Chapter 9 - Data management

Kate Cuschieri[1], Laila Sara Arroyo Mühr[2]

[1] Scottish HPV Reference Lab, Dept of Lab Medicine, Royal Infirmary of Edinburgh, 51 Little France Crescent, EH16 4SA

[2] The International HPV Reference Center, Karolinska Institutet, Stockholm, Sweden

*Correspondence: Kate Cuschieri, kate.cuschieri@nhs.scot

Contents

9.1 INTRODUCTION

It is essential that laboratories establish and maintain accurate and sufficiently detailed records to:

- Ensure the right result gets to the right person at the right time.
- Enable the review, audit, and improvement of systems.
- Comply with accreditation standards (particularly where a lab provides a diagnostic service).
- Create data sets that can be used to inform and improve health care.

A core part of the work of a laboratory is to record information on specimens received and the associated test and result. However, in addition to issuing individual reports/results, laboratories are usually tasked with providing reports for internal and external stakeholders on activity and outputs more broadly. This latter aspect is fundamental irrespective of whether the laboratory has a diagnostic or epidemiological function. Clearly, accurate record-keeping and management supports in maintaining the quality of laboratory outputs. Inadequate data management can lead to delays in reporting, incorrect result dissemination and an incapacity to identify activities that might drive improvement. Additionally, the time and resource involved in resolving issues around poor data management is not trivial.

Good data management starts by having clear answers to the following questions:

- What is the meaning of the information generated?
- What information needs to be disseminated?
- Who needs the information?
- How often do laboratories need the information?
- What are the limitations of the information?

The scope of the laboratory service will influence how these questions are answered. For example, the data sets and reports associated with a laboratory that provides testing for population-based surveillance exclusively (where results will not influence a clinical pathway) will be different to those that have a diagnostic remit. Nevertheless, once these questions have been considered efforts can be directed towards which information should be recorded.

9.2 THE USE OF DATABASES AND LABORATORY INFORMATION MANAGEMENT SYSTEMS

The next stage in data management is to decide how information can be physically stored and accessed. For this laboratory information systems (LiMS) are crucial. LiMS should ideally be electronic and have capacity to store information on sample ID/characteristics, sample origin, sample-processing, test results and reporting. Allocating a laboratory ID/accession number to a sample(s) can serve as a rapid index that can facilitate future searches of the LiMS (as well as being valuable for operational tasks). It is essential that the laboratories have reception SOPs that reflect what data fields are required and how data is recorded into the system.

LiMS can also be designed to restrict levels of access and authority depending on the task, for example, only individuals with the training and experience to authorise “final” results would have access to do so. LiMS should also ideally record all access at an operator level and training manuals and instructions for use should be clear and available. A record of this training should be retained in staff-files. Additionally, regular reviews of access permissions so that any issues with inappropriate or unauthorised can be identified and managed.

Although electronic systems of “ordering tests” are preferable, paper records (e.g., sample request forms) are still received by many laboratories; while information can be transcribed from these into the electronic system, the original record can serve as a valuable document (e.g., in the case of accidental transcription errors). Again, electronic systems for scanning and retrieval reduce the burden of filing, although retaining paper records for a defined period is acceptable. All laboratories should have a documented retention policy for records.

While there are a variety of electronic systems, the choice of exactly which system is used will be dependent on:

- Scope of the laboratory and whether it offers a diagnostic service.
- Hardware availability and capacity.
- Software and licensing costs.
- Level of IT support for a particular system.
- Level of complexity reflecting on local expertise, experience, and need.

At a minimum, the system chosen should allow rapid and robust retrieval of records; it should also be straightforward to collate core information - and straightforward to train individuals to do so. Additionally, after the system has been designed it should be intuitive and not rely on operators having significant expertise in information technologies. There should be a system in place for quality checking and it should be backed up. Ideally the system should allow:

- Full audit trail of specimen “journey” from receipt to result (with information on users who performed data entry)
- Capacity for receipt of human papillomavirus (HPV) test results through a middleware (this can avoid transcription issues and save a significant amount of time)
- Capacity for rule-based reporting to enable data to move to a particular level within the system hierarchy or a particular user outside the system.

When designing any system, it is imperative that those who have a detailed understanding of the nature of the disease, the disease-control objectives, the implications of the results and the laboratory-user community are involved from the outset. Version controlled user-manual or standard operating procedures for the system use should be available and well maintained.

Once a methodology and process for information flow is established it should be followed explicitly and compliance should be monitored and audited. Procedures should also be in place that allow users to alert the laboratory to any problems or issues with reporting. Often laboratories include the proportion of accurate and timely reports the laboratory procedures as a key performance indicator (KPI). When reviewed regularly, dips in KPI performance can identify system or training issues.

