Data Integrity : vers la mise en place d'un Data Management

Par Lionel PELLETIER - AKTEHOM
lionel.pelletier@aktehom.com

La publication en mars 2018 par la MHRA de son nouveau guide sur le sujet est l'occasion de rappeler l'objectif principal du *Data Integrity* : avoir confiance dans la qualité et l'intégrité des données générées, et être capable de reconstruire les activités. Cet article propose une interprétation des principales exigences et un éclairage sur leur mise en application. La maitrise du *Data Integrity* est devenue une préoccupation majeure des autorités de santé, et pas moins de 6 guidelines[1] ont été publiés depuis 2016 sur le sujet par différentes instances réglementaires. Ces lignes directrices, dont quelques-unes sont toujours sous forme de *draft*, montrent une certaine convergence dans les attendus. La détection par le passé de nombreuses violations des bonnes pratiques de fabrication en termes de maîtrise des données, ainsi que de fraudes caractérisées, ont conduit les inspecteurs à clarifier leur position sur le sujet, et à s'y former. Le nombre de *Warning Letters*, de *Non Compliance Reports* et autres injonctions sur ce point a ainsi considérablement augmenté ces dernières années.

Data Integrity

L'intégrité des données se définit communément comme étant la mesure dans laquelle les données critiques restent complètes, cohérentes et exactes (*complete, consistent and accurate*) tout au long de leur cycle de vie. Si cette notion n'est pas récente - l'IEEE la définissait déjà selon la même formulation dans les années 1990, et on trouve des *Warning Letters* émises par la FDA dans les mêmes termes au début des années 2000 - sa maitrise passe néanmoins par des dispositions nouvelles, du fait de la digitalisation grandissante des processus, de la mondialisation des activités et de la multiplication des systèmes informatisés par lesquels transitent les données.

Reliable Decision-Making

Pour qu'une décision soit robuste, l'intégrité des données qui la supportent doit être garantie. Quelle confiance peut-on accorder à une décision pharmaceutique prise sur la base de données erronées ?

Il est alors fondamental que le système qualité d'une organisation puisse permettre l'identification et la mise sous contrôle des points de vulnérabilité des données critiques, que ces données soient électroniques ou enregistrées sur un support papier.

ALCOA

Les autorités s'accordent à considérer que le *Data Integrity* est couvert lorsque les données respectent les exigences *ALCOA (Attributable, Legible, Contemporaneous, Original et Accurate)*. Cet acronyme résume les caractéristiques qui démontrent que les évènements subis par les données ont été correctement documentés

Attributs	Attentes	Exemples de moyens de contrôle
A - *Attributable* (Attribuable)	Il est possible d'identifier l'individu ou le système qui a généré ou modifié la donnée, ou traité l'activité	- Enregistrement d'un identifiant informatique - Signature manuscrite
L - *Legible* (Lisible)	Il est possible de lire ou interpréter la donnée après son enregistrement	- Bonnes pratiques documentaires - Contrôles sur les modifications des données - Audit Trail - Logbook - Archivage
C - *Contemporaneous* (Contemporaine)	La donnée est enregistrée au moment où elle est générée, ou au plus près de l'événement générateur	- Enregistrement des données sur le support définitif - HoroDatage - Synchronisation horaire - Limitation des mémoires temporaires
O - *Original* (Originale)	La donnée est conservée dans son état original ou en 'copie certifiée'. Elle conserve le contenu et le sens	- Conservation des données et métadonnées dans leur statut initial (ex : données dynamiques) - Systèmes de sauvegarde
A - *Accurate* (Exacte)	La donnée reflète exactement l'activité ou la mesure effectuée La donnée est vérifiée si nécessaire, les modifications sont expliquées	- Calibration des instruments de mesure - Revue des données et des métadonnées - Contrôles sur les données

Tableau 1. Attendus ALCOA

et que celles-ci peuvent être utilisées pour supporter une décision. On trouve également dans certaines publications la notion de *ALCOA+*, qui précise un peu plus les attendus en ajoutant les termes *Complete, Consistent, Enduring* (durable) et *Available* (disponible).

