resources for access.
- **Centralized vs. Distributed**: Centralized databases store data on a single CPU bound to a single certain physical/geographical location. Distributed databases, however, rely on a central DBMS which manages all its different storage devices remotely, as it is not necessary for them to be kept in the same physical and/or geographical location.
- **General vs. Specific**: General databases can be used by a wide range of applications. Specific databases are designed for use by a limited number of applications to serve a specific purpose or need.

Data Management

Examples of DBMS

	Product	Deployment	Access Controls/Permissions	Data Replication	Database Conversion	Multiple Programming Languages Supported	Relational
aws	Amazon Aurora ★★★★★ (1 review)	☁	✓	✓	✓	✓	✓
	TeamDesk ★★★★★ (38 reviews)	☁	✓	✓	✓	✓	✓
	Infanywhere ★★★★☆ (1 review)	☁	✓	✓	✓	✓	✓
caspio	Caspio ★★★★½ (60 reviews)	☁	✓	✓	✓	✓	✓
	SQL Sentry ★★★★★ (14 reviews)	🖵 ☁	✓	✓	✓	✓	✓

Data Exchange

In many cases, organizations require data stored in their DBMS to flow between their partners and stakeholders who may be using other applications.

A DBMS may not be able to function as intended, and provide any benefit to an organization, if data cannot be seamlessly exchanged in a standard format across an organization's network of partners and stakeholders.

Exchanging data in a seamless manner is critical to the successful operation of most organizations across all industries today. Data exchange drives service delivery, report generation, facilitating financial and commercial transactions, etc.

When done right, an Electronic Data Interface (EDI) is an unnoticeable part of an organization's data management program.

Data Exchange

Electronic Data Interchange (EDI)

EDI is defined as an application-to-application transfer of data between computers using industry-defined standards.

This is a comprehensive set of standards and protocols for the exchange of business transactions in a computer understandable format for organizations of all sizers across any industry – such as banking, healthcare, retail, airlines, travel, manufacturing, government, etc. EDI transactions are designed to be independent of the communications used by companies or the software technology that generates the underlying data for an EDI transfer.

This includes, but is not limited to, the following examples of data exchange:

- ✓ Transaction acknowledgment
- ✓ Financial reporting
- ✓ Inquiries
- ✓ Invoice order status
- ✓ Payments
- ✓ Pricing
- ✓ Billing
- ✓ Purchasing
- ✓ Receiving
- ✓ Scheduling
- ✓ Inventory management
- ✓ Currency exchange

Data Exchange

EDI standards

Four EDI standards exist:

- **UN/EDIFACT**: The only internationally-recognized standard, used mostly outside of North America.
- **ANSI ASC X12**: Used within North America.
- **TRADACOM**: Used by British retail companies.
- **ODETTE**: Used by European automakers.

Data Exchange

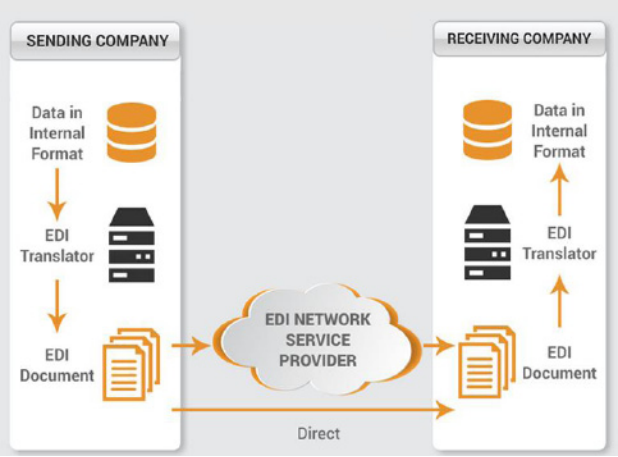
Features of EDI to consider

- **Purely Electronic**: As the name suggests, EDI removes and replaces traditional paper-based processes for any data-oriented transactions. Example: no more filling out order forms, signing it, and sending it off to purchasing.
- **Expediency**: Eliminating paper-based processes saves a lot of time and resources for an organization. It significantly reduces operational overhead.
- **Easy to gather and store data**: EDI makes it possible to help capture data being exchanged via your DBMS. This creates your data repositories that can further facilitate organizational efficiency.
- **Usability**: EDI makes it much easier to deliver data driven services to internal and external stakeholders, customers, and other end-user than traditional paper-based processes.

