



iPRES 2021 Beijing • China 19-22 Oct. 2021  
17<sup>th</sup> International Conference on Digital Preservation



# Executable Archives

Software integrity for data readability and  
validation of archived studies

20 October 2021, iPRES 2021

Dr Natasa Milic-Frayling  
CEO & Founder, Intact Digital Ltd

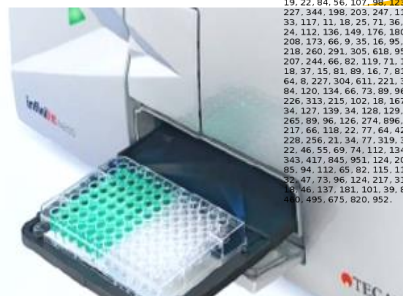
Dr Marija Cubric  
University of Hertfordshire, UK



- Raw data, collected from instruments must be archived and remain readable for a specified period of time (often decades)
- Research studies must be reproducible directly from archived raw data.

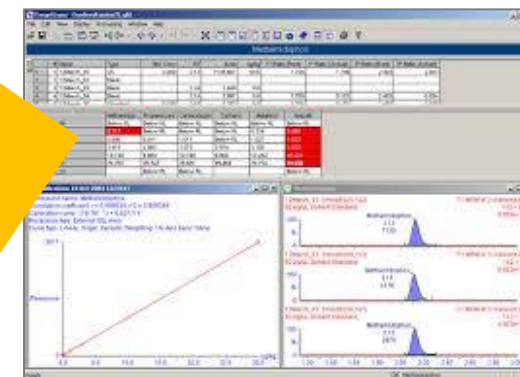


Specimen processing



Raw data

317.8, 92.73, 112.89, 67.318, 28.96, 107.41, 631.78, 146.397, 118.98, 114.  
 246, 348, 116, 74, 88, 12, 65, 32, 14, 81, 19, 76, 121, 216, 85, 33, 68, 15, 108, 68,  
 77, 43, 24, 122, 96, 117, 36, 211, 301, 15, 44, 11, 46, 89, 18, 136, 68, 317, 28, 80,  
 82, 304, 71, 43, 221, 198, 176, 310, 319, 81, 99, 264, 380, 56, 37, 319, 2, 44, 53,  
 28, 44, 75, 90, 102, 37, 85, 107, 117, 64, 88, 136, 48, 134, 99, 175, 89, 315, 226,  
 78, 96, 214, 218, 311, 43, 89, 51, 90, 75, 128, 96, 33, 28, 103, 84, 65, 26, 41, 246,  
 84, 270, 98, 116, 32, 59, 74, 66, 69, 240, 15, 8, 121, 20, 77, 89, 31, 11, 106, 81,  
 191, 224, 328, 18, 75, 52, 82, 117, 201, 38, 23, 217, 27, 21, 84, 35, 34, 109, 138,  
 49, 77, 88, 1, 81, 217, 64, 55, 83, 136, 251, 269, 311, 96, 54, 32, 120, 18, 132, 102,  
 219, 211, 84, 150, 219, 275, 312, 64, 10, 106, 87, 75, 42, 21, 29, 37, 10, 44, 18,  
 126, 115, 132, 160, 181, 203, 76, 81, 299, 314, 357, 351, 96, 11, 28, 97, 318, 238,  
 106, 24, 93, 3, 19, 17, 26, 60, 73, 88, 14, 126, 138, 234, 286, 297, 321, 365, 264,  
 19, 22, 84, 56, 107, 98, 82, 111, 234, 136, 7, 33, 43, 40, 13, 28, 46, 42, 107, 196,  
 227, 344, 198, 203, 247, 116, 19, 8, 212, 230, 31, 6, 328, 65, 48, 52, 59, 41, 122,  
 33, 117, 11, 18, 25, 71, 36, 45, 83, 76, 89, 92, 31, 65, 70, 83, 96, 27, 33, 44, 50, 61,  
 24, 112, 136, 149, 176, 180, 184, 143, 171, 205, 296, 87, 12, 44, 51, 69, 80, 34, 41,  
 208, 173, 66, 9, 35, 16, 95, 8, 113, 175, 90, 56, 203, 19, 177, 183, 206, 157, 200,  
 218, 260, 291, 305, 618, 951, 320, 16, 124, 78, 65, 19, 32, 124, 48, 53, 57, 94, 96,  
 207, 244, 66, 82, 119, 71, 11, 86, 77, 213, 54, 82, 316, 245, 303, 86, 97, 106, 213,  
 18, 37, 15, 81, 89, 16, 7, 81, 39, 96, 14, 43, 216, 118, 29, 55, 109, 186, 472, 213,  
 64, 8, 227, 308, 611, 221, 364, 818, 375, 128, 296, 1, 118, 93, 76, 10, 15, 23, 119, 71,  
 84, 120, 134, 66, 73, 89, 96, 230, 48, 77, 26, 101, 127, 936, 218, 439, 178, 171, 61,  
 228, 313, 215, 102, 18, 167, 269, 216, 218, 66, 59, 48, 27, 19, 13, 82, 48, 182, 119,  
 34, 127, 139, 34, 128, 129, 74, 63, 120, 11, 54, 61, 73, 92, 180, 66, 75, 101, 124,  
 265, 89, 96, 126, 274, 896, 917, 434, 461, 235, 890, 312, 413, 328, 381, 96, 105,  
 217, 66, 138, 22, 77, 64, 42, 12, 7, 55, 24, 83, 67, 97, 109, 121, 135, 181, 203, 219,  
 228, 256, 21, 34, 77, 319, 374, 382, 675, 684, 717, 864, 203, 4, 18, 92, 16, 63, 82,  
 22, 46, 55, 69, 74, 112, 134, 186, 175, 119, 213, 416, 312, 343, 264, 119, 186, 218,  
 343, 437, 845, 951, 124, 209, 49, 617, 856, 924, 936, 72, 19, 28, 11, 35, 42, 40, 66,  
 85, 94, 112, 65, 82, 115, 119, 236, 244, 186, 172, 112, 85, 6, 56, 38, 44, 85, 72,  
 26, 47, 73, 96, 124, 217, 314, 319, 221, 644, 817, 821, 934, 922, 416, 975, 10, 22,  
 18, 66, 137, 181, 101, 39, 86, 103, 116, 138, 164, 212, 218, 296, 815, 380, 412,  
 401, 495, 675, 820, 952.



