

Rohdaten / „Raw Data“ und Elektronische Daten::

Wie sind Rohdaten definiert?
Welche Daten muss man aufzeichnen und speichern?

Ein alt-bekanntes Thema ist immer wieder die Auseinandersetzung mit Rohdaten und den Anforderungen bezüglich der Validierung bzw. des Daten-Managements (Integrität). Eindeutige Definitionen und Hinweise sind selten, der Interpretationsspielraum weit gestreut, und die Ansichten unterschiedlich.

Nach EU GMP Leitfaden heißt es im Kapitel 4 – Dokumentation – folgend:

1. Records include the raw data which is used to generate other records.
2. For electronic records regulated users should define which data are to be used as raw data.
3. At least, all data on which quality decisions are based should be defined as raw data.

Als Referenz zu den drei Anforderungen gilt die deutsche Übersetzung durch das Bundesministeriums für Gesundheit:

http://www.bmg.bund.de/fileadmin/dateien/Downloads/Gesetze_und_Verordnungen/Bekanntmachungen/A/AMWHV/Anlage_1_zur_Bekanntmachung_-_Kapitel_4_Dokumentation.pdf

(Kommentar: Leider ist die Übersetzung nicht besonders gelungen, daher bleiben wir beim englischen Original)

Detailbetrachtung der 3 Anforderungen:

1. Aufzeichnungen, welche einen Nachweis über Maßnahmen zur Sicherstellung der behördlichen Anforderungen beinhalten ("provide evidence of various actions taken to demonstrate compliance"), können Rohdaten enthalten oder werden verwendet, um andere Aufzeichnungen zu erzeugen. Als Beispiel wird der *Batch Processing Record* (Chargenverarbeitungsprotokoll) genannt.
 - In aller Regel ist dieses Protokoll tatsächlich eine Sammlung von Daten aus Produktions-, Monitoring- und Labor-Systemen usw.
 - Die Daten können messtechnisch direkte oder indirekte Messwerte enthalten, automatisiert oder händisch übertragen werden, als Anhang zum Protokoll geführt werden, in anderen Systemen zwischengespeichert oder berechnet werden usw..
2. Definition – Benutzer, die definiert haben, dass das Masterdokument (siehe Kapitel 4.1) als elektronisches Dokumententyp existiert, müssen definieren, welche Daten tatsächlich als Rohdaten gültig sind.
 - Die meisten Daten werden i.R heutzutage elektronisch aufgezeichnet, verarbeitet und archiviert.

- Diese Anforderung setzt natürlich voraus, dass man alle seine GMP-relevanten Datensätze (Bewegungs-, Stamm-, Metadaten) kennt, bzw. analysiert hat und welche Entscheidungen auf diesen Daten getroffen werden.
 - Beispiel: Das o.g. Chargenprotokoll wird aus Daten aus 5 unterschiedlichen (validen) Systemen (bzw. Aufzeichnungen = Datenquellen) erzeugt.
3. Qualitätsentscheidungen: Daten müssen als Rohdaten definiert werden, wenn daraus Qualitätsentscheidungen getroffen werden.
- Solche Entscheidungen könnten typischerweise eine Produktfreigabe oder Sperrung durch die Sachkundige Person o.ä. sein. Dabei wird der Messwert gegen eine Sollvorgabe (mit definierten Grenzen, ggf. aus der Zulassung) verglichen. Zum Beispiel, es wird ein Sollwert von 10,0 definiert, obere Grenze 10,1 und untere Grenze von 9,9. Der zulässige Bereich wäre in diesem Beispiel definiert auf 0,3 (hier ohne Einheiten).
 - Die Qualitätsentscheidung hätte einen erheblichen Einfluss auf die Produktqualität und auf das Patientenrisiko.
 - Wäre der berechnete bzw. übertragene Messwert falsch (nicht valide), könnte eine Entscheidung positiv-negativ oder negativ-negativ sein:
 - a. Tatsächlicher Wert liegt im Gültigkeitsbereich, (falsch) ermittelter Wert zeigt eine Grenzwertverletzung – Charge wird gesperrt – kein Patientenrisiko, aber Businessrisiko vorhanden.
 - b. Tatsächlicher Wert liegt nicht im Gültigkeitsbereich, (falsch) ermittelter Wert zeigt **keine** Grenzwertverletzung – Charge wird freigegeben – hohes Patientenrisiko und Businessrisiko vorhanden.
 - Eine tiefere Betrachtung würde uns zur Statistik und zur Wahrscheinlichkeitsrechnung bringen, bzw. die Eintrittswahrscheinlichkeit bzw. Nichteintrittswahrscheinlichkeit (siehe z.B. Rowohlt Taschenbuch: Der Hund, der Eier legt).