Data Exchange

How does EDI work?

- **Export**: Data is exported from the database.
- **Convert**: This is then converted into a standard format with an EDI translator.
- **Transfer**: An EDI service provider passes it along to the recipient.
- **Convert**: It is converted by the recipient's EDI translator into a standard format that fits their organizational need.
- **Import**: This is then imported into that system for use.



Data Exchange

Benefits of EDI

- **Improved reporting performance**: Reporting time can be cut down from days/weeks to minutes.
- **Savings**: As mentioned previously, removal of manual paper-based processes reduces overhead costs and time.
- **Accuracy and Error Reduction**: EDI makes reporting on data more accurate and less prone to errors due to its availability across organizations in a standard format.
- **Customer Service**: Having data on-hand in a standard format that is available across multiple organizations has a big impact on customer service activities. Such as a travel agency working with an airline to resolve a customer booking issue at the airport.
- **Improved supplier and client relationships**: With faster processes for invoicing, billing, purchasing etc., relationships with key suppliers and clients get a boost due to reduced lead times for fulfilling orders, completing payments, and providing services.

Data Exchange

Challenges of EDI

- **Technology Investment**: There is a cost associated with the implementation of this technology in terms of software, hardware, and administration. Furthermore, each trading partner that a company wants to use EDI with may require resources to set it up, and this can be cost-prohibitive for smaller companies or companies without technical resources. This negates any savings on overhead, especially so if the implementation is also poorly conceived and executed.
- **Change Management**: It is hard to change organizational culture. It won't happen overnight, and it will lead to a transition period with redundant processes, leading to frustration and inefficiency when dealing with partners and can strain relationships.
- **EDI Standard Selection**: There are different EDI standards and formats. For large organizations with many functions and services spread over multiple geographic locations, often it isn't enough to select just one standard. Due to the scale of operations and data required to be integrated, this leads complications of it own.
- **System Complexity**: EDI increases the complexity of an organization's IT environment substantially. Especially if multiple EDI standards are in use.

Data Exchange

Methods for EDI transfers

There are plenty of methods to conduct EDI transfers. However, there are only two methods that stand out:

- **Value-Added Network (VAN)**: This is most common method used across the world for EDI transfers. It enables companies to send a transmission which is then reviewed by the VAN and sent to the correct recipient.
- **Applicability Statement 2**: Championed by Walmart, who requires all vendors to use this, the AS2 has risen as a major transfer method given the scale of Walmart's operations and their large number of relationships with suppliers. In AS2, documents are transmitted across the internet and the security of the document is achieved by encryption and the use of digital certificates.

Some other methods for achieving EDI are listed below:

- ✓ FTP, SFTP, and FTPS
- ✓ Email
- ✓ HTTP and HTTPS
- ✓ AS1 and AS4
- ✓ OFTP and OFTP2
- ✓ Mobile EDI

Data Exchange

EDI service provider examples

Although EDI has been around for decades and enabled enterprises to communicate with their partners, many have started to shift from traditional on-site EDI service providers to cloud-based EDI offerings. Following is a list of a few on-site and cloud-based EDI service providers:

- **Mulesoft**: Their Anypoint B2B solution offers the possibility of modernizing an organization's EDI solutions with the help of APIs. Utilizing the Anypoint B2B, organizations can connect with any parties regardless of their systems and applications. Mulesoft offers rapid onboarding of new stakeholders, several popular protocols, and message logging. It is an ideal choice for those that have an extensive trading network and need to manage multiple solutions simultaneously.
- **Dell Boomi**: This Dell EDI gateway supports exchange and transformation of messages, workflow configuration, testing and monitoring of transactions. Similarly to Mulesoft, those that have a sizeable business-to-business ecosystem may prefer Boomi compared to more traditional methods.

Data Exchange

EDI service provider examples

- **SPS Commerce**: A web-based service, it comes with an easy-to-use interface that an organization can use to set up connections to transmit messages quickly. The transmission happens through VAN, AS2, or other communication protocols. The solution is suitable for multiple fulfillment models for businesses of all sizes.
- **Youredi**: This is a cloud-based integration platform, enabling connectivity across any systems and applications whether they are situated in the cloud or on-premise. Once all systems and applications are connected, the data transfer can happen in real-time (or it can be scheduled or happen in batches).
- **CLEO**: This tool is geared towards streamlining B2B and EDI processes whether the systems and applications are on-premise or in the cloud. Cleo promises to automation EDI processes to connect, transform, and route EDI and non-EDI document types between any applications. Their service can process X12, EDIFACT, and TRADACOMS standards, transform and orchestrate data in these supported formats, automate data flows across systems, and accelerate on-boarding of new stakeholders. Additionally, they provide their customers business intelligence on the message transfers.

