

Giesecke & Devrient  
**Pitfalls in Serialization Backends**

White Paper



Giesecke & Devrient

## Management Summary

When setting up the overall infrastructure to make pharma production processes compliant with multiple serialization requirements worldwide it is of importance to consider a set of relevant, however not obvious constraints already from the beginning. This white paper touches questions of performance and scaling, security requirements, business continuity topics, and last but not least total cost of ownership.

## What is it all about ?

Serialization of individual, in general indistinguishable items is not a new idea. Think for example of license plates of cars. Identical type and color might make two cars so similar that you would not be able to tell one from the other. By recording the unique license plate, however, they become two different representations of one type of object. The same holds true for machinery and spare parts, luxury watches. Banknotes, credit cards, tickets and passports are serialized and personalized as well.

Maybe you already asked yourself the question: Why does serialization of pharmaceutical products become mandatory exactly these days? One part of the truth may be: because it is technically possible ! Memory and storage prices decrease exponentially since decades. And while it has been feasible to manage the comparably small volume of serial numbers for spare parts or watches by selfmade data bases – or worst case manually – it would have been quite a challenge to handle annually billions of object IDs and their related lifecycle events.

Times have changed. Storage is cheap now, computers become faster and evermore connected and as a consequence a huge wave of data is rolling towards us. The disruptive new dimension of data processing carries a danger: the human mind tends to underestimate the new quality of the challenge. It is probably insufficient, if you have a solution for the small size problem and just plan to multiply this by – say – hundred or thousand. The term “emergence” is used to describe that systems tend to exhibit new characteristics when scaling them up.

Let us consider an example: While it is common and beneficial practice to carry out backups of stored data in order to prevent data loss in cases of hardware failure or hazards, this concept is simply outdated for very large data volumes. Of course, you can save all the data onto tapes or disks, but the necessary time and the number of media will scale at least proportional to the data volume. If you need one hour for the backup today you will need one hundred or even one thousand hours tomorrow. How often do you plan to run this type of backup job ? How fast do you think you can manage the data restoration after failure ?

This white paper aims at spotting upcoming challenges in order to avoid pitfalls which in the simplest case are nasty and expensive, in the worst case they are prone to destroy data, shut down the production and ruin the reputation of a brand.

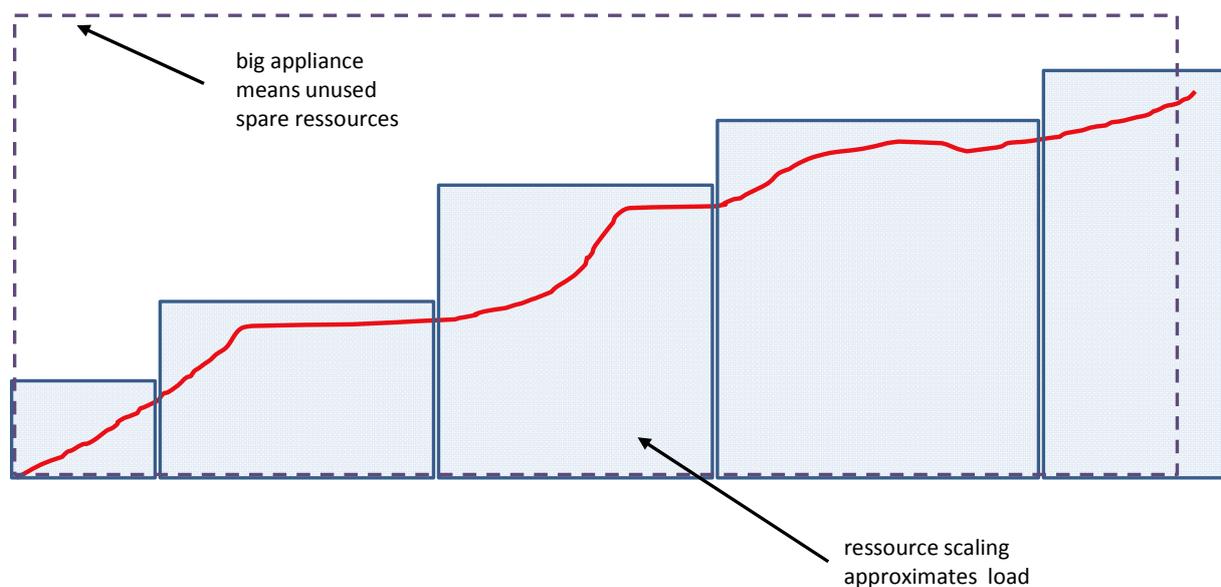
## Serialization is not done when the codes have been printed.