Messtechnische Betrachtung:

Nehmen wir nun an, dass der übermittelte Wert nach einer validierten Methode an einem qualifizierten Messinstrument (z.B. „driftfrei“, kalibriert, Wartungsvertrag) bestimmt wurde. Die einzelnen Messwerte (z.B. 100 Werte) wurden physikalisch / analytisch direkt mit dem Messinstrument oder einer Messeinrichtung bestimmt und diese können mathematisch über eine Formel (exaktes mathematisches Model, ohne Variablen – z.B. thermische Faktoren) zu einem indirekten Messwert verarbeitet werden. Man unterscheidet zwischen direkter und indirekter Messungen und der Messhäufigkeit, z.B. zur Reduktion von Messfehlern.

Die Messgröße muss nicht unmittelbar Gegenstand der Messung sein. Sie kann auch über physikalisch bekannte oder festgelegte mathematische Beziehungen aus Größen bestimmt werden. In aller Regel werden physikalische Messgrößen bestimmt (SI Einheiten) und es werden digitale Messmethoden angewendet.

Beispiel – pH Meter - Glaselektroden: Eine Halbzellenreaktion an der Glasmembran bildet ein elektrisches Potenzial aus, welches in direkter Abhängigkeit zur H⁺-Ionen-Konzentration steht. Mit Hilfe eines Voltmeters (Spannungsmessung) kann nun eine Potentialdifferenz zwischen Glaselektrode und Bezugselektrode gemessen werden. Man erhält einen Messwert in Volt. Das Volt ist eine abgeleitete Einheit, das heißt, sie kann mit Hilfe der sogenannten Basiseinheiten kg, m, s und A ausgedrückt werden. Ein Rechenchip errechnet über die Nernstgleichung die H⁺-Ionen-Konzentration und man erhält einen Wert in pH-Einheit. Die Nernstgleichung beinhaltet die Faraday-Konstante, die ebenso eine relative Unsicherheit (vernachlässigbar gering) enthält. Die ermittelten Spannungswerte in Volt sind laut Definition keine Rohdaten!

Eine exakte und richtige Messung beinhaltet in Sinne einer guten Ingenieurspraxis eine dokumentierte und eindeutige Festlegung des Messprinzips, der Messmethode, des Messobjekts (hier: Probe), dokumentierte Messverfahren und Bewertung der Einflussgrößen bzw. eine Reduktion der Messfehler. Der Messaufbau besteht aus einem Messgerät, Messeinrichtung oder einem Messwerk; bei Kombinationen oder Reihen spricht man auch über eine Messkette. Prinzipiell kann die Messkette datenorientiert betrachtet werden, d.h. vom Messobjekt bis hin zum Messergebnis. Der anderen eingabenorientierte Prozessfluss könnte durch die (geschulten) Person oder Personen repräsentiert sein, welche das Messobjekt zur Verfügung stellen (Probenzug), die Probe(n) vorbereiten, das Messgerät verwalten (reinigen etc.), nach der Messmethode messen oder einzelne Messschritte einleiten oder umsetzen (Einflussgröße durch manuellen Eingriff).