Data Visualization

What is next, now that you have a robust DMBS and an EDI solution in place?

We need data visualization to bring our data to life. Because a visual summary of information makes it easier to identify patterns and trends than looking through thousands of rows in a database.

Data is much more valuable when it is visualized. Charts and graphs make communicating data findings easier.

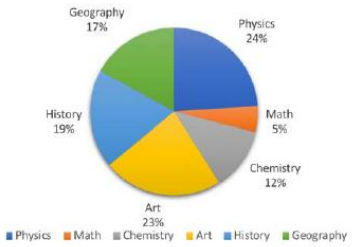
When done right, data visualization is the crucial final step of data analysis. Without it, important insights and messages can be lost.

Data Visualization

Data vs. **Information**

```
['Subject',
'count'],  ['Physics',
22], ['Math', 10],
['Chemistry', 12] ['Art',
24], ['History',
15], ['Geography', 21] ,
```

</> Raw Unorganized



Processed Structured

Data Visualization

Bringing data to life

Before putting together a visualization, you need determine what your visualization is trying to achieve.

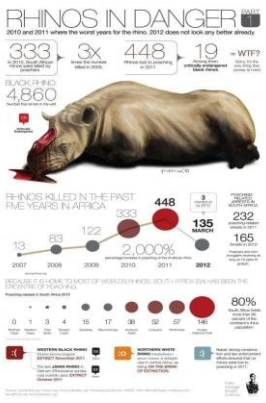
Almost all purposes of data visualization can fall within two categories:

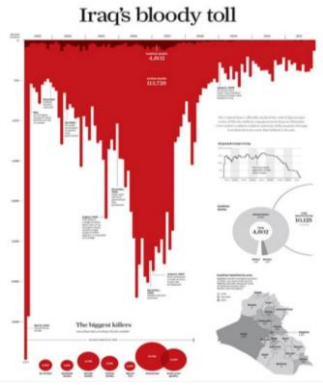
- **To drive an emotional response:**
  - Objective can be to drive an immediate call-to-action, or spark a "viral" discussion, for a social, environmental, or a political cause, etc.
- **To drive an analytical response**:
  - Does not require an immediate response, and the objective is often to spark calm dialogue, or somber reflection, which can lead to – but not necessarily have any – actionable outcome.