Speciality software for interpreting and analyzing raw data

## ALCOA+ Data integrity requirements

### A

#### Attributable

Document must clearly identify who has created and contributed to them, and be protected against falsification or forgery of those details

### L

#### Legible

Stored documentation must be legible and easy to read

### C

#### Contemporaneous

Documentation should demonstrate and support contemporaneous record-keeping

### O

#### Original

Storing original copies of documentation guarantees accuracy and confidentiality

### A

#### Accurate

The processes and procedures by which companies record and keep their documentation up to date must ensure accuracy and reliability

### +Complete

All documentation must have an audit trail to show no data has been deleted or lost

### +Consistent

Documentation must be date and time stamped and stored in such a way to prove it has been assembled in the expected sequence

### +Enduring

Data must be available for as long as the regulation requires

### +Available

Data must not only exist, they must be accessible too, when and where required for reference and auditing purposes

- Archived study data in electronic form must remain immutable, readable and ‘dynamic’, i.e., interactive
- Reproducibility of study results requires software to remain functional and usable—**software integrity**

## Challenges

- Software is subject to rapid obsolescence if not regularly updated
- Updated software cannot guarantee the same functionality and output as originally used software version.

- OECD provides guidance on meeting regulatory requirements and adhering to the Principles of Good Laboratory Practices (GLP) when using computerized systems.



Organisation for  
Economic  
Co-operation and  
Development

## 6.16 Archive.

*When legacy systems can no longer be supported, consideration should be given to the importance of the data, and if required, to **maintaining the software for data accessibility purposes**. This may be achieved by maintaining software in a **virtual environment**.*

*OECD SERIES ON PRINCIPLES OF GOOD LABORATORY PRACTICE AND COMPLIANCE MONITORING, [Number 22](#). Advisory Document of the Working Party on Good Laboratory Practice on GLP Data.*

- Entire archived study is a ‘digital object’ to be preserved. It includes
  - Raw digital data and derived information
  - Study reports based on the processed data and include some of the visualized information.
- Significant properties of the digital object are intrinsically tied to the processing of data and observed through the software use.

Approach: Two preservation principles are considered **sufficient to preserve significant properties of the archived study**:

**Stored data integrity** – verified through checksum of electronic data files

**Software integrity** – verified functionality of the legacy software.

## ARCHIVIST

- Study records and raw data are stored and preserved by archivists, following the instructions of scientists.

## RESEARCHER

- Studies are reconstructed by scientists. Use of the specialty software is outside the scope of archivists' competences.

## IT SPECIALIST

- Software installation and management is a matter for IT specialists, particularly use of legacy operating systems that present security risks.

How to achieve preservation and reliable use of archived research study, considering

- The dependence on obsolete software and
- Separation of concerns and competencies among three types of specialists?

What would be an effective design of an **Executable Archive**, a system and services that extend the current electronic data archiving solutions and practices, recognizing

1. The fundamental **dependence of electronic data access on software**
2. A need for **long-term software installation support**.