Le *Data Integrity* est ainsi un requis essentiel du système qualité pharmaceutique, et l'attendu principal est que les principes *ALCOA* soient respectés pour toutes les activités régies par les GxP (*voir tableau 1*).

Les fabricants, distributeurs ou exploitants doivent donc être en mesure de détecter les failles, dans l'organisation ou les systèmes, qui peuvent mener à une altération des données, et entrainer une décision erronée. Ces détections prennent en compte **les altérations volontaires** ou involontaires des données, et s'appliquent aux **supports électroniques et papier**.

Les différentes lignes directrices émises sur le *Data Integrity* ont introduit une terminologie qu'il est important d'assimiler, bien que certaines définitions puissent varier légèrement d'un texte à l'autre. L'interprétation correcte de certains termes conduit à une meilleure maitrise de l'intégrité des données.

Métadonnées

Les *métadonnées* sont les informations liées à une donnée, qui apportent un élément de contexte et permettent de mieux en comprendre le sens. L'intégrité de ces métadonnées doit être assurée. Par exemple, le lien entre une donnée et son horodatage (la date et heure de son acquisition) est toujours maintenu, et cet horodatage respecte lui-même les attendus *ALCOA*.

Enregistrements statiques et dynamiques

Les instances réglementaires, et en particulier la FDA, distinguent les enregistrements statiques des enregistrements dynamiques.
Sur un enregistrement statique, les données sont figées, et n'ont pas de raison d'être modifiées. C'est le cas, bien souvent, d'un enregistrement papier ou d'une image électronique. Un ticket de pesée, par exemple, ne nécessite généralement pas de traitement par un utilisateur pour être exploitable et contient a priori toutes les informations nécessaires à l'interprétation du résultat.

Ce n'est pas le cas d'un format d'enregistrement dynamique qui, lui, autorise une interaction avec un utilisateur en vue de son exploitation. Sur un chromatogramme, par exemple, les paramètres d'intégration peuvent être modifiés, un pic apparaissant alors plus ou moins large. Le caractère dynamique d'un enregistrement doit ainsi être préservé, afin de pouvoir retrouver le même résultat sur la base des mêmes données. Dans le cas du chromatogramme, cela signifie qu'une impression après intégration peut ne pas être suffisante pour interpréter un résultat, et qu'un accès aux données stockées électroniquement doit être maintenu.

Audit trail

L'audit trail est un journal d'évènements sécurisé, traçant de manière horodatée les modifications apportées sur un système, et généré par le système lui-même. L'objectif de l'audit trail est de pouvoir reconstituer les événements liés à toute création, modification ou suppression d'une donnée critique.

L'audit trail est considéré comme une métadonnée car il permet de connaitre le "quand", le "qui", le"quoi" et le "pourquoi" d'une modification.

Sauvegarde et archivage

La sauvegarde est une copie des données, des métadonnées et des paramètres de configuration conservée dans l'optique d'une restauration en cas de perte des données originales.

L'archivage est le stockage des données sur le long terme, dans l'optique de pouvoir consulter les données tout au long de leur période de rétention.
L'intégrité des données sauvegardées, comme archivées, doit être maintenue.

Data Lifecycle

Le *Data Lifecycle* (cycle de vie des données) est l'ensemble des phases du processus par lequel les données sont enregistrées, traitées, revues, rapportées, conservées, récupérées et soumises à des revues. Il s'étend depuis la génération ou l'acquisition d'une donnée ou d'un ensemble de données, jusqu'à leur destruction ou effacement.