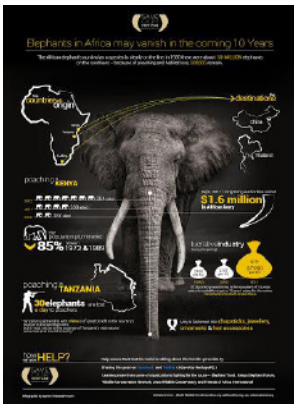
37

Data Visualization

Examples of visualizations seeking an emotional response



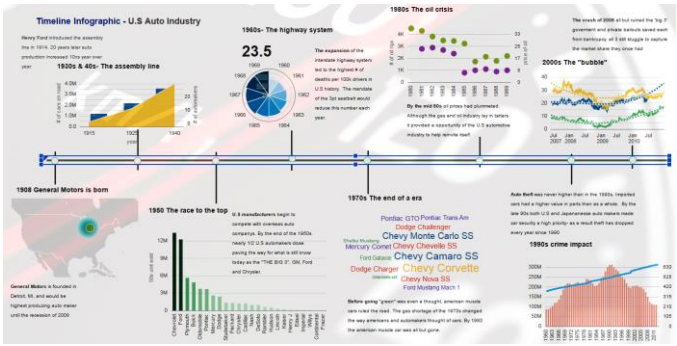


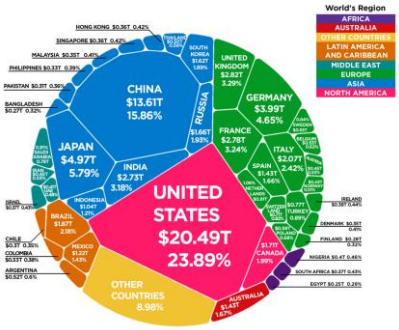


38

Data Visualization

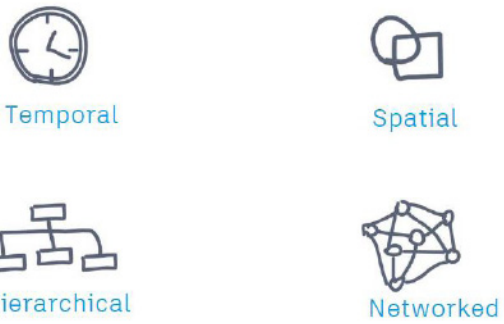
Examples of visualizations seeking an analytical response





Data Visualization

Visualizing various data and relationships



Data Visualization

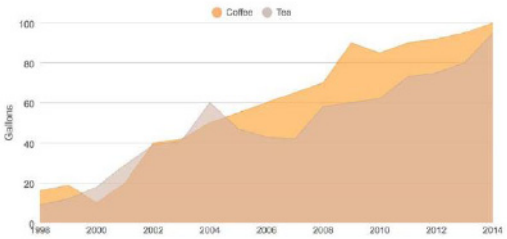
Visualizing various data and relationships: Temporal

Temporal

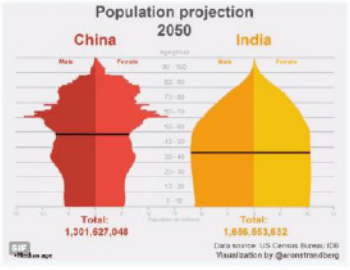
Data that represents a state in time, or what happens over a defined time-period.

The following are examples of how to visualize temporal data:

Area Charts



Animations



Data Visualization

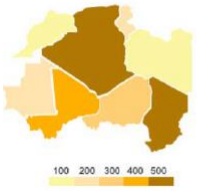
Visualizing various data and relationships: Spatial

Spatial

Data that relates to the physical world.

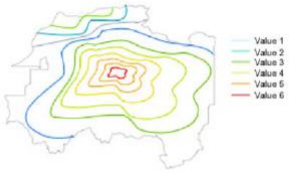
The following are examples of how to visualize spatial data:

Choropleth Map



Dot Map

Contour Map



Data Visualization

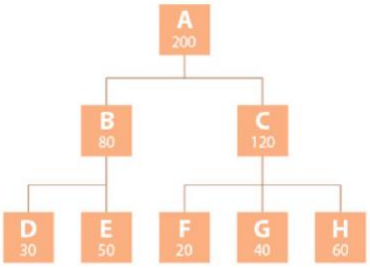
Visualizing various data and relationships: Hierarchical

Hierarchical

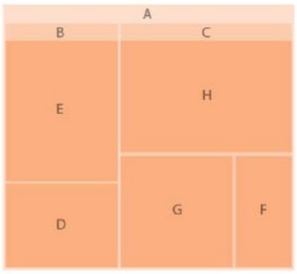
Data that relates to positions in a defined hierarchy.

The following are examples of how to visualize hierarchical data:

Tree Diagram



Tree Map



43

Data Visualization

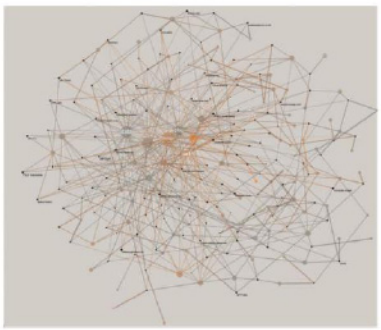
Visualizing various data and relationships: Networked

Networked

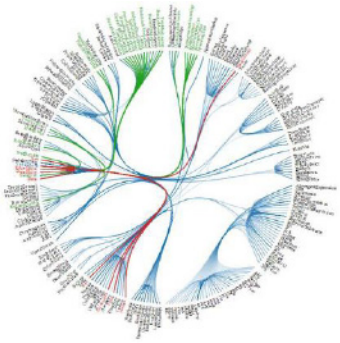
Data that connects to each other.

