Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Scientific Data Management and Archiving with AnIML

Dr. Maren Fiege
Waters GmbH

©2012 Waters Corporation

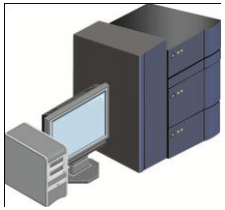
Overview

- Why Scientific Data Management?
- Why Standard Data Formats?
- Why AnIML?

©2012 Waters Corporation

Data Generation

Sampling



Analytical Data Acquisition

Data Processing

Results + Reports

Instrument Data

Processed Data

Result Data

©2012 Waters Corporation

Common Concerns

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

- Multiple locations and applications
- Integrating, comparing, reporting
  - different sources
  - different techniques
  - different vendors
  - different software
- Sharing data among peers
- Balancing scarce resources effectively

©2012 Waters Corporation

Regulatory Requirements

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

- Copies of Electronic Records for Inspection
  - Accessibility for Inspection
  - Integrity of Content and Meaning
  - Human Readable Form
  - Standard Portable Formats
- Retention and Maintenance of Records
  - GxP demands Archival of Records for Extended Time
  - Requirement to keep the Raw Data
  - Ability to Reprocess

©2012 Waters Corporation

Record Retention Requirements

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Industry Segment, Regulator and Type of Record	Typical Retention Period
Pharma: Good Laboratory Practices; records related to a new drug application (NDAs)	Date of submission plus 5 years
Health Care: medical records	Life of patient plus “n” years
Drug/device study records	Marketing application plus 2-3 years
Government records	20-50 years, or permanent
Copyright records (all organizations)	Life of copyright = 95 years or as business needs dictate
Patent records and supporting data	Application plus 17 years

