entimo®

Dimensions of Standards' Versioning: From Variables to Program Code

Astrid Wolter

Entimo AG | Stralauer Platz 33-34 | 10243 Berlin | www.entimo.com

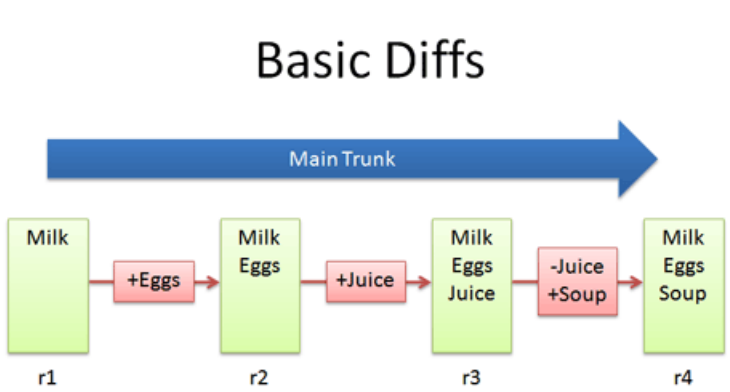
04 November 2015

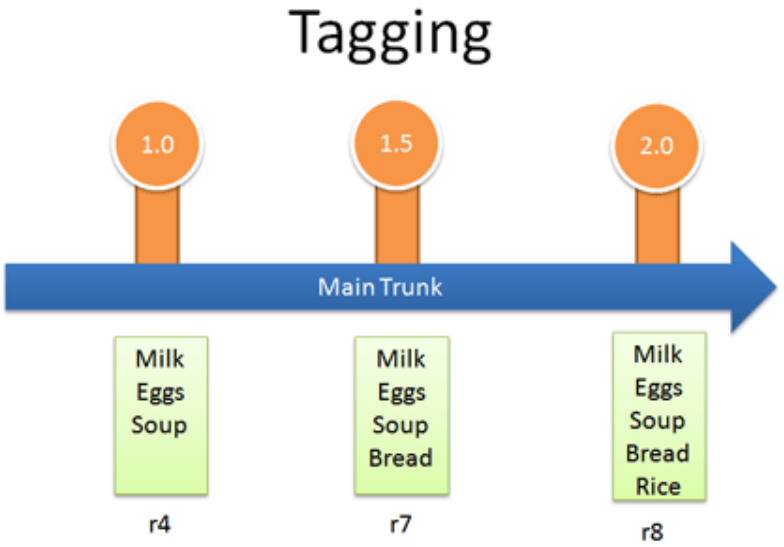
- Paradigms
- Best practices in software development
- Artifacts
- Challenges
- Solutions
- Conclusions

- **Version Control Synonyms:**
  - Revision Control, Source Control, Versioning…
- **Definition:**
  - Wikipedia: “Management of changes…”
- **Paradigms:**
  - Lock-modify-unlock (e.g. Visual SourceSafe)
  - Copy-modify-merge (e.g. SVN, CVS, Git, ClearCase…)

Classic VCS are file-based!

- **Concepts:**
  - Revision
  - Release, minor/major version
  - Head, trunk/main, branch
  - Diff
  - Conflict
- **Actions:**
  - Add
  - Check out (for edit)
  - Check in
  - Restore, revert version
  - Version tree
  - Tag, label
  - Merge
  - Locking, lock break…