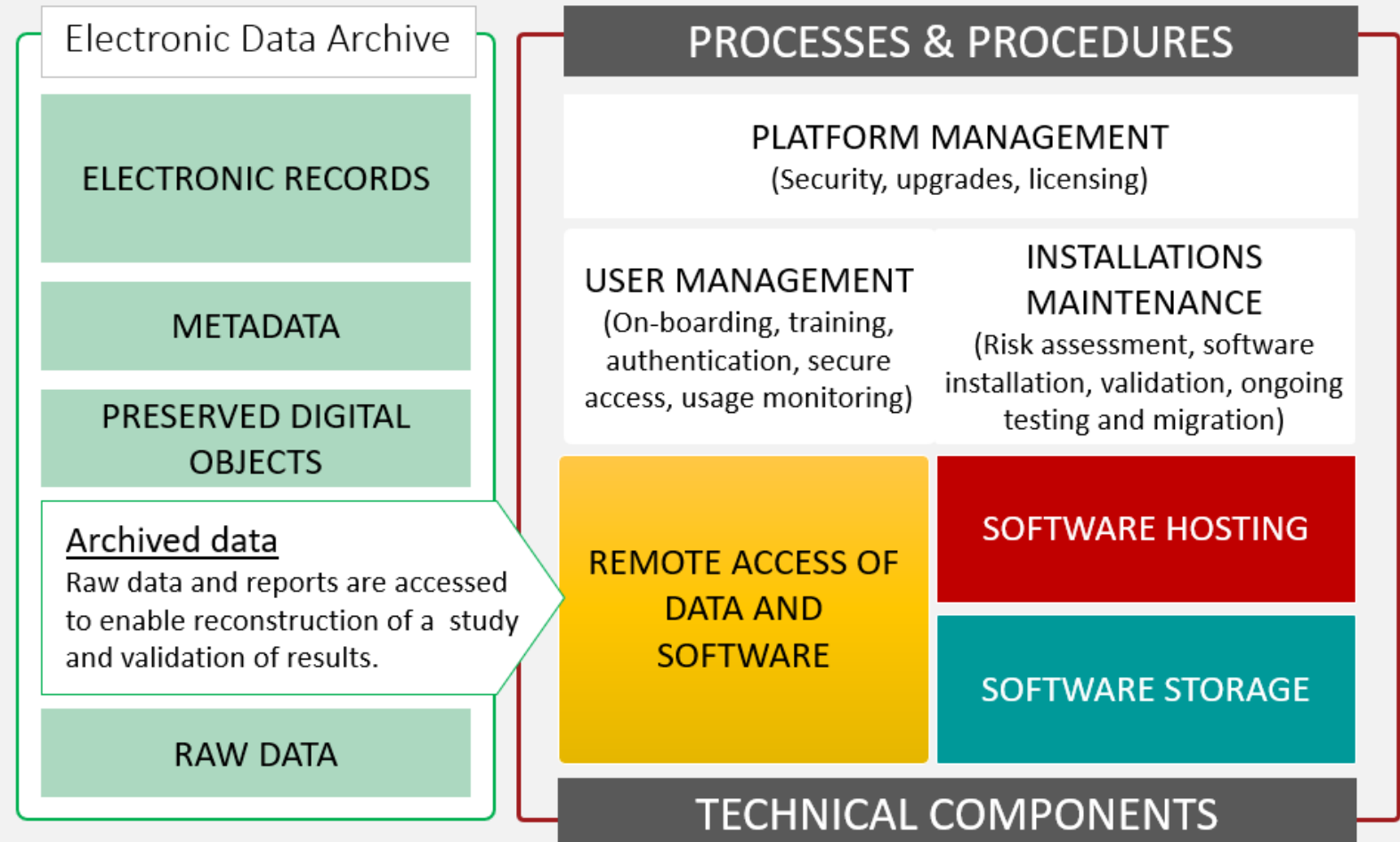
## APPROACH

- **Define an Executable Archive Framework**, informed by the software engineering and IT practices and scientific archive requirements
- **Explore designs** of technical infrastructure and services within the framework.



Electronic data archiving is expended with

- Complementary **technical components** to support software installation, hosting and use
