

Long-term readability of data: Legacy software virtualization case-study

Natasa Milic-Frayling, Intact Digital Ltd



BACKGROUND

Reproducibility of studies is essential for validation of scientific results. Studies involve raw data collected from instruments and software applications used to analyse data.

Authenticity of the study results depends on the integrity of data and trustworthiness of software installations used for data analysis.

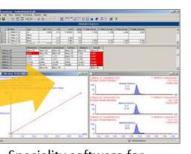
Long-term reproducibility of studies requires long-term management of data and long-term maintenance of software installations.



Data Integrity

Data Integrity is the extent to which data are complete, consistent and accurate throughout the data lifecycle.

- Complete All the required data needs to be collected and stored
- **Secure** Data must not be destroyed
- **Unaltered** Data must not be changed
- **Confidential** Data must not be disclosed to unauthorised individuals
- **Usable** Study metadata needs to provide context for experts reproducing study results.

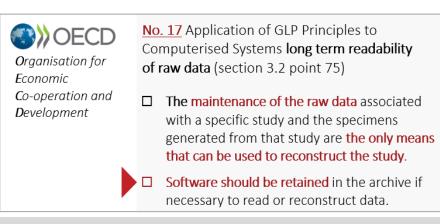


Speciality software for interpreting and analyzing raw data

Software Integrity

Software Integrity is the extent to which a software installation is functional, reliable, and usable throughout its lifecycle, from operational use (i.e., applied in new studies) to study reproducibility (i.e., applied to archived data).

- Functional Software installation must remain operational
- Unaltered Software must not be changed
- **Secure** Software installation must not present risks from cyberattacks and confidentiality breaches
- Accessible Software installation needs to be easily accessed for repeated use.



Specialty

software

PROBLEM

System decommissioning process removes instruments from active use and operations of an organization.

Software associated with the instrument is also decommissioned and removed from active use. Software is, however, required for data readability.

Currently, there is no systematic and principled approach to maintaining decommissioned software for data readability and study reconstruction.

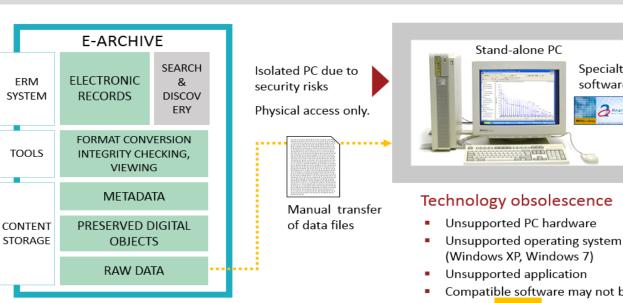
METHOD

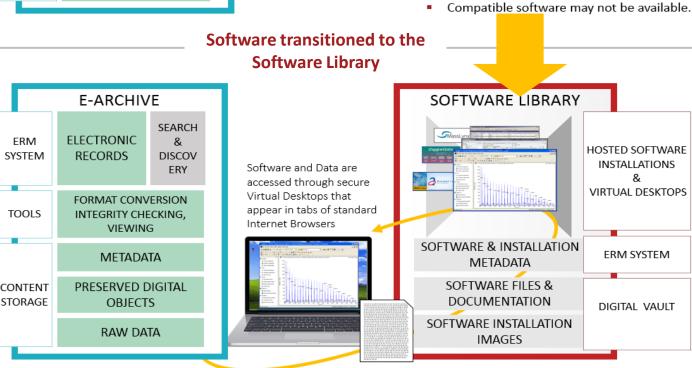
Software transitioning is a process of creating and validating software installations for long-term use in virtualized computing environments.

Through a systematic validation process, the installations are approved for study reproducibility and long-term data readability.

Software Library (SL) refers to a platform that hosts and provides access to the installed software through virtual desktops. Software Library complements Data Archives.

Software Integrity methods in the Software Library are aligned with the Data Integrity methods in e-Archives.





CASE STUDY

Analyst 1.4.2 software by ScieX has been used by a pharmaceutical organization since 2006.

The instrument set-up and the software installation followed the best practices and produced documentation:

- Installation qualification (IQ)
- Operational qualification (OQ)
- Performance qualification (PQ)
- **Re-qualification** after the initial IQ, OQ, PQ and in accordance with a user's Standard Operating Procedure (SOP) requirements.

The instrument has been decommissioned and study data stored in e-Archive. Analyst 1.4.2 installation on an isolated PC with Windows XP has been used to reproduce studies.

A Software Library (SL) instance was set up to transition Analyst 1.4.2 software from a lab PC, i.e., to create and validate a virtualized software installation using IQ, OQ and PQ documentation.

Besides virtual installations the Software Library includes:

- Digital Vault Stores software executable and installation files. Not accessible to end users for integrity reasons.
- Transfer Desktop Enables secure transfer of data into the Software Library using MOVEit platform generally used by the pharmaceutical organization for data transfers.

Installation and validation process

Set up a VM with MS Windows XP and SP3. Original IQ followed to create Analyst installation.

Dependency on 1.4.1 and 1.4.0 detected → Installed Analyst

1.4.2 with the dependent software. An authentication step detected \rightarrow changed registry settings.

Installation completed -> SL_IQ documentation. Migrated the isolated Analyst 1.4.1 installation to the Virtual

Desktop (VD) delivery environment. Provided secure VD access to the Analyst 1.4.2 installation.

> Applied subset of OQ procedure, relevant for study reproduction → SL_OQ

> > documentation.

Analyst 1.4.2 re-installed following SL_IQ process.

DEV

TEST

PROD

To conduct SL_OQ, used Transfer Desktop to access MOVEit with compressed study data. Downloaded study data to U: drive → Study data visible across desktops. Compressed data copied to C: drive of the Analyst Desktop. Applied integrity checks using checksum software. Uncompressed the study data in Project folder.

SL_OQ applied

Analyst Installation validated for study reconstruction. Data deleted from the SL.

> Subset of SL_OQ and data selected for re-validation -> SL_PQ documentation.

Analyst 1.4.2 re-installed within the hosting and delivery platform, following SL_IQ, and evaluated by applying SL_OQ. Software installations accessed through the SL portal.

> SL_PQ applied before starting study reconstruction.

Study reconstruction

Testing and PQ were performed using a copy of archived data: 1GB compressed; 9GB uncompressed. Data transfer from MOVEit to SL: 1 min; from Transfer Desktop to the Analyst Desktop: 2 min.

Checksum: < 20 sec. Decompression of 1GB: 10 min.



Figure: Example using data from a sample included with Analyst 1.4.2 installation.

Copyright©2016-2020 Intact Digital Ltd



Tel: +44 7870 566 404 Email: contact@intact.digital

Intact Digital Limited www.intact.digital