The following are examples of how to visualize networked data:

Network Diagram



Chord Diagram



44

Data Visualization

Stages of developing data visualizations

45

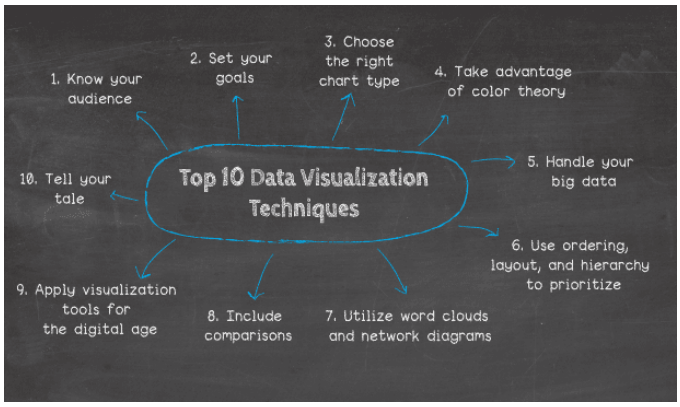
Data Visualization

Impactful visualizations

The following 10 tips should serve as a valuable guideline while working to develop any type of visualization from your data.

Following these guidelines will help you effectively achieve your objectives.

You can read more about these tips here: https://www.datapine.com/blog/data-visualization-techniques-concepts-and-methods/



46

Data Visualization

Tools used for visualization

The following tools are examples of software used for developing and producing data visualizations. This isn't an exhaustive list, but it should help you get started.

- ✓ Power BI
- ✓ Tableau
- ✓ D3
- ✓ HighCharts
- ✓ Echarts
- ✓ Leaflet
- ✓ Vega
- ✓ deck.gl
- ✓ FineReport
- ✓ Infogram
- ✓ Google Charts
- ✓ Polymaps

47

Data Visualization

Tools used for visualization

- **Power BI**: Developed by Microsoft, this is a very power set of business analysis tools that can help provide insights for the organization. It can connect hundreds of data sources, simplify data preparation and provide instant analysis. Organizations can view reports generated by Power BI on web and mobile devices.
- **Tableau**: This is a business intelligence tool for visually analyzing data. Users can create and distribute interactive and shareable dashboards, depicting trends, changes and densities of data in graphs and charts. Tableau can connect to files, relational data sources and big data sources to get and process data.
- **D3**: A JavaScript library based on data manipulation documentation. D3 combines powerful visualization components with data-driven DOM manipulation methods.
- **HighCharts**: This is a chart library written in pure JavaScript that makes it easy and convenient for users to add interactive charts to web applications. This is the most widely used chart tool on the Web, and business use requires the purchase of a commercial license.

48

Data Visualization

Tools used for visualization

- **Echarts**: An enterprise-level chart tool from the data visualization team of Baidu. It is a pure JavaScript chart library that runs smoothly on PCs and mobile devices, and it is compatible with most current browsers.
- **Leaflet**: This is a JavaScript library of interactive maps for mobile devices. It has all the mapping features that most developers need.
- **Vega**: A set of interactive graphical grammars that define the mapping rules from data to graphic, common interaction grammars, and common graphical elements. Users can freely combine Vega grammars to build a variety of charts.
- **deck.gl**: This is a visual class library based on WebGL for big data analytics. It is developed by the visualization team of Uber.

Data Visualization

Tools used for visualization

- **FineReport**: An enterprise-level web reporting tool written in pure Java, combining data visualization and data entry. It is designed based on “no-code development” concept. With FineReport, users can make complex reports and cool dashboards and build a decision-making platform with simple drag-and-drop operations.
- **Infogram**: A fully-featured drag-and-drop visualization tool that allows even non-designers to create effective visualizations of data for marketing reports, infographics, social media posts, maps, dashboards, and more.
- **Google Charts**: A free data visualization tool that is specifically for creating interactive charts for embedding online. It works with dynamic data and the outputs are based purely on HTML5 and SVG, so they work in browsers without the use of additional plugins.
- **Polymaps**: A dedicated JavaScript library for mapping. The outputs are dynamic, responsive maps in a variety of styles, from image overlays to symbol maps to density maps. It uses SVG to create the images, so designers can use CSS to customize the visuals of their maps.

Data Publication

What is the value of your data and awesome visuals if it isn't published?

Data publishing is the act of releasing data in published form for use by others. It is a practice consisting in preparing certain data or data set for public use thus to make them available to everyone to use as they wish. This practice is an integral, and important, part of the open science movement. While publishing data is not often seen in private enterprises, governments and research institutions frequently do publish their data.

