Database preservation for industry customers

Florian Hartl (CSP GmbH & Co KG Company, Munich, Germany)

Florian Hartl is a software developer. He has been doing archiving projects for the automotive industry for over 8 years now. He is working with CHRONOS, a software from his employer CSP. Currently he is working as a key account manager and does project management for archiving projects.

Hartl did not bring a slide presentation but talked and showed a specimen of network connected automotive production tool: a screwdriver. Equipment used in the automotive industry produces a lot of data. This data may need to be kept for a lifespan of usually 30 years. The standard retention period for compliance reasons in this business is 25 years, but it is also good to have a time buffer of five additional years. The industry does not necessarily know what they need to do with the data in the future, but it is needed for legal reasons in case there is an accountability problem with a car in the future. What is important to them?

- Safe storage – stored not just in one place but all over the world
- Revision security – write once, read many – no feature for changing or deleting the files. Important that they can guarantee they have not changed the data.
- Documentation and knowledge – they must keep everything together in the documentation so that a future user has all the knowledge they need to analyse and understand the data.

Why is knowledge so important? Hartl has worked with old COBOL files and it was challenging, as the knowledge was not preserved with the data. Nobody could tell him anything about the data and it was created 10 years before he was born.

The Diesel Emissions Scandal triggered more demand for proving that data collected during the production period had not been manipulated at any later stage. Hartl's company CSP has thus grown from 30 to 120 people due to increasing demand from the automotive industry. CSP expects there will be more data in the future. The data that the depicted screwdriver type generates in a 15-year lifetime is about 15 terabyte. In the next five years, they expect screwdrivers to generate more than 15 terabyte per year.

Archiving with CHRONOS progresses in four steps:

- The first step is knowledge – gather producers and IT people and find out what they need. You must talk to multiple people in factories and the IT sector to find out which data is important to be collected and stored.
- If you have done a good identification phase, the archiving runs smoothly. CHRONOS uses a format that is similar to SIARD (Fitzgerald 2013). It would be simple to map between them.

- Then the data in the DBMS that were archived in the process are being replaced by links to the archive.
- After 30 years, long-term preserved data is deleted completely from the archives.

Data are split up into data files and metadata. Hartl said his company was not keen on encrypting data, as it is not easy to use the data without the ability to decrypt. Plain text files are best for storing the data. Hash values are also kept seeing if the file changes. CSP has a system that allows staff to check this on the fly.

CHRONOS is also used in the financial sector and other industries. Hartl concluded with a side-glance on the storage systems behind CHRONOS. The company uses a storage system called EfficientNodes that prevents changes on the hardware side based on hash values.

Figure 8: Industry screwdriver as used in the automotive industry.