- Corresponding **processes and procedures** for managing installation, validation, maintenance and use of the software.





Executable Archive framework is used to design and implement **Software Library** with functionality complementary to Electronic Data Archive.

## Electronic Data Archive + Software Library → Executable Archive

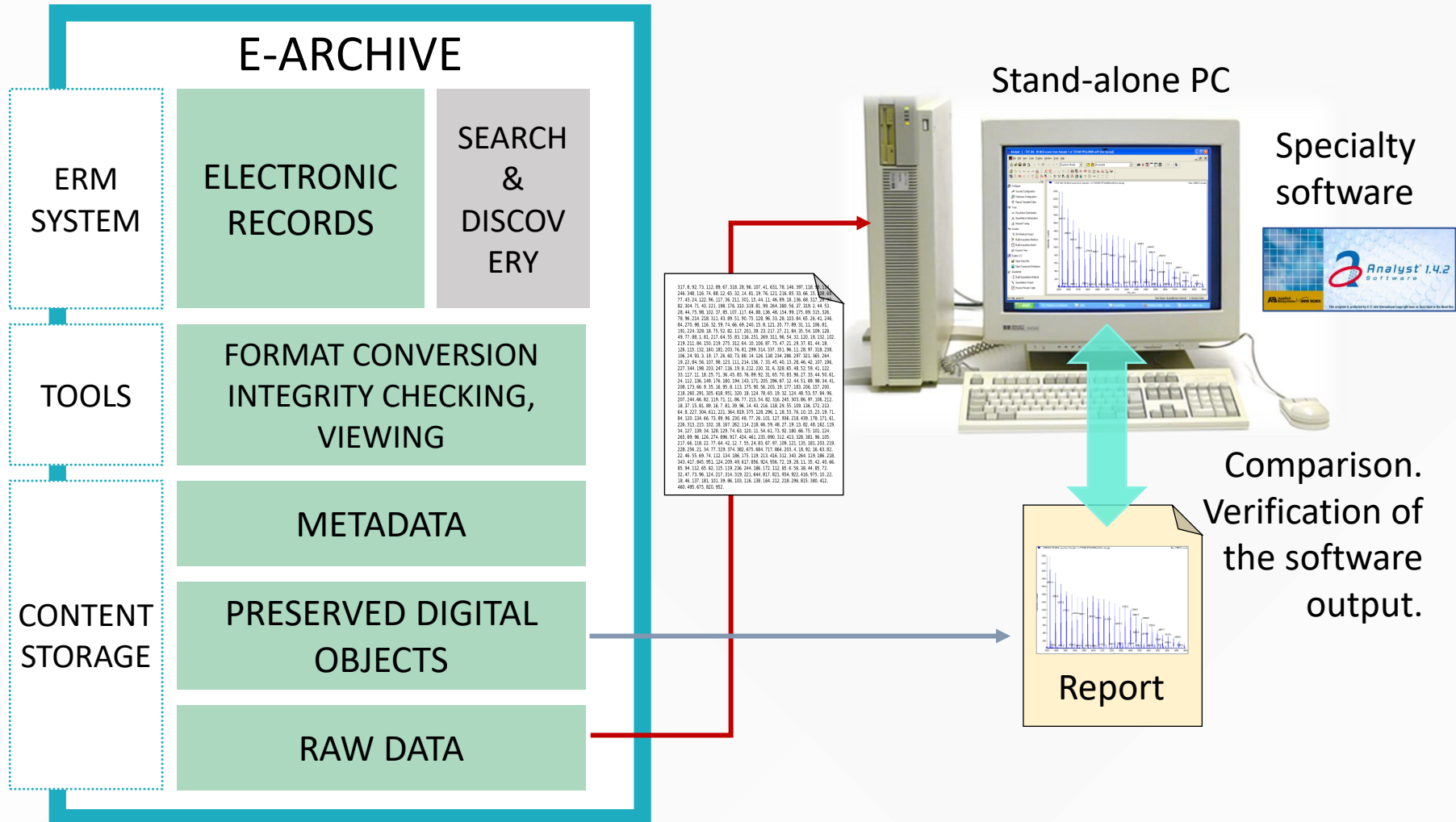
### Software Library platform and services