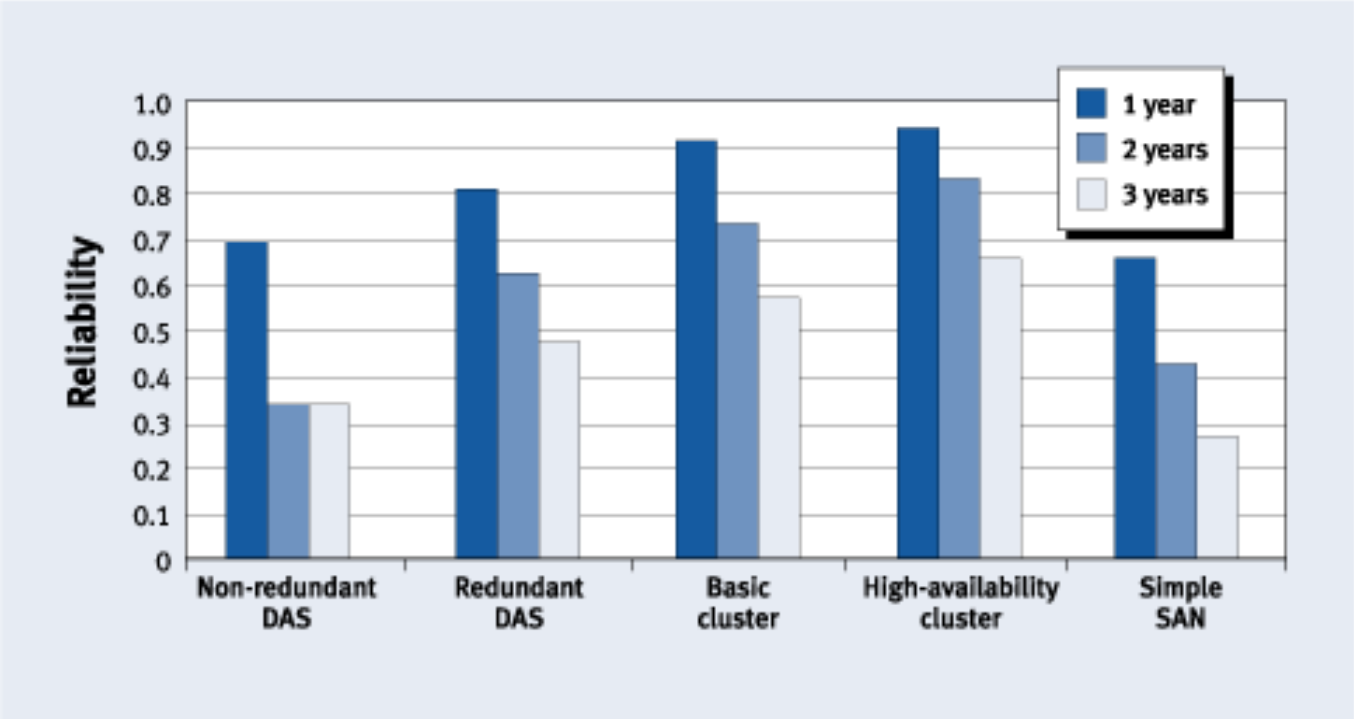
©2012 Waters Corporation

Lifespans

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Software lifetime: approx. 9 years*

Hardware depreciation: 3-5 years



Comparison of reliability over three years

*Software lifetime and its evolution process over generations, Tamai, T., Torimitsu, Y., Conference on Software Maintenance, 1992, Proceedings

©2012 Waters Corporation

Other Considerations

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

- Cost and effort of repeat analysis
  - May not even be possible!
- Litigation
- Instrument data not retrievable if laboratory or manufacturer goes out of business
- Other government requirements

©2012 Waters Corporation

Data Integration

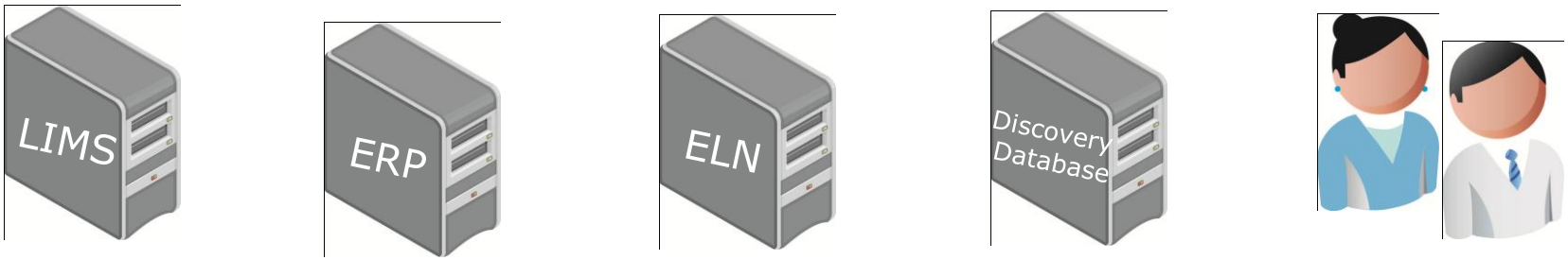
Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Instrument Data

Processed Data

Result Data

?



©2012 Waters Corporation

Capturing and Cataloging Scientific Data

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Instrument Data

Processed Data

Result Data

File Data (proprietary)