Zum Messen gehören (z.B. nach DIN 1319):

1. Eindeutige Definition der Messaufgabe (Messproblem) und der Messgröße: Die Aufgabe, das Messobjekt und die physikalische Messgröße sind (in einer SOP) festzulegen.
2. Festlegung der Maßeinheit für das Ergebnis: Die Einheit und ihr Einheitenzeichen sind in der Regel gemäß dem SI festgelegt, wählbar sind Vorsätze für Zehnerpotenzen (ebenfalls gemäß DIN 1301-1). Es gibt auch dimensionslose Größen, z. B. Winkel, Brechzahl, Anzahl, deren Werte ohne Einheit angegeben werden.
3. Zusammenstellen der Randbedingungen: Als Randbedingungen sind z.B. Eigenschaften des Messobjektes und der Umgebung zu beachten. Die Umgebung muss daher qualifiziert sein (Raum, Einrichtungen).
4. Wahl einer Messeinrichtung oder eines Messgerätes: Ausgehend vom Messprinzip und der Messmethode wird ein Messverfahren entwickelt, das in einer Messeinrichtung verwirklicht wird. Vielfach steht bereits ein

- auf dem Markt verfügbares Messgerät für die Messaufgabe zur Verfügung.
5. Kalibrieren von Messeinrichtung/Messgerät: DIN EN ISO 9000 ff fordert die Rückführbarkeit aller Messungen auf nationale Normale. Dieses wird durch das Verfahren der Messmittelüberwachung gesichert. Dazu soll ein Messgerät in regelmäßigen Abständen kalibriert (eingemessen) werden. Falls der Messwert nicht mit dem Wert der Messgröße innerhalb vorgegebener Fehlergrenzen übereinstimmt, ist das Gerät vorerst neu zu justieren (einzustellen).
 6. Festlegung des Messablaufs: zeitliche bzw. örtliche Abfolge der Messungen: z. B. Reihenfolge von Einzelmessungen, Wiederholungen, Messreihe unter geänderten Bedingungen; räumliche Verteilung der Messpunkte (Messstellen), Messprofile, regelmäßiges Raster usw.
 7. Durchführen der Messung und Ermitteln des Messergebnisses: Es können eine Messung oder auch mehrere unter denselben Bedingungen gewonnene Messungen derselben Größe (Vergleichs-/Wiederholmessungen) durchgeführt werden. Dann sind Mittelwert und Standardabweichung zu berechnen. Ferner können Messungen verschiedener Größen erforderlich sein, aus denen der Messwert der gesuchten Größe nach festgelegten mathematischen Beziehungen zu berechnen ist.
 8. Berücksichtigung der Auswirkungen von Einflussgrößen: Korrektur von systematischen Messabweichungen. Je nach Umständen gehört hierzu auch eine Reduktion, d. h. eine Korrektur auf einheitliche Bedingungen.
 9. Ermitteln des vollständigen Messergebnisses: Ein vollständiges Messergebnis besteht aus dem Messwert (gegebenenfalls Mittelwert aus einer oder mehreren Messreihen oder dem berechneten Wert aufgrund anderer Messungen), ergänzt durch quantitative Aussagen zur Messunsicherheit; als Letztere dienen Fehlergrenzen, die mit den Mitteln der Fehlerrechnung oder der Fehlerfortpflanzung zu bestimmen sind.

Nun sind Messungen im Allgemeinen mit einem Messfehler behaftet (systematischer, zufälliger Fehler, Fehlergrenze des Messgeräts, Quantisierungsfehler). Übrigens können systematische Fehler auch durch wiederholende Messungen nicht aufgehoben werden.

Die vom Messgerät erfasste Größe kann grundsätzlich nicht richtig angezeigt werden, bzw. je nach mathematischem Model aus mehreren bestimmen Größen aus unterschiedlichen Messungen (Geräte), entstehen Messunsicherheiten und eine Fehlerfortpflanzung. Dies kann dazu führen, dass die Gesamtfehlergrenze (die einzelnen, relativen Fehlergrenzen addieren sich in aller Regel) unter Umständen größer sein könnte, wie die vorgegebenen Grenzwerte des definierten Sollwerts für die Qualitätsentscheidung. Dies ist zwar in aller Regel eher unwahrscheinlich, sollte aber durchaus betrachtet werden.