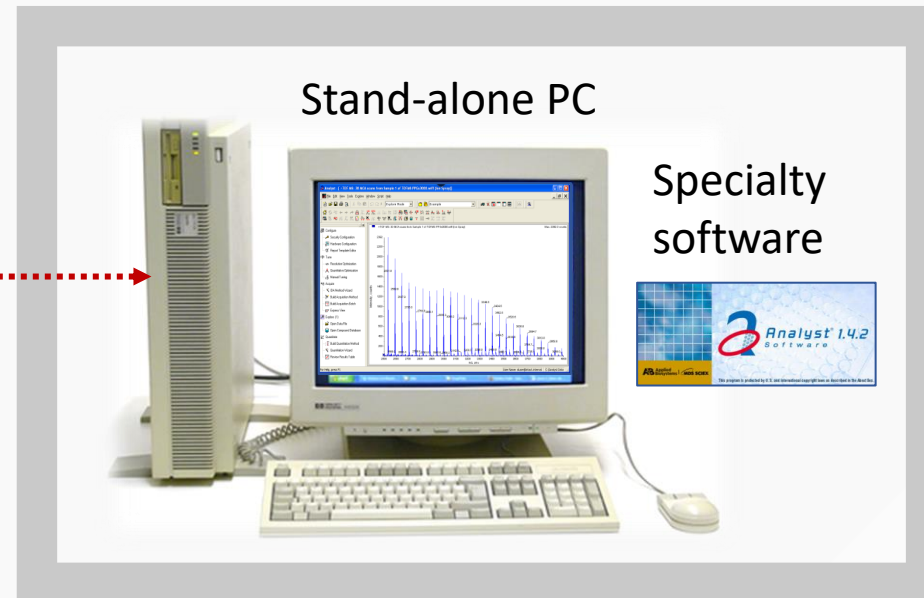
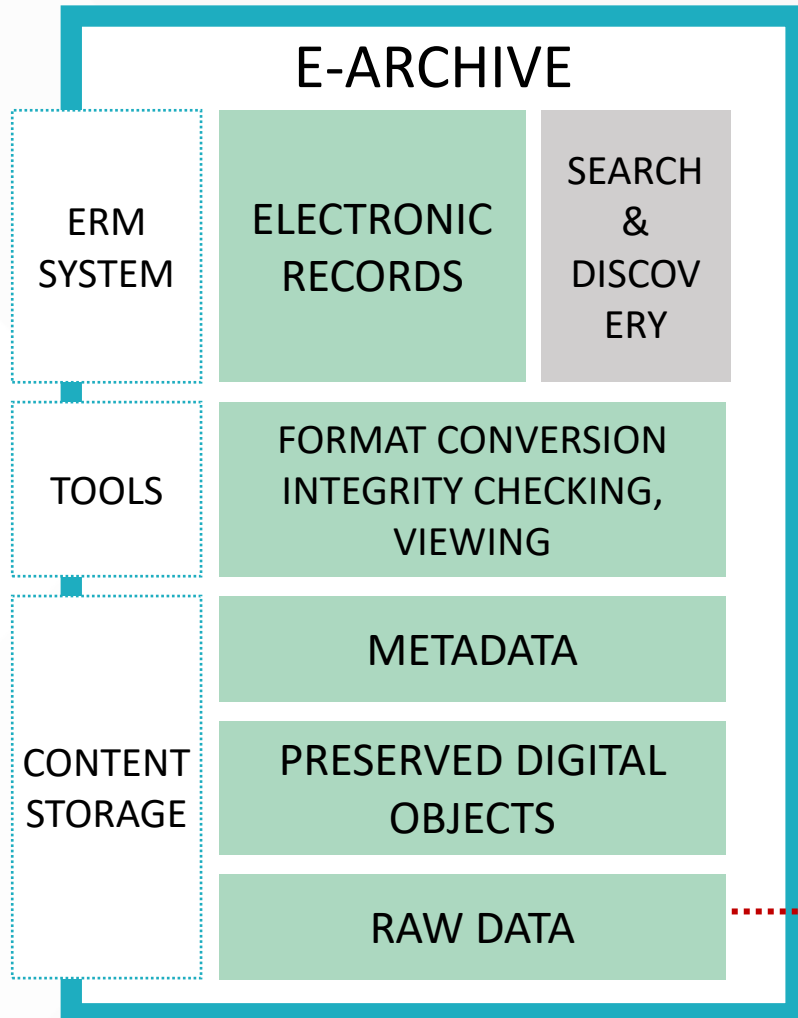
- Host the collection of validated software installations
- Provide secure connections to data repositories
- Enable reliable use of software and data.