Interactive graphics, often developed as a part of data visualization, allows readers to delve into a story's underlying data without burdening the reader to digest large volumes of information spread across multiple databases. These are often frequent features on websites such as those of the *New York Times* – and is one form of data publication.

When done right, data publication promotes wider data sharing and reuse, as well as provides credit to those that share their data. Depending on what is being shared, it can open new markets and create an audience.

Data Publication

Why publish data?

- **Credit and Recognition**: Publishing with data provides citable, and often peer-reviewed credit for created datasets. It allows the publication of valuable datasets that may not be well-suited for traditional research journals and grant recognition to those who may not qualify for authorship on traditional papers.
- **Discoverability and Reusability**: Published data helps maintain an indexed ecosystem of discoverable data sources, such as those in Web of Science, PubMed, PubMed Central, MEDLINE, Scopus and Google Scholar, as well as being available and discoverable on nature.com, etc. Publishing data leads to standardized data formats for reporting. Standardization makes research data easier to find and reuse, allowing data to be utilized in future experiments and research.
- **Peer-review, Awareness, and Accessibility** : Publishing with a Data Descriptor will enable the data generated to be more widely available to your peers and community. Published data is peer-reviewed and evaluates the technical quality and completeness of associated datasets with standards upheld by an Editorial Board of recognized experts from a broad range of fields.

Data Publication

Sample tools used to publish data

- **Dryad:** An open-source, research data curation and publication platform. Datasets published in Dryad receive a citation and can be versioned at any time. Dryad is integrated with hundreds of journals and is an easy way to both publish data and comply with funder and publisher mandates.
- **Figshare:** This is a multidisciplinary repository where users can make all of their research outputs available in a citable, shareable and discoverable manner. Figshare allows users to upload any file format to be made visualizable in the browser so that figures, datasets, media, papers, posters, presentations and file sets can be disseminated.
- **re3data**: A global registry of research data repositories that covers research data repositories from different academic disciplines. It presents repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions.

53

Data Publication

Online publishing

Apart from formally publishing data and associated visualizations in academic, peer-reviewed, journals, you can opt to publish online to avoid a lengthy and time-consuming peer-review process.

Here are some tools and platforms used to publish online:

- ✓ Wikipedia
- ✓ Reddit
- ✓ Social media (Facebook, Twitter, YouTube, etc.)
- ✓ Industry association websites
- ✓ News websites (Politico, Quartz, etc.)
- ✓ Topical websites (Gizmodo, TechCrunch, etc.)
- ✓ Independent blogs
- ✓ Self-published website
- ✓ Public cloud service (Google, Amazon, Microsoft)

54

Data Publication

Online publishing advantages

- **Cheap**: Most online platforms will let you publish for free. There is no shortage of content publishers. Especially if using social media outlets or pushing content to public cloud service providers. It is also relatively inexpensive to create and maintain your own website today and is used by many individuals to publish their data.
- **Easy**: You can publish your data and visualization with little technical knowledge using free-to-use open-source tools that require no more than a regular computer and an internet connection.
- **Few restrictions**: Apart from ensuring what you are about to publish does not violate local laws or your organization's expectations, you can publish anything you want. With a little technical knowledge about VPNs and encryption, even those few restrictions can be circumvented to publish online.
- **Global reach**: Content published online has the potential to reach anyone, anywhere, on earth with an internet connection the moment it is published.

Data Publication

Online publishing challenges

- **Lack of credibility**: Because anyone can – and many do – publish what they want online, there are major issues around the credibility of a vast majority of data published online without peer-review – often posing as fact. Therefore, it is harder for someone publishing data online to establish credibility with key audiences and stakeholders that might ignore data that isn't being published through a traditional peer-reviewed process.
- **Harder to control the message**: Once published online, people can take the content and make modifications to it as they please. This can be counterproductive to the original message the author was trying to convey, and they would have no control over the outcome. Especially if that data was posted on social media or other free publishing platforms online.
- **Double-edge of global reach**: Not everyone will be pleased with the story your data tells through an appealing visualization. Without a peer-review process to moderate the message, neutralize any offensive conclusions, or highlight blind spots in your analysis or insight, the author will be exposed to the harshest form of peer-review – the one done by the general public.

Questions?

Contact for further information:
Charles Sebukeera: charles.sebukeera@un.org

UN environment
United Nations
Environment Programme