As a matter of fact, printing individual codes onto each and every single item running through a packaging line at 120 m / min is a challenge for printing equipment and quality assurance devices. But this is, of course, not enough. Aggregation, ERP connectors, line management and last but not

least formally correct reporting to national authorities build a machinery where one gear has to fit into another. A complete serialization solution has to serve all these – partially supplier dependent - interfaces and bridge the gaps. Here, flexibility is of primary importance in order to protect investment into a solution. Changing national reporting standards, upgrading or replacement of IT in the packaging line and updates of the ERP software or data model will lead to adaptation in the serialization and track and trace software. Modular concepts help to confine the adaptation work and the validation effort.

## Scaling is always too big, but how much ?

IT resources for mission critical process steps need always to be sufficiently large and powerful in order to cope with heavy duty situations. So sizing the architecture is always about “too big”. The important parameter is granularity. How good can you meet the ideal size of storage and calculation power ? If one option turns out to be too small, how big is the next possible option ? Or how small ? Do you need a complete additional high end appliance ? Do you have to adapt the data model or the partitioning and replication concept ? The smaller the size of one additional device - the more precise the approach towards the ideal capacity thus avoiding unused spare capacity which only may be necessary on a later stage. This is particularly true for big data applications where the unadapted reserve would be really large and expensive.



## Security is the basement, not the garden fence.

Serialization data is sensitive data because it represents valuable and sensitive items. Accidental or intentional manipulation of data sets will cause inconsistent information and can compromise patients' health. It is self-evident that access to the data has to be subject to user and rights management. A precise definition of roles and the attributed rights allows to confine data handling according to the need-to-know principle. Generation of serials, mass operations, change of item status and characteristics, deletion of data, blacklisting, remote access and much more need to be assigned as read or write access rights to those players and stakeholders who fulfill specific tasks and process steps like product managers, production managers, CMO, quality assurance, auditors, system administration etc. In particular the creation of a new user is a sensitive task. Furthermore password policies need to be defined and implemented.

But that is still not all to be considered. A possible security leak consist in theft of valid but still unassigned serial numbers which could be sold to unauthorized third parties for introducing counterfeits into the supply chain. This might only be possible for insiders, however, it may be possible. Intrinsic logical connections between item status, reported events and business steps can spot such inconsistencies and avoid forwarding of counterfeit serials to official verification entities.

As a further consequence, a fundamental rule must be the 100 % monitoring and logging of all activities in an audit trail which must be unaccessible for manipulations and tamper proof.

### **Playing the lottery does not make you an expert in random numbers.**

“Randomization” or “randomized serialization” as set forth in the GS1 standards implements the item individualization by choosing random serial numbers or more precisely “pseudo-random”. This latter difference being of significant importance as computer generated random numbers are still generated following generation rules, however pretty complex ones. This means nothing less than they are predictable once the generation rule is known. If the generation logic for the randomized codes can be re-engineered it will be easy to produce counterfeit products with virtually correct coding.

The way how randomized serial numbers are generated including the usage of pseudo random generators involves a secret. Here the experience in cryptography advises strongly to rely on publicly known and tested methods and algorithms whose security is based on a single secret element (“key”) rather than to use sophisticated secure procedures. The latter must be entirely hidden from any unauthorized access to stay secure (“security by obscurity”). Modern crypto technology provides at low cost both, high quality random numbers and certified security for the generation of unpredictable codes.

Furthermore, in the strict sense of randomness a random number can occur twice. Irrespective of the fact that this event should be extremely unlikely, it is of principal importance for serialization. One basic requirement is uniqueness of the serialization code. Double or multiple codes - be it due to errors in the code generation or due to perturbations in the marking process and administration – will render the system useless and even more importantly will destroy the users’ confidence in the concept. This will ultimately undermine the acceptance. So the generator must be prevented to generate the same number twice by appropriate measures.

### **Oops ...**

As already touched in the introduction, classical backup strategies appear no longer adequate for large data volumes. On the other hand, no compromises can be allowed with respect to availability of data and their analysis. Hardware failures like dying disc drives, accidents like fire, flooding or extreme weather events and last but not least human errors may not lead to data loss and subsequent process perturbations.

Modern data base concepts allow to store data redundantly according to predefined rules in a specified number of copies and spatially distributed over physically separated sites. Database

management dynamically cares about sufficient redundancy and balances the load for the existing data nodes. Sudden breakdown of one node will lead to a short reorganization without any impact on the operation and unrecognized by the ordinary user. After replacing the defective unit the database system re-balances the data storage again without noticeable effects.