A transition of software into the Software Library is a standard procedure that involves:

- Software installation
- Software validation, in the context of a task (e.g., a study reconstruction)
- Software maintenance.

# Requirements: Current practice





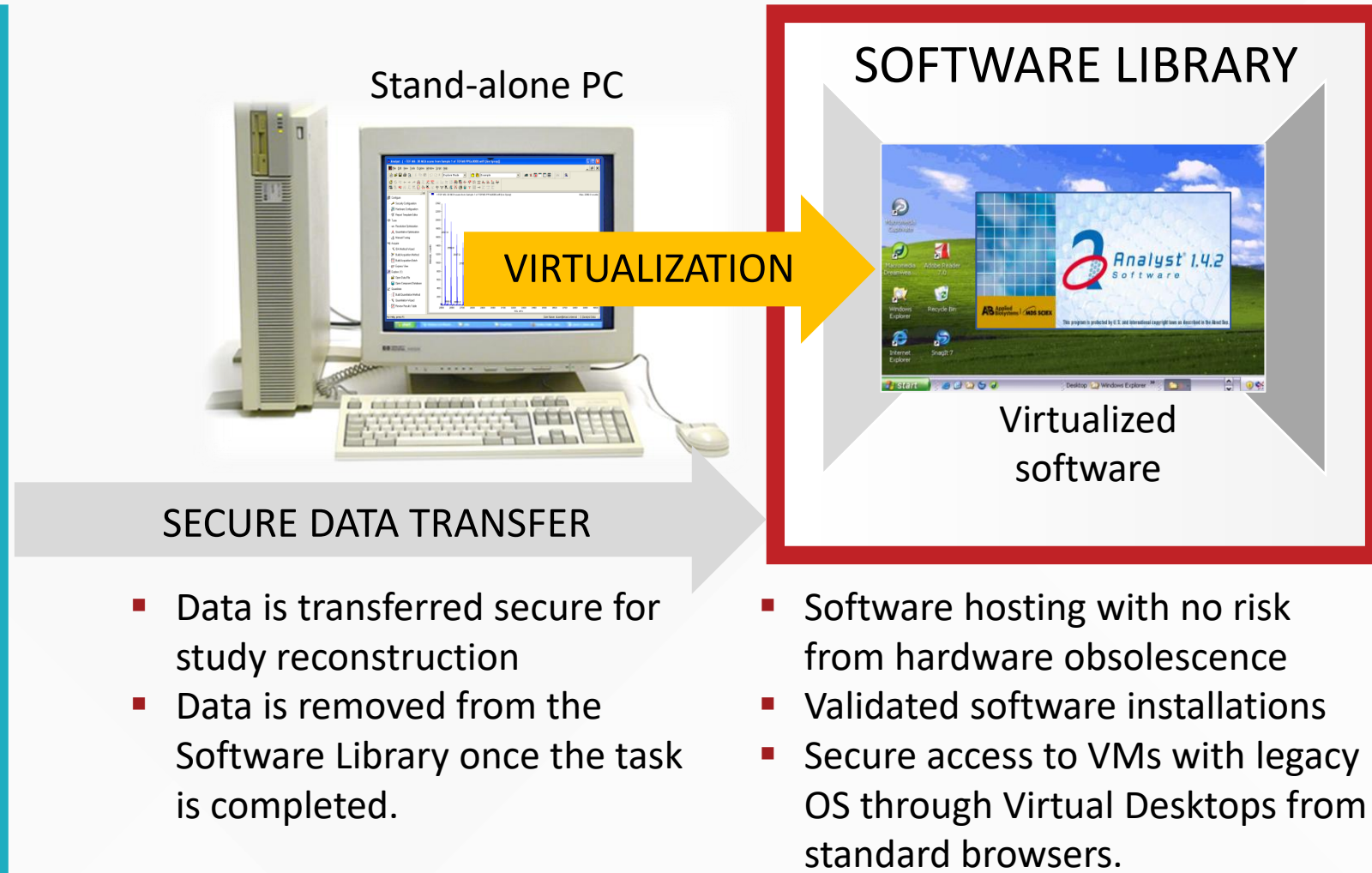
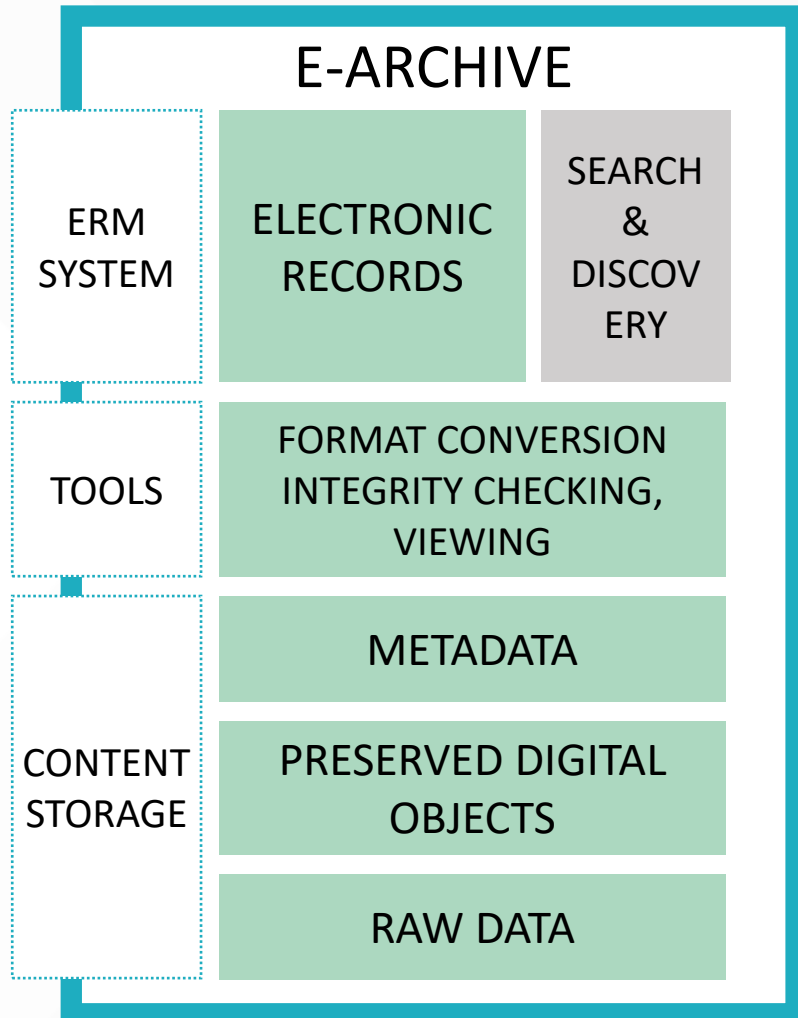
Physical  
transfer of  
data files