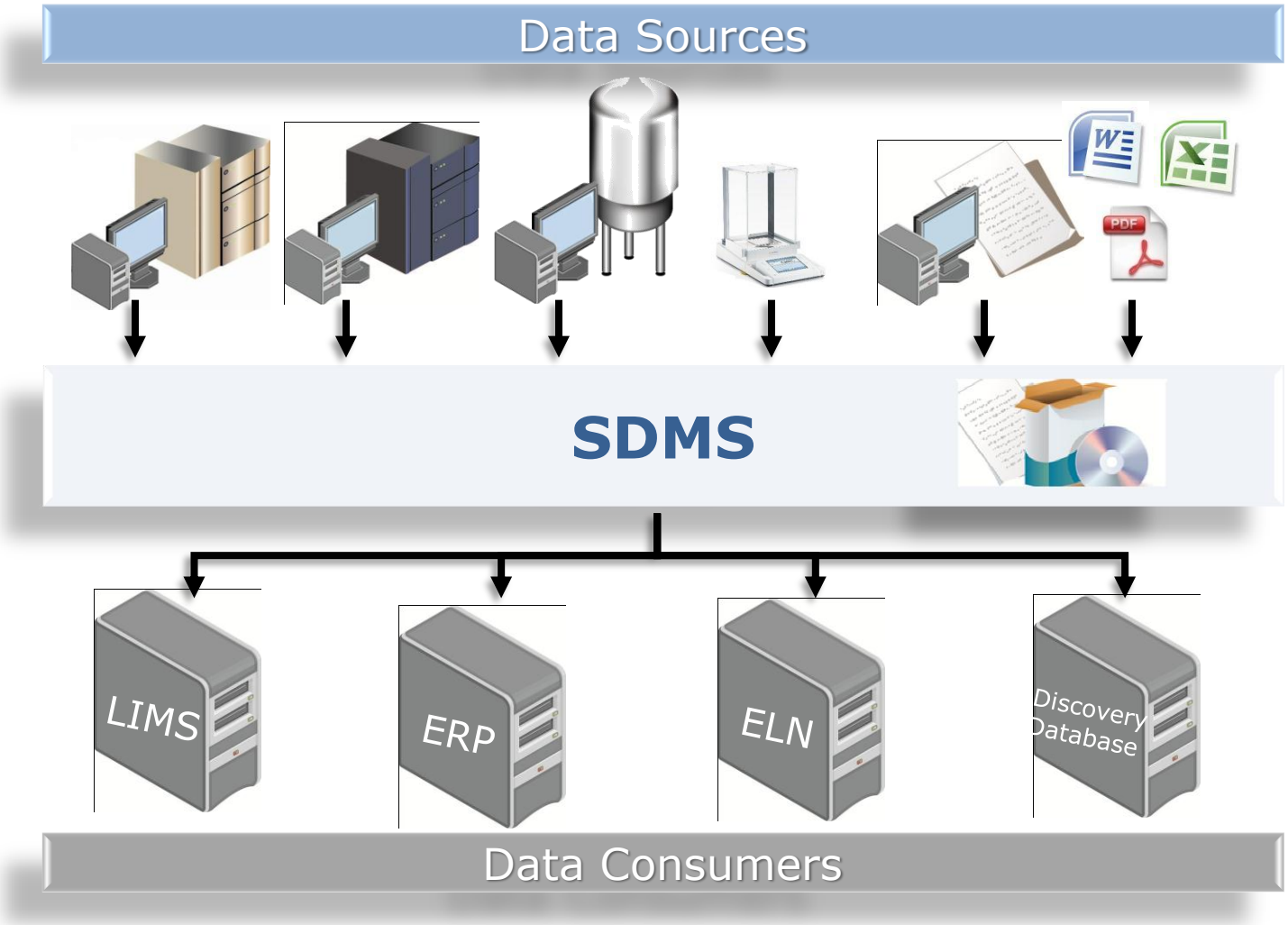
Printed Reports/ PDFs

Scientific Data Management System

©2012 Waters Corporation

Data Export/Re-use

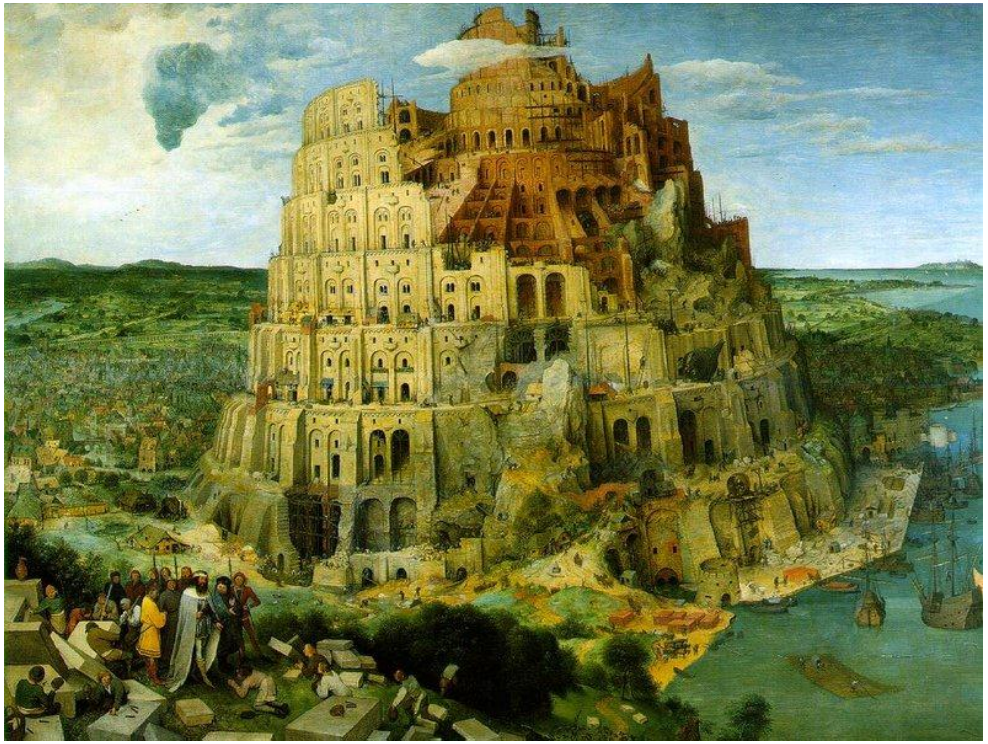
Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

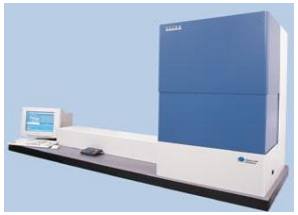


©2012 Waters Corporation

Why standard formats?

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™





©2012 Waters Corporation

Standards for Report documents

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

- Past: Paper, microfilm / microfiche
- TIFF - can be large, not easily searchable.
- EMF - fully scalable and re-usable
- EMF and PDF
  - more compact, often with a better quality
  - Metadata can be embedded
- PDF - usually device-independent
- PDF/A (ISO-19005-1, 2005)

©2012 Waters Corporation

Standards for Analytical Data

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

- IUPAC JCAMP-DX: since 1988
- ASTM ANDI (NetCDF): since 1992
- HUPO mzData
- mzXML (Proteomics MS)
- mzML