L'intégrité des données doit être garantie tout au long de leur cycle de vie. Ceci implique la maitrise du *Data Lifecycle*. En d'autres termes, toutes les étapes de chaque donnée ou enregistrement critique sont identifiées et comprises (ex : création, stockage, transfert, modification, archivage), de façon à être en mesure de détecter tout risque de modification ou d'altération. La revue du *Data Lifecycle* permet ainsi de comprendre toutes les manipulations subies par les données (ex : calculs, exclusions), et de remonter jusqu'aux données brutes. Cette revue fait partie du système de *Data Governance*.

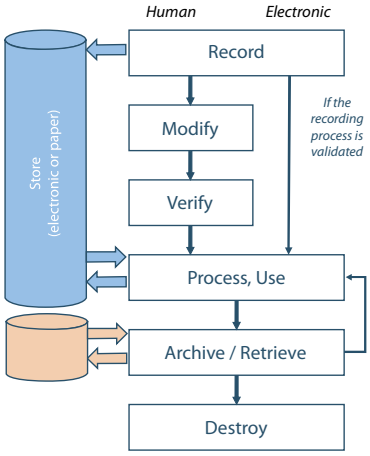

Figure 1. Data Lifecycle

⇢

Data Governance

La *Data Governance* correspond à l'ensemble des dispositions qui visent à garantir que les données, quel que soit le format dans lequel elles sont générées, sont enregistrées, traitées, conservées et exploitées de façon à garantir un enregistrement complet, cohérent et exact tout au long de leur cycle de vie.

La *Data Governance* fait partie intégrante du système qualité pharmaceutique et s'appuie sur les 3 piliers que sont le **comportement**, **l'organisation** et **la technique.**

Un comportement aligné sur les principes de *Data Integrity* implique notamment :

- la compréhension de l'importance du sujet par l'ensemble du personnel concerné (au moyen par exemple d'un code de conduite, ou d'un code d'éthique),
- l'implication du management,
- la remontée et le traitement correctes des déviations.

Les dispositions organisationnelles peuvent être, par exemple :

- une approche fondée sur le risque,
- la mise en place de procédures
- la formation du personnel
- la ségrégation des rôles,
- la revue des données et les vérifications de routine,
- les revues périodiques et la surveillance du système.

Les dispositifs techniques concernent, entre autres, les sujets suivants :

- l'utilisation de systèmes informatisés ou automatisés validés,
- l'implémentation d'*Audit Trails*,
- la sécurisation des supports d'enregistrement,
- le contrôle des accès,
- la mise en place de systèmes de sauvegarde et d'archivage.

La revue du système de *Data Governance* permet d'évaluer la bonne interaction, au sein des services, entre les comportements, les mesures organisationnelles et les dispositifs techniques.

Le sujet *Data Integrity* ayant été assimilé pendant longtemps à une problématique informatique, les dispositifs techniques sont bien souvent en place, qu'il s'agisse de fonctionnalités informatiques de sécurisation et de contrôle, ou d'automatisation de processus, destinée à limiter les erreurs liées aux interventions humaines.

La mise en place de solutions techniques n'est cependant pas suffisante si elles ne sont pas exploitées de manière appropriée et efficiente. L'*audit trail,* par exemple, doit faire l'objet d'une revue, pour garantir qu'aucune modification non contrôlée n'a été effectuée. Cette revue ne se fait pas de façon exhaustive sur tous les événements survenus sur le système mais se focalise sur les écarts observés dans le fonctionnement du processus d'obtention de la donnée, et ce uniquement sur les données critiques. La présence d'un *audit trail* au sein d'un système ne garantit donc pas à elle seule l'absence de modifications contrôlées, encore faut-il que les règles d'exploitation de celui-ci soient définies et suivies, comme la mise en place d'une procédure de revue systématique de certains évènements, ou l'utilisation validée d'un rapport de revue par exception.

L'aspect organisationnel du *Data Integrity* suppose une certaine maturité sur le sujet, avec une bonne implication des métiers et du management, et une prise de conscience des risques au niveau de l'entreprise, ce qui nécessite parfois une évolution de la culture.

