Guidance for WHONET data management and security

Guidance for WHONET data management and security

Background

Individuals and organizations responsible for the collection, collation, analysis, interpretation, and reporting of public health surveillance data have a responsibility to data sources and indeed to patients and to data owners to ensure the security and confidentiality of information managed. In this document, we provide basic principles, guidance, and resources to aid surveillance program coordinators in making appropriate decisions for the management of data files from the WHONET desktop application a particular focus on: 1) data storage by an organization; and 2) data transfer between organizations.

The document should not be seen as a comprehensive review of all security procedures. Organizations will need to consider local institutional requirements and guidelines and relevant national policies and regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Rule (GDPR) of the European Union. Readers should also consider the distinct confidentiality issues between data of human, animal (especially in the food production sector), food, and environmental sources.

The confidentiality of isolate-level details is of paramount importance in isolates of human origin, while the inappropriate management and reporting of aggregate statistics has bearing when applied to healthcare organizations, food producers, and international trade. Data may either accurately or inaccurately suggest issues of poor laboratory test quality, emergence of new resistance, and outbreaks which could have important consequences on the perception of healthcare services and inappropriate interventions, and a decrease in trust in the appropriateness of data management and use practices may compromise the sustainability and impact of the surveillance program.

1. Storage of WHONET data files

WHONET data files may be created in one of three ways: 1. manual data entry into the WHONET application; 2. import of data exported from laboratory information systems, generally using BacLink; and 3. direct creation of WHONET files by a laboratory information or other system, though few vendors or data managers have implemented this third approach. The below comments apply to WHONET data files, intermediate data files exported from laboratory information systems, and any possible data outputs such as reports, tables, and charts whose contents might be considered to be confidential.

Once created, these files need to be stored in a secure location. The three most common options to consider are:

- Local data file storage on individual laptops and centrally managed organizations servers
- Internet-based storage, generally in web applications developed by the organization or a local vendor
- Cloud-based storage, generally through contractual relationship with an international cloud service provider such as Amazon Web Services, Google Drive for Business, among others

1.1 Local data file storage

Files are stored either on the user's password-protected local hard drive or preferably on a central password-protected and managed institutional server within a Local Area Network (LAN). All computers potentially hosting confidential data should be encrypted with convenient industry standard protection tools for whole-disk encryption such as BitLocker, McAfee, and VeraCrypt.

Though storage on local desktop and laptop hard drives is common, there are risks to data security, including theft or loss, and to data integrity, including data corruption through hard drive defects or computer virus infection. There are many advantages to the server approach:

- Central management by an organization's IT department of user access authorizing certain individuals within certain departments access to the WHONET application and to WHONET and other relevant data files.
- IT departments generally (but not always) have excellent strategies for manual or automated periodic (*e.g.* daily, weekly, or monthly) backups of confidential files and procedures for virus protection
- Data entered by laboratory staff is immediately available to any authorized user for data analysis and use, avoiding the inconvenience and security risks of manual mechanisms for transferring data.

With regards to laptop, desktop, and server storage, organizational policy should dictate requirements for user access covering usernames and password credentials managed by the Windows User Account Control (UAC) linked to hiring and employment policies, for example requiring passwords of certain strength and periodic updates and the removal of employee accounts after an individual has left the organization.

In many organizations, it is common to storage confidential data on simple external storage devices such as USB memory sticks and CDs. However, such practices should be limited or completely restricted because of the intrinsic risk in storing confidential data on devices that are so easily misplaced or stolen, corrupted with viruses, or confused with multiple versions of the same file stored on different devices. Indeed storage of confidential data on such mobile devices is forbidden by many organizations unless the device itself is encrypted, but encryption does not resolve the issues of loss, damage, or corruption.

- Backup strategy: Regardless of the storage methods chosen, generating routine backups is an important part of any data management strategy. Ideally the backup strategy would be automated, and the backup file should itself be encrypted and stored on a different physical disk or location in the cloud service to prevent accidental deletion or loss due to mechanical disk failure.

1.2 Internet-based data file storage

Except for the cloud storage solution described in the next section, the desktop version of WHONET does not support internet-based data file storage, as would be supported in by a full client-server platform within a web application. Consequently, we will not provide guidance for this approach within this document, and implementation can only be considered by groups with significant prior experience, expertise, and commitment.

For individuals interested in developing such a web application, there are a number of national and international surveillance initiatives which do import local WHONET files or WHONET export files into web applications developed and managed by professional IT development teams. The first three systems process WHONET files directly, while the fourth utilizes simple text files exported from WHONET.