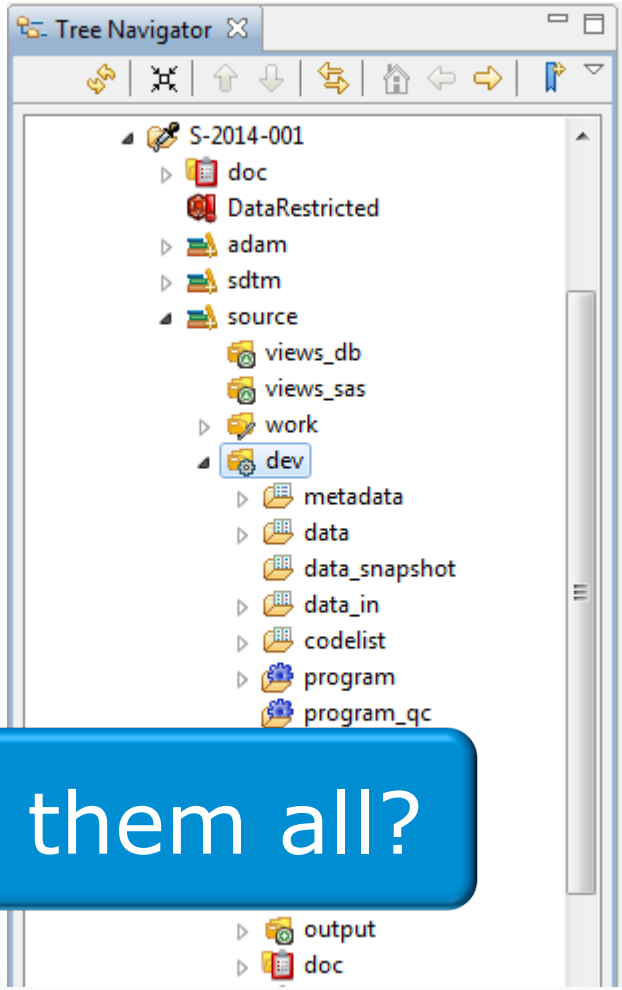




- **Best practices in software engineering:**
  - Descriptive commit message
  - Make each commit a logical unit
  - Incorporate changes frequently
  - Coordinate with your co-workers
  - Never refill/re-justify paragraphs
  - Don’t use long lines
  - Don't commit generated files
  - Understand your version control tool…

Is pharma world different?

- **entimICE repository content (sample):**
  - Data
  - Metadata
  - Codelists
  - Checks
  - Programs
  - Mappings
  - Outputs /TLFs
  - Logs..



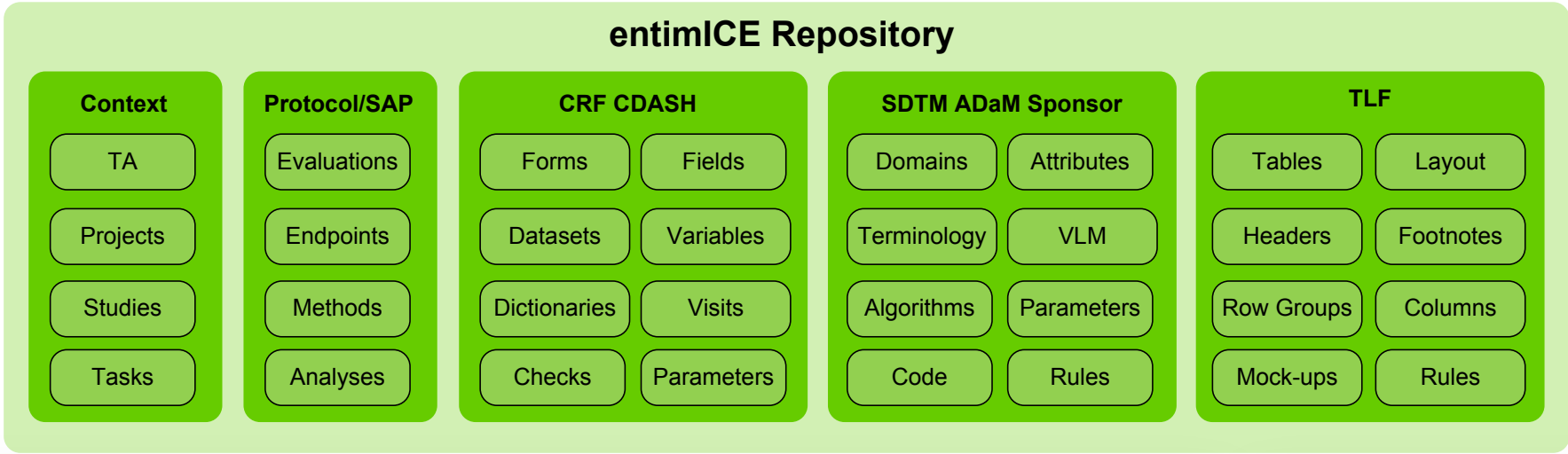
Can classic VCS handle them all?