La *Data Governance* suit une approche basée sur le risque, pour identifier les données critiques (*Data Criticality*) et les risques d'altération qui leur sont associés (*Data Risk*), afin d'ajuster les efforts de contrôle au juste nécessaire, et de manière équilibrée avec les autres activités qualité. Le niveau d'effort à apporter pour maîtriser les données est fixé en fonction de la criticité de celles-ci et de leur impact sur les CQA (*Critical Quality Attributes*) ou les données libératoires. De même, la graduation de l'effort est proportionnelle à la possibilité de détecter une altération des données. L'efficacité des dispositions en place est enfin pilotée et revue périodiquement.

Data Criticality

Pour déterminer la criticité d'une donnée, les questions suivantes se posent :

- "Quelle décision la donnée influence-t-elle ?"
- "Quel est l'impact des données sur la qualité ou la sécurité du produit ?"

Les décisions sur lesquelles les données ont une influence diffèrent en importance, et l'impact des données sur une décision varie également. L'effort se focalise sur les données les plus critiques.

Data Risk

Le degré de vulnérabilité d'une donnée, en regard d'une modification non contrôlée (volontaire ou non), est évalué au travers d'une analyse de risques. Les facteurs à prendre en compte pour cette analyse peuvent être la complexité du processus analysé, son degré d'automatisation ou la subjectivité de l'interprétation des résultats. L'évaluation est réalisée au périmètre du processus exploitant, et non pas uniquement sur des fonctionnalités ou technologies informatiques, de façon à prendre en compte les interactions avec les utilisateurs ou les interfaces entre les systèmes, à chaque étape du cycle de vie.

La validation des systèmes informatisés, si elle est toujours nécessaire, n'est plus suffisante aujourd'hui pour garantir une maitrise des risques sur l'intégrité des données.

C'est ici que réside la principale évolution de l'approche à considérer : la mise en place d'un *Data Management* (*voir figure 2*).

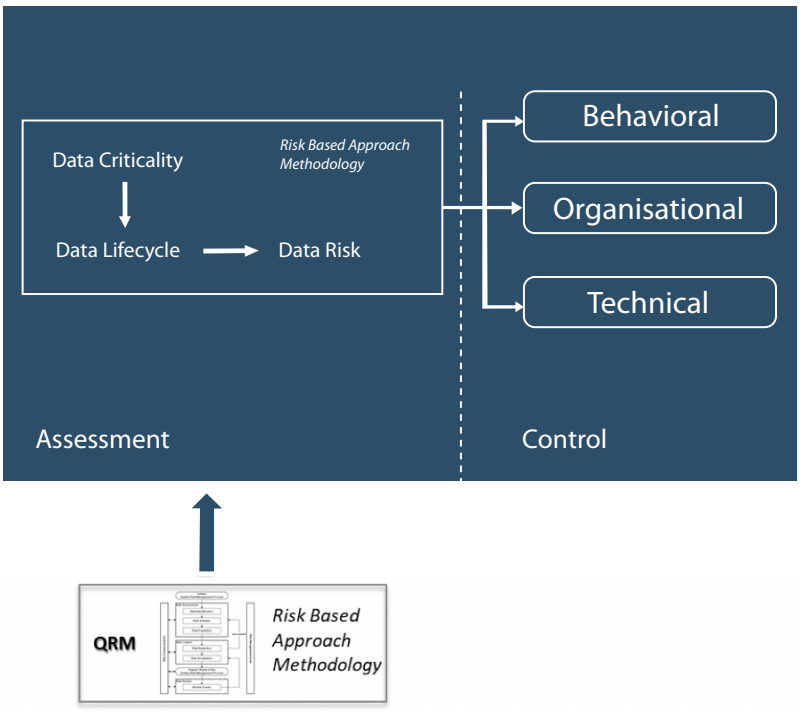

Figure 2. Data Risk

Data Management

Initialement, la maitrise du *Data Integrity* reposait sur la validation des systèmes informatisés. Chaque système suivait un cycle en V, au cours duquel les fonctions étaient définies, développées et testées, et le rapport de validation portait la compliance du système. Puis l'approche a évoluée vers une maitrise des risques en intégrant l'environnement du système et son cycle de vie. Cette Risk-Based Approach, portée notamment par le GAMP5[2] , prenait en compte la vérification de la mise en place de divers aspects organisationnels pour l'exploitation (formations, procédures …), mais restait située au niveau du système.