- PROVENRA (Venezuela). Programa Venezolano de Vigilancia de la Resistencia a los Antimicrobianos, provenra.com.ve.
- Cubo Bacteriológico (El Salvador). "Bacteriological Cube".
elsalvador.bvsalud.org/vitrinas/cuales-son-estas-bacterias-resistentes
resistenciabacteriana.salud.gob.sv/login.php
- AMRCloud (Russia). amrcloud.net/en
- EARS-Net (European Union): WHONET creates isolate-level data export files that designated national representatives in Europe upload yearly to the EARS-Net web portal, atlas.ecdc.europa.eu/public/index.aspx?Dataset=27&HealthTopic=4.

1.3 Cloud-based data file storage

An increasingly attractive option for many organizations is the use of commercial cloud services for data storage as offered by Amazon Web Services, Google Drive for Business, DropBox for Business, and others. Most cloud storage providers allow for legally binding Business Associate Agreements (BAAs) and Data Sharing Agreements to be signed which provide a formal relationship between parties indicating the type of data to be stored and detailing the responsibilities associated and legal requirements. This is the approach generally taken by users of the widely used DHIS2 public health surveillance platform.

Advantages include the following:

- The options are increasingly convenient for users to implement with high levels of security.
- Storing data in the cloud provides greater protection from loss, allows for easy file sharing, version histories, and access to data from remote locations.
- A national data coordinator could provide individual folders to each satellite facility. This would allow for automatic data collection and synchronization without manual intervention.
- A satellite facility would simply copy their data to the designated folder on their computer when they are prepared to share it, and the cloud service would take care of the details.
- This model can scale to any number of users and any amount of data, while alleviating the national coordinator from having to manage this data themselves.

As a web-based solution, there are still security risks related to hackers, virus infestations, and poor user account and password management, so a decision to explore a cloud-based solution should only be considered by organizations with good practices for user account management and mitigation practices to decrease security risks.

2. Data file transfer

This section covers two topics:

- Preparation of data for file transfer, which could include data filtering, encryption, removal, and/or aggregation
- Transfer of files from on organization to another

2.1 Preparation of data for file transfer.

- Aggregate statistics: Word and Excel outputs
  Aggregated statistics which do not identify individual isolates or patients may satisfy the needs of a surveillance program. The risks are reduced when sharing this type of information in the absence of isolate- and patient-identifying data.

- Isolate data: filtered data, raw data/limited data/de-identified data
  It is often the case that isolate and patient data must be shared, and the aggregate approach is insufficient. In these cases, it may not be necessary to share the "real" or "raw" data with the other partners. As a general principle, and often required by law, it is always desirable to share the minimum amount of data necessary to accomplish the task at hand.

  For example, certain columns in the data file such as the patient's name or date of birth may be omitted from the data entirely, or a patient's identification number can be sufficiently obscured through a technique called "hashing".

  Hashing a patient's identification number would produce a new value to use in place of the patient's real identification number in such a way that is mathematically infeasible to determine the original number, but still uniquely identifies the patient in the data set. (A one-way cryptographic hash).

  Certain date values can be omitted altogether if they aren't necessary to share, while other important dates can be "shifted" to obscure the original values but preserve the approximate date. For example, WHONET provides a feature to create an export with shifted specimen collection dates, where the user specifies how many days to add or subtract from the real value.

  Another option which may be appropriate for certain fields is to recode the data. For example, a hospital may not wish to share their true laboratory code with a national system, so the true code can be recoded to a numeric value. Lab A -> 0001, Lab B -> 0002, etc.

2.2 File transfer mechanism

- Physical copies on USB, CDs, etc.

Physical copies of data on USB drives, CDs, or other portable media are inherently subject to easy loss or theft. As such, they should only be used when there are no other alternatives, and the data in which they contain must be encrypted.

- By email – secure email, unsecure mail, encrypted attachments (sharing passwords)
  Your organization may support sending this type of data via secure email. The biggest security risk in this case is on behalf of the sender. It is easy to mistype an email address while sending data for example. Apart from this risk, it will often also require manual collection and organization of the data files by the recipient, which can be very tedious depending on the number of senders.

- SecureFTP, WinSCP, Filezilla. For Linux – SSH
  There are various server/client software available to facilitate direct secure file transfers. In this scenario, the receiving site would maintain an SFTP server and provide individual credentials to each sender. Each user or site should be designated their own folder on the receiving server which cannot be accessed by any other client user. Each client user will then use an SFTP client software like those mentioned above to transmit their file to the central server securely.

- Web portal for file uploads, e.g. Viet Nam, JANIS, EARS-Net/CAESAR, GLASS
  A web application can be employed or developed by your organization to allow remote users to upload data via a secured website portal.
  This application would need to be independently vetted for security.