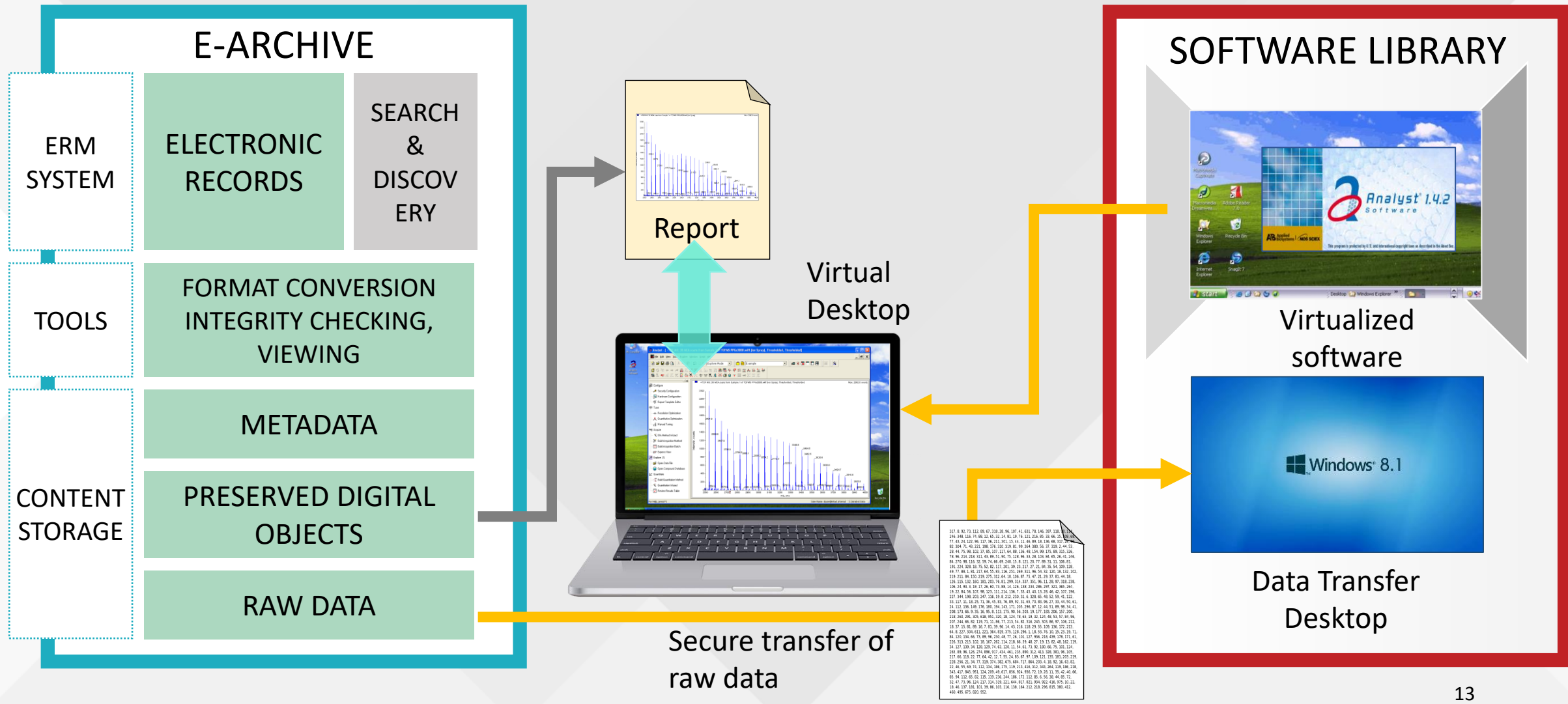
## Technology obsolescence

- Unsupported PC hardware
- Unsupported operating system (Windows XP, Windows 7)
- Unsupported application
- Compatible software may not be available.

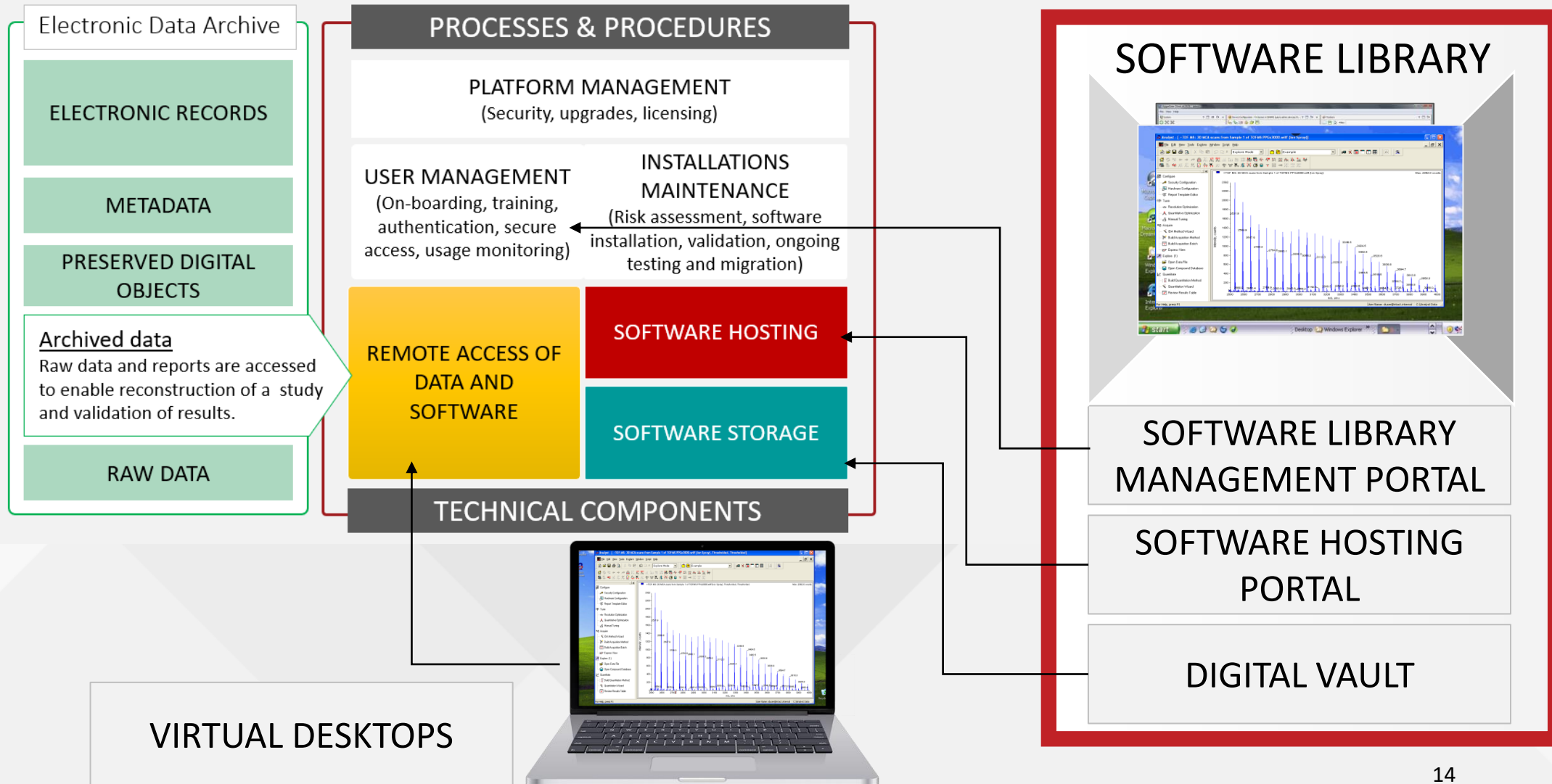
Isolation, to mediate  
security risks

Restriction to physical  
access only.