Aujourd'hui, la multiplication des systèmes informatisés, bien souvent interfacés entre eux, ainsi que leur niveau de configuration élevé, corrélé à la complexité des flux, rendent cette approche obsolète. Il devient nécessaire de passer d'un modèle focalisé sur les systèmes à un modèle focalisé sur les données, avec la mise en place d'un *Quality Data Management*, structuré et piloté par le risque, et garantissant une cohérence de traitement au sein d'une organisation. Ce modèle porte une sensibilisation continue à l'intégrité des données sur l'ensemble des processus critiques, en considérant les données électroniques, les données papier, les interfaces entre les deux au cours du cycle de vie, et non plus seulement sur les systèmes critiques.

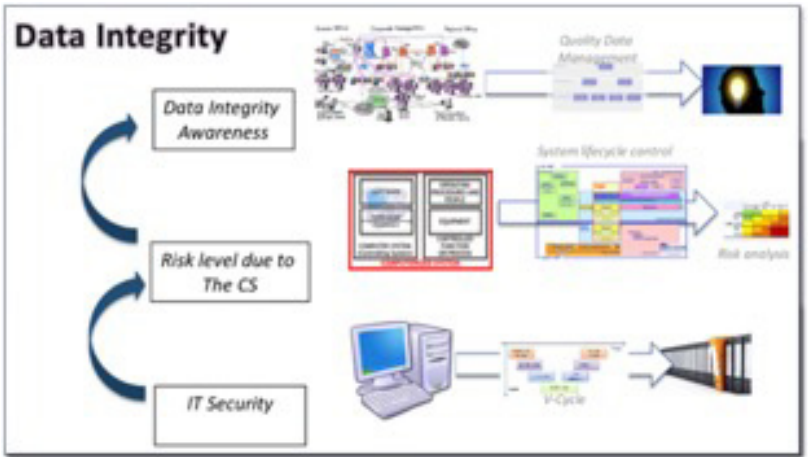

Figure 1. Data Management

Le *Data Management* est piloté par un référent, le *Quality Data Manager*, dépendant de la Qualité, et en lien avec tous les services gérant ou manipulant des données critiques supportant des décisions pharmaceutiques. Ce référent, garant des principes et des règles de *Data Integrity*, en pilote l'implémentation, apporte son support aux opérationnels et tient le management informé des risques les plus élevés.

Conclusion

Les bonnes pratiques de Data Management et de Data Integrity donnent confiance dans la robustesse des décisions pharmaceutiques et font partie intégrante du système qualité. Les risques d'altération des données doivent être pilotés au même titre que les risques sur les produits, car la qualité d'un produit pharmaceutique est étroitement liée à la qualité de ses enregistrements de traçabilité.

Bibliographie

[1] EMA Questions and answers: Good Manufacturing Practice - Data Integrity (August 2016)
PIC/S Draft Guidance PI 041-1: Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments (August 2016)
FDA Draft Guidance for Industry: Data Integrity and Compliance with CGMP (April 2016)
WHO Technical Report Series No. 996, 2016 - Annex 5: Guidance on Good Data and Record Management Practices
MHRA GxP Data Integrity Guidance and Definitions, Revision 1, March 2018
CFDA Draft Drug Data Management Standard, October 2016
[2] GAMP5, A Risk-Based Approach to Compliant GxP Computerized Systems, ISPE, 2008