9.3 ENSURING THE QUALITY OF DATA

Sometimes it is helpful to think about data in terms of "quality dimensions" - i.e., what key characteristics ensure data is of good quality. These characteristics need to reflect the scope of the service; however, there are likely universal. The Data Management Association UK (DAMA UK) recommend the following accuracy, consistency, validity, completeness, timeliness, and uniqueness. The components are explained in detail in **Table 9-1**.

Table 9-1 Data Quality dimensions

Component	Comments
Accuracy	Accuracy is achieved when the data reflects reality. This can refer to core demographic and/or clinical information as well as items like external reference numbers. Initial plausibility checks at the **start of the process** can help identify issues. For example, if the age of someone who has been tested for the diagnosis of a cervical lesion is 1 year old or 140 years old then this is clearly wrong. Real-life information can change in real time, for example, changes to national screening or vaccination policies will affect which samples are eligible for laboratory testing. It is important to keep track of data that is likely to change over time. High data accuracy allows the generation of reports that can be trusted and permits confident decision-making.
Consistency	Consistency is achieved when data values do not conflict with other values within a record or across different data sets. For example, date of birth for the same person in two different data sets should be the same. Consistent data improves the ability to link data from multiple sources. This, also increases the usefulness of the data set.

Validity	Validity is the extent to which the data meets the expected format, type, and range. For example, a sample collection month should be between one and twelve. Having valid data means that it can be used with other sources. It also helps the smooth running of automated data processes. Note that valid values do not equal accurate values.
Completeness	Data can be thought of as complete when all the data required for a particular use is present and available for use. It does not equate to 100% of data fields being complete, rather it is about determining what data is mandatory/minimum and what is optional. A laboratory should define minimum data sets. Note that complete values do not equal accurate values.
Timeliness	Timely data are those that are available when anticipated and required. A laboratory should set clear guidance on when results/reports can be expected. For labs offering a diagnostic service this is usually described as the expected "turnaround time" in days. For other non-diagnostic work-streams such as for national epidemiology, a frequency of reporting should be set (e.g., quarterly) so stakeholders can anticipate and plan analysis accordingly. Timeliness can help ensure swift management at a person/patient level. Additionally, timely reporting of aggregate, population data helps identify patterns and trends that may be actionable. While delays in reporting are unwanted a balance between accuracy and timeliness needs to be reached and should be considered fully when a laboratory sets its turnarounds for reporting.
Uniqueness	Data can be referred to as unique if it appears only once in a data set. A record can be a duplicate even when some fields are the same but some different. For example, two patient records may be associated with different referral settings or general practitioners, but if they both refer to the same patient there is duplication. Unnecessary duplication can cause inaccuracy and delays in reporting. Care should be taken when combining data sets as this can lead to a risk of duplication. Where possible data should be checked for uniqueness.

Adapted from Government Data Quality Hub: Meet the data quality dimensions - GOV.UK (www.gov.uk)[1]

9.4 RECORDING INFORMATION ON SAMPLES AND ASSOCIATED LABORATORY RESULTS

Recording receipt of specimens is detailed in chapter 4 on Collection and handling of specimens for HPV testing. Typical information to be collected and recorded during the receipt, processing and testing of samples for downstream HPV testing is shown below. Please note that the list contains *examples* of data fields, their specific nature will depend on the remit of the laboratory.

As mentioned earlier, laboratories should define minimum data sets to support the generation of accurate and timely results. User-facing information (website, user manual, standard operating procedures (SOPs)) should clearly communicate what the minimum data set for sample reception is and that failure to provide this could result in delay(s) in reporting or an inability to test.

- Demographic information (name, date of birth, participant/patient code)
- Location of collected sample (i.e. clinic, community)
- Person(s)/entity requesting the test
- Location/person for result-return
- Date sample collected
- Type of sample collected (anatomical sample type, clinician vs self-collected, preservation medium)
- External laboratory reference number(s) (if sent from another laboratory)
- Date and time sample received at laboratory
- Reason for request
- Clinical details
- Confirmation that sample is eligible for testing, if not record reason for rejection
- Type of HPV test allocated
- Date(s) of HPV test
- Lot number of HPV test and controls
- Result of HPV test
    - this should include whether the sample was valid for testing
    - if the test is a genotyping test classification of “risk” status should be clear on the report
- Date result authorised and identify of individual/team for contact

Please read Chapter 4 (Collection and handling of specimens for HPV testing) for further details on specimen recording information.

9.5 INFORMATION GOVERNANCE AND DATA SECURITY

Ensuring that health data, is securely and appropriately maintained, accessed and analysed is crucial.

Information governance is the overall approach to management of data within an organisation. It includes (but is not confined to) record management, information security and protection, compliance, data governance, data quality, risk management, privacy, data storage, audit, analytics, IT management and archiving.[2] Laboratories should have policy(s) on information governance and the detail of these will be informed by remit of the lab and the local, legal basis and framework for processing the data. Training on local information governance procedures should be mandatory and there should be a clear policy and process for the identification and management of data breaches. Periodic assessments must be conducted to validate compliance with standards so that any adjustments to protocols or systems can be promptly implemented.