## EXECUTABLE ARCHIVE FRAMEWORK





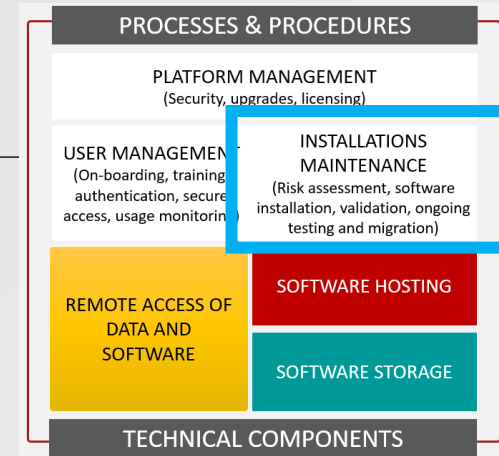
# Procedures: Installation management

## INSTALLATION QUALIFICATION (IQ → SL-IQ)

- IT** Create a sandboxed VM environment for software installation
- IT** Upload of software and software installation documentation
- IT** Document the process of installing the software in the VM
  - ▶ Follow the original **Installation Qualification** (IQ) used to create installations in the Lab.
  - ▶ Create **Software Library IQ** (SL-IQ) for the specific software.

## OPERATIONAL QUALIFICATION (OQ → SL-OQ)

- IT** Configure Virtual Desktops (VD) to support data transfer and software use.
- RS** Review the Operational Qualification (OQ) of the original software installed in the Lab
- RS** Document the testing of the virtualized software installation
  - ▶ Select and test software features that support the study reconstruction task
  - ▶ Document the operational qualification process SL-OQ for virtualized software installations.



### Roles

- IT** IT Specialist
- RS** Researcher
- AR** Archivist



# Procedures: Software integrity checks

## PERFORMANCE QUALIFICATION (PQ → SL-PQ)

RS

Create a **Performance Qualification (PQ)** test and select representative test data

- ▶ Select a minimal set of steps to establish the **software integrity**
- ▶ Perform the test with the original software (in the Lab) and virtualized software

RS

Document the PQ procedure for the virtualized installation to create SL-PQ

- ▶ Apply SL-PQ test before importing the study data.

## STUDY RECONSTRUCTION

AR

Use Transfer Desktop to transfer data into the Software Library environment

- ▶ Data is accessible to all the VMs available to the user

RS

Activate the desktop with specific software and gain access to data

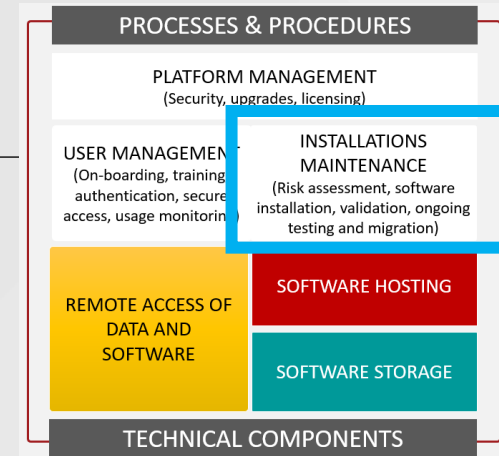
- ▶ Move the data to the VM if it needs to be used locally by the software

AR

Access the study documentation and reconstruction process.

RS

Reproduce the results and document the study reconstruction process.



### Roles

IT

IT Specialist

RS

Researcher

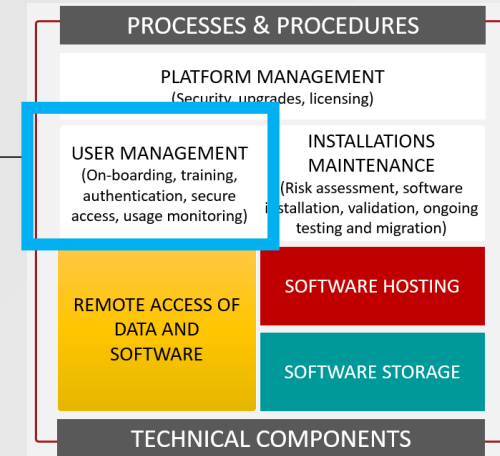
AR

Archivist

# Human factor management

Software Library supports a systematic capture of knowledge about software installation, maintenance and use through

- Regular user training
  - Refresher course every 6 months
  - Involvement in internal compliance audits and software verification annually or semi-annually and in regulatory inspections (every 2 years)
- Detailed documentation
  - Software management is documented by describing steps and capturing screenshots.



# Concluding remarks

- Archiving of digital content cannot be separated from long-term care of software

Data integrity + Software integrity → Preserved content  
(significant properties)

- Executable Archive Framework supports a design of software management systems that complement data archives.

Electronic Data Archive + Software Library → Executable Archive

- Software Library platform and services support activities of three expert roles: IT staff, archivists and researchers.

Optimization: Keep the Software Library ‘aligned’ with the stored data.

# Thank you!

DIGITAL



*Software is the **Yang of Digital**.*

*Interacting with **stored content files**, the **Yin of Digital**, it brings to life ephemeral existence of digital creations.*

Dr Natasa Milic-Frayling

[natasamf@intact.digital](mailto:natasamf@intact.digital)

[www.intact.digital](http://www.intact.digital)