Was sind GMP Rohdaten:

Die sogenannten Rohdaten bestehen demzufolge aus zwei Datentypen:

- Dem Messwert (Ergebnis der Messung) selbst, z.B. „ph-Wert = 8,9“
- Den Metadaten des Messwerts, das heißt die Angabe der Verfahren, Methode, Einflussgrößen, Probenzug, usw. (Rückführbarkeit / Transparenz)

Wird der Messwert nun in das Chargenprotokoll oder in eine Laboraufzeichnung übernommen, dient diese als Grundlage für die zu treffende Qualitätsentscheidung. Der Messwert als alleinige Angabe ist nicht ausreichend – häufig wird das Messgerät mit angegeben. Vollständig wären diese aber mit den Angaben (oder entsprechende eindeutige Referenzen) auf die Messmethode (z.B. Ph. Eur / USP), verwendete (valide, dokumentierte) Verfahren und Personen.

Praktische Auslegung: Der Umfang der Laboraufzeichnungen sollten angemessen sein – zumindest Referenzen sollten angegeben werden (Rückverfolgbarkeit sicherstellen). Ziel und Zweck eines Qualitätssystems ist die Sicherstellung der Qualitätselemente, wie z.B. Messverfahren, Methodvalidierung, Personalschulung – dies entspricht dem prinzipiellen Teil der Metadaten eines Messwerts. Dabei dient ein prospektives und retrospektives Qualitätsrisikomanagement der Sicherstellung eines guten Messkonzepts.

Bleiben wir nun bei diesem Beispiel und betrachten exemplarisch die gesamte Messkette (inklusive Route des Messobjekts) des gesamten Prozesses:

1. Probenanweisung und Probenzug
2. Lagerung und Transport der Probe (= Messobjekt)
3. Vorbereitung der Probe /Geräte zur Analytik nach Messmethode / Verfahren
4. System – d.h. Messgeräte – ist qualifiziert, kalibriert, etc. und vorhanden
5. Messgerät (Genauigkeit bzw. Fehlergrenze): Der Einzelwert wird mit einem Messgerät mit einem relativen Fehler von 1 % bestimmt.
6. Es werden 100 Messwerte erzeugt (Annahme: Wiederholende Messung = für gleiche Probe). Die mathematische Formel besteht aus dem Mittelwert dieser Messung (indirekter Messwert) und einem bestimmten Einzelwert aus einer anderen – direkten - Messung (z.B. Mittelwert wird mit Einzelwert multipliziert)
7. Die Berechnung erfolgt über ein (validiertes) MS Excel Sheet (Tabellenkalkulationsprogramm):
 - a. Die 100 Messwerte werden händisch (copy & paste) übertragen.
 - b. Der Einzelmesswert wird ebenso manuell, händisch (von einem Papier-Ausdruck) übertragen.

8. Das Ergebnis der Berechnung wird wieder händisch z.B. in einer Laborsystem (z.B. LIMS) übertragen.
9. Nach Freigabe aller Laborergebnisse wird das Analyseergebnis im LIMS freigegeben und über eine IT Schnittstelle an eine Chargenprotokollierungssystem (ERP) automatisiert gesendet.

Das Auftreten einer fehlerhaften Übertragung (systematischer Fehler) ist mit hoher Wahrscheinlichkeit deutlich höher oder tritt öfters ein, als die Betrachtung von Messunsicherheiten (bzw. Fehlerfortpflanzungs-Regeln). Sprich, dass beispielweise die falsche Probe gezogen wird, Ablesfehler, Daten falsch übertragen werden oder versehentlich nicht in die richtigen Spalten vollständig kopiert werden usw.

Zurückkommend auf die Rohdaten, die ja in diesem Beispiel die 101 Einzelwerte sind:

- ➔ Schritt 1 bis 5 sind Metadaten der Messung (Teil der Rohdaten)
- ➔ Schritt 6: „erste“ Rohdaten (Messwerte)
- ➔ Schritt 7: Rohdaten und Formel (=Methode; Metadaten)
- ➔ Schritt 7: Excel Sheet (vereinfacht, ugs. für Tabellenkalkulationsprogramm) = Metadaten (Methode) führt zum berechneten Messergebnis (entspricht der Entscheidungsbasis)
- ➔ Schritt 8: Methode - hier hohe Wahrscheinlichkeit eines systematischen Fehlers beim Übertrag
- ➔ Schritt 9: Datenformat geändert?