Systems & processes that can support the security of data are as follows:

- Access Controls: The laboratory information management system (LiMS) is designed with granular access controls. Staff members are assigned specific roles and access permissions based on their job responsibilities. Access to patient records and test results is limited to authorized personnel, including trained laboratory technicians and authorized clinicians.

- Encryption Protocols: Electronic data, encompassing patient demographics, test results, and clinical details, undergoes encryption. This can support the security and confidentiality of transmitted data within the laboratory network and to external entities.

- Audit Trails: The LiMS maintains comprehensive audit trails that logs every access and modification made to the data. This includes details on who accessed the data, when the access occurred, and the nature of the changes made. Regular audits are conducted to monitor and review these trails, ensuring accountability and transparency in data handling.

- Physical Security Measures: Rigorous physical security measures are implemented to prevent unauthorized access to paper records and other physical media. Access to areas housing records is restricted, and surveillance systems are in place to monitor these spaces.

- Compliance with Privacy Regulations: The laboratory adheres to local and international privacy regulations. Regular training sessions are conducted to educate staff on the significance of patient confidentiality and the legal implications of data breaches.

- Incident Response Plan: In the event of a suspected or actual data breach, the laboratory has a well-defined incident response plan. This plan to include immediate containment measures, investigation procedures, notification protocols, and steps for mitigating the impact on affected individuals. The implementation of these comprehensive security measures not only ensures compliance with privacy regulations but also fosters trust with patients and stakeholders.

Striking the right balance of ensuring secure and robust governance is in place while ensuring that real-world data is used gainfully to improve health is sometimes challenging, particularly in the face of evolving legislation.[3] Communication with local governance experts can support the development of procedures and safeguards. Data management frameworks can be helpful in this regard also, such as the “Data management and competency framework” published by the World Health Organization (WHO) (Western Pacific Region) which is designed to serve as a “practical

tool to provide both a structure and methodology to enable health information workforce employees (who have cause and need to engage with health data), their line managers and human resource managers to define the competencies required for the identified data management roles within their organizations".[4]

Data Validation and Quality Control:

The laboratory should employ a stringent data validation and quality control process to ensure the accuracy and reliability of all information generated. Regular calibration of equipment, validation checks at various stages of the testing process, and continuous monitoring of data quality are integral components of our quality assurance framework. This proactive approach enhances the trustworthiness of the data, contributing to the overall reliability of our testing outcomes. Additionally, maintaining comprehensive documentation of all data management processes is paramount to laboratory´s commitment to excellence.[5] This includes keeping accurate meticulous of system configurations, updates, and user activities.

Data Sharing and Collaboration:

Notwithstanding governance considerations, collaboration and data exchange with external laboratories and agencies is integral to the laboratory's mission; sharing data can allow impactful epidemiological studies and foster collaborative, international research initiatives.[6] This commitment not only enriches the collective understanding of HPV but also cultivates transparency and knowledge sharing within the broader scientific community. When such exercises are planned agreed data transmission/exchange formats and data transfer agreements help ensure integrity and security of the information.

Disaster Recovery and Business Continuity:

In anticipation of unforeseen events, the laboratory should have developed and implemented a comprehensive disaster recovery and business continuity plan that includes provision for data capture and retrieval.[7,8] Regular data backups, strategic off-site storage solutions, and well-defined protocols for rapid system recovery collectively ensure minimal disruption in the face of unforeseen events.

9.6 REFERENCES

1. https://www.gov.uk/government/news/meet-the-data-quality-dimensions

2. Information Commissioner´s Office (IC). UK GDPR guidance and resources, last accessed October 1, 2024)

3. Jones MC, Stone T, Mason SM, Eames A, Franklin M. Navigating data governance associated with real-world data for public benefit: an overview in the UK and future considerations. BMJ Open 2023;13:e069925.

4. World Health Organization. Data management competency framework, last accessed October 1, 2024).

5. Ammenwerth, E, de Keizer, N. An inventory of evaluation studies of information technology in health care—Trends in evaluation research 1982–2002. Methods of Information in Medicine 2019;48:218–226.

6. Adler-Milstein J, DesRoches CM, Kralovec P. Electronic health record adoption in US hospitals: progress continues, but challenges persist. Health Affairs 2015;34:2174-2180.

7. Chute, CG, Beck SA, Fisk TB, Mohr DN. The Enterprise Data Trust at Mayo Clinic: A semantically integrated warehouse of biomedical data. Journal of the American Medical Informatics Association 1996;3:422–433.

8. Syn SY, Kim S. Characterizing the research data management practices of NIH biomedical researchers indicates the need for better support at laboratory level. Health Info Libr J 2022;39:347-356.