- AnIML

©2012 Waters Corporation

Proprietary vs. Standard Formats

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

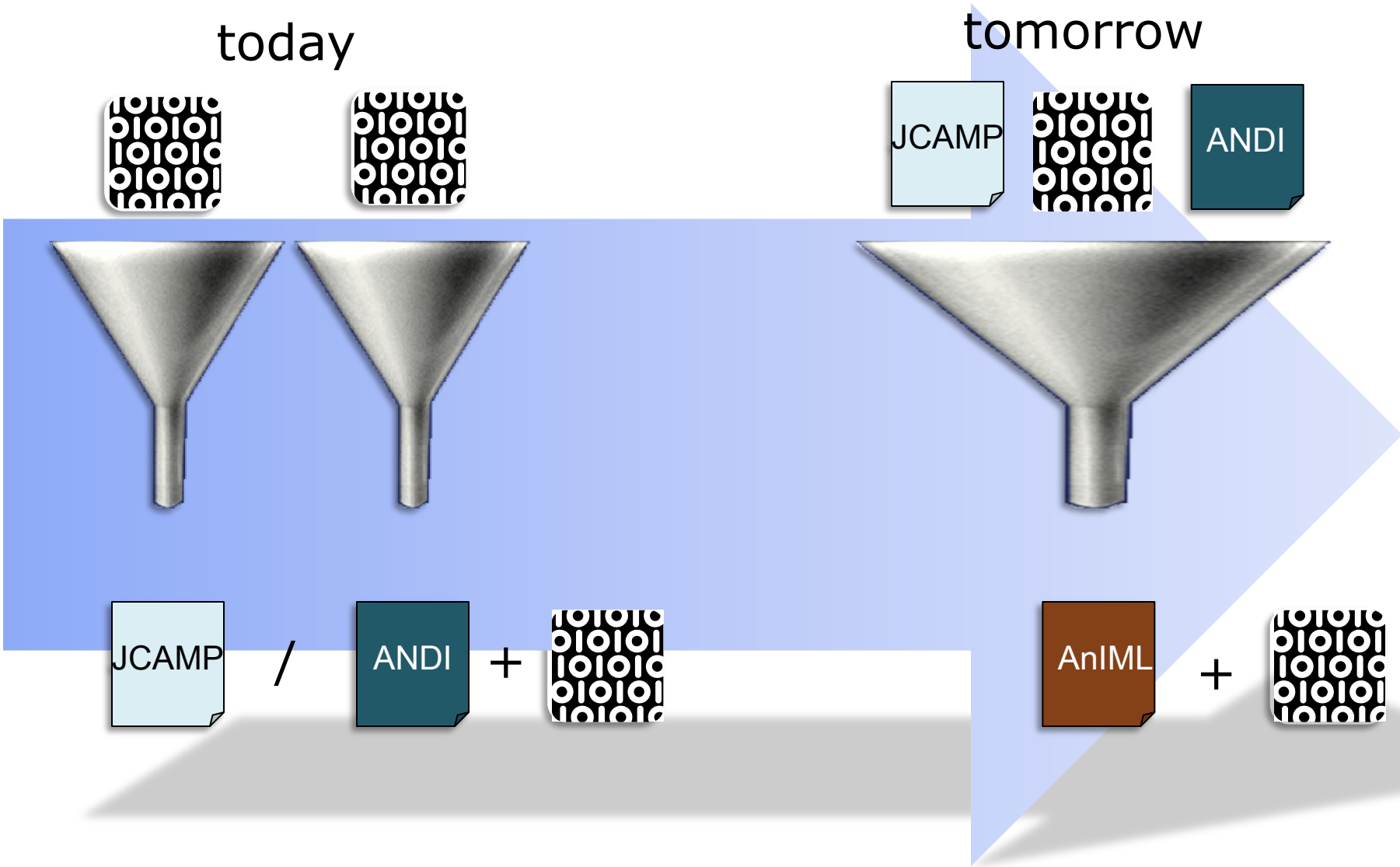
- Why don't vendors use Standard Formats?

Proprietary Formats	Standard Formats
Binary	ASCII-based (e.g. XML)
Compact	Verbose
Fast to Read/Write	Slow to Read/Write
Data Acquisition and Processing	Data Sharing and Long Term Stability

©2012 Waters Corporation

Conversion

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™



©2012 Waters Corporation

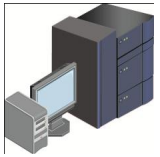
Why AnIML?

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

AnIML Data Set

- Workflow Information
- Metadata
- Instrument Data
- Processed Data
- Result Data
- Audit Trail, Signatures

Sampling



Instrument Data ←

Processed Data ←

Result Data ←

©2012 Waters Corporation

AnIML is…

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

- … XML-based
  - Human readable
  - Long term stable
  - Non-proprietary
  - Easy to index via metadata
- … Technique-agnostic
  - One format fits many
  - Easily shareable
  - Viewable with a generic viewer without the original application!
- … Easy to re-use and integrate
- … Compliant-ready

©2012 Waters Corporation

Summary

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

- Data Management
  - centralizes your data
  - Makes it accessible for sharing and integration
  - Keep original (binary) data and Reports
- Standard formats make it
  - Shareable
  - Viewable
  - Future-proof
- AnIML covers (nearly) all analytical data
  - With benefits!

©2012 Waters Corporation

Sources of Information

- The AnIML Web Site
  - http://www.animl.org
- The SourceForge Analytical Information Markup Language Project: Summary Page
  - http://sourceforge.net/projects/animl

©2012 Waters Corporation

Q&A

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Next:

Burkhard Schäfer, BSSN Software:

"AnIML in a Fully Integrated Laboratory"

©2012 Waters Corporation