### **I have it somewhere, wait a second ...**

One more „emerging“ feature when scaling up data volumes is the rising response time to queries. A common joke asks: how do you eat an elephant ? The winking answer: slice by slice ! Well, that may be true, however pretty time consuming. Modern database solutions tend to answer: employ a sufficiently large number of hungry people. In technical terms: parallel processing. Data is stored on a larger number of autonomous data nodes. Queries are performed by all instances in parallel. Increasing the data volume requires adding more data nodes, however, the query response time stays practically constant as the calculation power also has been scaled automatically.

Furthermore, the classical database architectures rely on a number of structured tables which grow with the number of entries and the number of data fields. For track and trace applications structured tables are not optimal. Product characteristics and event data which all relate to one item have various formats, data types, sizes and structures. One product item may experience many inspections and re-aggregations others very little to none. For special cases pictures or geolocations etc. need to be stored and so on.

These phenomena often addressed by the big data paradigm “Volume, Velocity, Variety” demonstrate the intrinsic power of next generation database technologies.

### **Access from Everywhere**

A frequently overlooked aspect is the user interface. Many stakeholder with different needs and requirements access the serialization and track and trace software in order to fulfill their roles and tasks with respect to production, packaging and commissioning. Even more, their workplaces are distributed over the manufacturing site or over several sites, in particular when contract manufacturers are involved. The software interface which grants access to the necessary functionalities like starting a production lot, performing data analysis, sending reports or manual registering voids or aggregations has to offer quasi ubiquitous possibilities to log in, authenticate and perform special operations. Ideally this software is completely web based. This means you need simply a computer, a browser and a network access to use the software package. Any CMO or logistics service provider can – provided enrollment and sufficient access rights – contribute item events and serial management, even over large geographically distances. No software rollout is required. No special system requirements are to be considered. No incompatibilities arise. No effort for software updates has to be budgeted.

An even more versatile aspect is the data base access for queries and feedback by mobile devices. In particular for all tasks related to storage and transport of products this will reduce the overhead for data management.

## Money talks

Serialization and track and trace solutions mean investment. Investment means risk. As mentioned earlier, flexibility is one way to deal with risk. Other aspects are sustainability, standardization and avoidance of lock-in situations.

Consider for example the aspect of license fees for software and technologies. Proprietary solutions often require a higher level of investment into royalty payment and lead to a one-to-one relation between supplier and buyer which cannot easily be replaced by another option. Open source components, in contrary, are supported by a large community and usually the technologies are free.

In the case of modern database concepts there exists another cost efficiency argument: the software and data storage runs on commodity hardware which is significantly cheaper than sophisticated specialized appliances. In particular, maintenance effort can be reduced and price reductions apply by purchasing company wide standard hardware. Upgrading the installation to the currently necessary extend without unused large spare capacities also reduces capex and risk.

## Make or buy ? Or utilize ?

Apropos capex: an interesting alternative – provided a respective company policy allows that - is the outsourcing of the track and trace solution to a service provider. On-premise there need to be installed special site interface devices. The data storage, analysis and reporting can be done as a cloud application.

## The best is yet to come...

Ok, we are done. Serialization runs, data storage and analysis is organized and reporting works as expected. Lean back and relax ? Wait a second: once you have the complete data about packaging and logistics processes, you can likely create more benefit out of it. Big data algorithms will help you to spot recurring obstacles or hick-ups in routine processes. Root-cause analysis allows to identify their origins and the related standard operation procedures can be improved. Modelling of processes and predictive analysis can be used to optimize throughput and productivity.

Track and tracing and product serialization thus develop from a mandatory investment to an opportunity.

## A best practice approach

The above mentioned pitfalls have been identified during intense work here at Giesecke & Devrient in the field of identity management and authentication and reliable answers to the challenges have been developed and introduced into the Giesecke & Devrient track and trace software product sertraQ. This software is regularly maintained to comply with changing or extended regulations for serialization code generation and reporting, it adheres to industry standards for maximum interoperability and employs state of the art technologies for fast processing of very big data volumes. Access to a web based demo installation can be provided to give a first impression of the look and feel.



For further discussion please contact:

Giesecke & Devrient  
Dr. Michael Fiedler  
Prinzregentenstraße 159  
81677 Munich  
Tel. +49 89 4119 - 0  
[michael.fiedler@gi-de.com](mailto:michael.fiedler@gi-de.com)



Giesecke & Devrient