Ist das richtig so?

„At least, all data on which quality decisions are based should be defined as raw data“: Ein definierter Entscheidungsträger macht eine Qualitätsentscheidung auf Grundlage des Messergebnis in Schritt 8 oder 9, das er hier höchstwahrscheinlich im LIMS System sehen würde (z.B. Sachkundige Person bei der Chargenfreigabe).

Häufig wird nun diskutiert, ob die Rohdaten existent sind auf Ebene:

- ➔ Messgeräte (Schritt 6) – und leiten daraus ab, dass die Daten dort gespeichert und valide vorliegen müssen (über den gesamten Aufbewahrungszeitraum) – dies wäre exakt für dieses Beispiel richtig.
- ➔ Excel Sheet (Schritt 7) – abgeleitet daraus, dass das Excel Sheet validiert sein muss – dies wäre exakt für dieses Beispiel richtig.
- ➔ Das LIMS und ERP System müssen validiert sein.

In aller Regel würde ein Entscheidungsträger nicht die Rohdaten im Excel Sheet verifizieren oder einsehen müssen. Er verlässt sich auf die korrekte Bestimmung der Daten und dass nötigenfalls die Daten nachvollziehbar wären und dass die Messung korrekt abgelaufen ist. Das ist soweit auch korrekt und konform und bedeutet, dass die Daten auf dem Messgerät tatsächlich als Rohdaten zu definieren sind und validiert sein müssen.

In diesem Beispiel haben wir einige Übergänge (bzgl. Daten):

- Messgeräte ins Excel Sheet (manuell, händisch, über copy&paste)
- Ergebnis Excel Sheet ins LIMS (manuell, händisch, über Ausdruck des Excel Sheets d.h. Papier ins LIMS System)
- LIMS ins ERP System

Bei diesem Beispiel unterscheiden wir einen manuellen Eintrag vom Papier aus oder wenn Daten in einem Fremdsystem (z.B. Excel Export der Datenergebnisse möglich) über die Kopierfunktion (ugs. copy & paste) möglich wäre. Dieser Faktor wird deswegen verwendet, da er nicht unerheblich ist und meist vergessen wird.

Nun könnte man argumentieren, dass der gesamte Prozess doch ein wenig umständlich ist – korrekt, aber durchaus in der Praxis so anzutreffen ist. In vielen Labors wirkt(e) die Verwendung von Excel Sheet durchaus attraktiv – diese sind schnell erstellt, kostengünstig und anpassbar.

Was passiert eigentlich, wenn das Messgerät nun einen sogenannten Ringspeicher hat (z.B. für 1000 Messwerte). Sollte man dann versuchen, die Daten irgendwie auf eine z.B. SQL Datenbank zu spiegeln – oder ähnliches? Aufwand? Nachhaltig?

Was wäre eigentlich, wenn man hier den Prozess gesamthaft anpassen und jede Art von manuellen (Daten-)Übergänge eliminieren würde. Nehmen wir mal an, dass die Messwerte vom Messgerät direkt in ein „intelligentes“ LIMS (oder Data Warehouse) übertragen werden, dort berechnet werden und wie gehabt automatisiert in die ERP Ebene gesendet werden. Wäre die Schnittstelle zwischen Messgerät und LIMS qualifiziert, würden eigentlich die Rohdaten nicht mehr im Messgerät selbst verbleiben müssen (könnten also dort gelöscht werden). Und würde man die Übertragung vom LIMS ins ERP sogar noch mit den Rohdaten erweitern, wären dies letztendlich im finalen System ebenso verfügbar (und damit auch alleinig gültig). Dies hätte die Vorteile, dass:

1. keine redundante Daten im Unternehmen zu verwalten wären
2. die zufälligen Fehler der manuellen Datenübertragung gegen Null wären
3. das Messergebnis relativ einfach im finalen „Entscheidungs“-System betrachtet werden können, inklusive der Rohdaten (=Datengrundlagen entspricht den Entscheidungsgrundlagen)
4. Recherchen (z.B. OOS) oder Rückfragen bei Inspektionen schneller beantwortet werden können

Dies sind alles Punkte, die für eine umfangreiche „Datenautomatisierung“ sprechen würden. Und dann wären dort noch die Kosten zu bewerten und werden aktiv diskutiert – es scheint im ersten Augenblick eine hohe Investition zu sein. Dies muss man im Detail analysieren – das Ergebnis wird erheblich Kosteneinsparungen zeigen.

Aber wie kann so etwas implementiert werden – nicht im technischen Sinne – hierfür gibt es Lösungen, sondern im gesamthafften Ansatz:

1. Der häufigste Fehler besteht darin, dass die Bewertung der Daten, ob diese GMP relevant sind oder nicht, nicht durch die Entscheidungsträger selbst getroffen werden (sondern z.B. durch den Geräteverantwortlichen, Qualifizierungsteam, etc. auf einzelner Geräte/System Betrachtungsbasis):
 - a. Involvieren Sie die Entscheidungsträger, die definieren müssen, welche Daten für Sie entscheidungsrelevant (d.h. damit GMP relevant) sind
 - b. Betrachten Sie Daten- bzw. Messketten von der Erzeugung bis zur finalen Darstellung – an diesen sind in aller Regel mehrere Systeme beteiligt.
 - c. Analysieren Sie die Daten-Übergänge zwischen den Systemen (Papierübertrag, copy&paste, Schnittstelle); ggf. liegen die Rohdaten auch bei einem externen Dienstleister (Prüfung von Qualitätsvereinbarungen nötig).
2. Wichtig ist nun, ob der Entscheidungsträger seine Entscheidung auf Papier dokumentiert oder elektronisch darstellt (z.B. über eine elektronische Signatur). Konsequenterweise müssten bei einer elektronischen Freigabe auch die Daten elektronisch verfügbar sein (auch die Rohdaten). Aber nur konsequenterweise, aus logischen und betriebswirtschaftlichen Gründen, aus regulatorischen Gründen könnten diese auch auf Papierform vorliegen (Achtung: Overkill Gefahr).
3. Thema richtig darstellen: Es ist eigentlich keine Frage der Rohdaten-Definition – sondern das Thema heißt eigentlich Datenmanagement. Der „Projektauftrag“ eigentlich eine Datenanalyse (plus Aufzeichnungen).
4. Nötigenfalls kann man auch im ersten Schritt festlegen, welche Entscheidungen eigentlich getroffen werden. Eine einfache Auflistung, welche qualitätsrelevante Entscheidungen im Unternehmen überhaupt getroffen werden, schafft oft viel Transparenz. Diese Qualitätsentscheidungen beginnen beim Personalmanagement, Lieferantenfreigabe, Ausgangsstoffe, Dienstleistungen, Wartungsverträge, Änderungen, Abweichungen, Materialfreigaben, etc.

In der Inspektionssituation könnte man ja gefragt werden, wenn man eine Entscheidung getroffen hat (Chargenfreigabe), auf welcher Grundlage diese getroffen wurde. Die Grundlage sind die relevanten Messergebnisse auf dem z.B. Chargenprotokoll, wenn diese den Sollvorgaben entsprechen. Die Messergebnisse sind ein Resultat aus Messwerten (Rohdaten-Aufzeichnungen) – „Records include the raw data which is used to generate other records“. Natürlich müssen diese vorhanden, valide und nachweisbar sein. Die Messkette mit allen Zwischenschritten der Aufzeichnungen müssen daher mindestens bekannt sein.

GMP relevante Entscheidungen basieren auf möglichst präzisen Methoden (nach Stand von Wissenschaft und Technik) – nicht auf einem Bauchgefühl. Daher ist es unabdingbar, dass man seine Rohdaten identifiziert, bestimmt, festlegt und entsprechend deren Integrität und Validität effizient sicherstellt.