- **Data types:**
  - SAS datasets
  - Data warehouse tables
  - CSV data
  - XML data…
- **Challenges:**
  - Mixed storage types (files and database)
  - Small, big and very big data

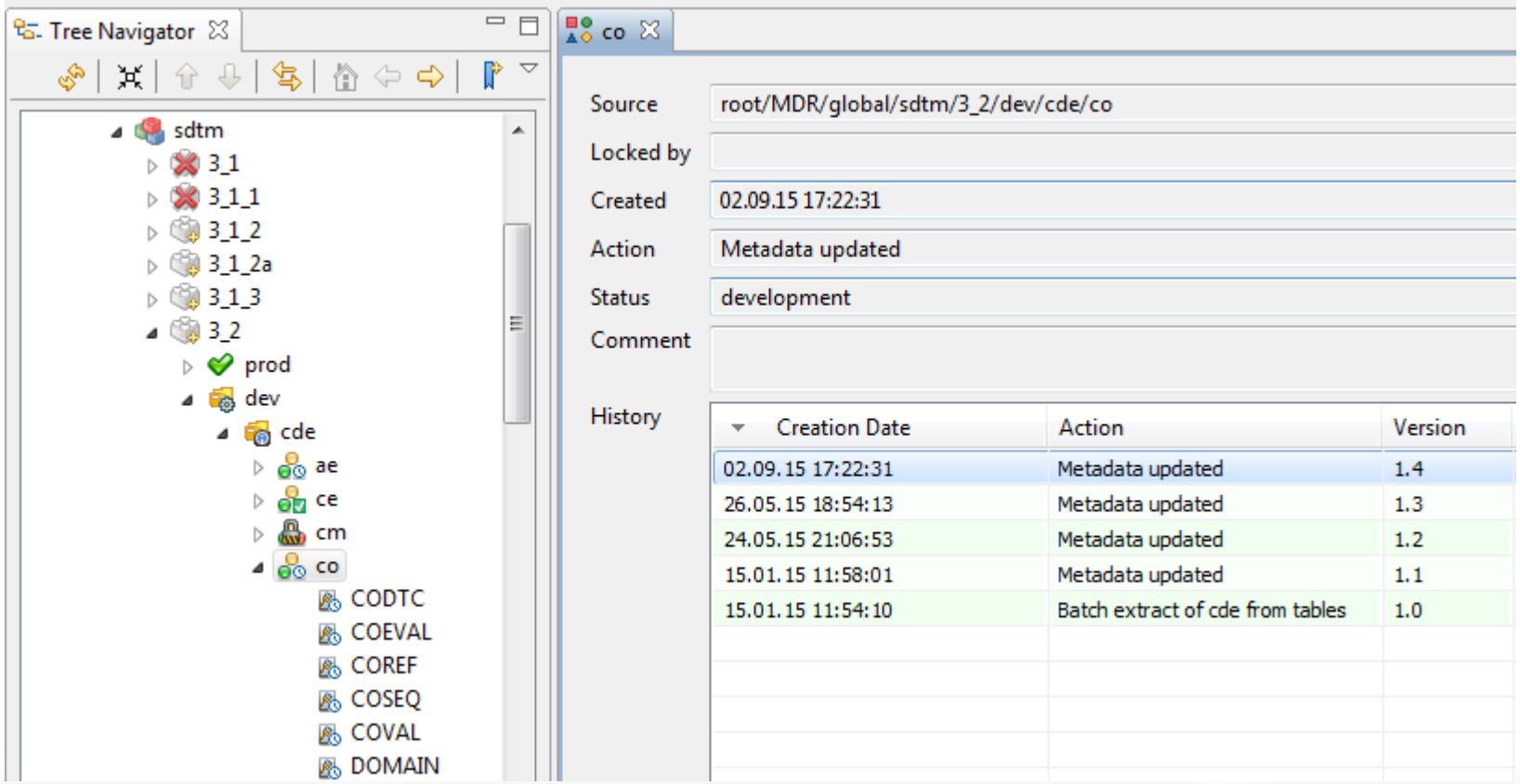
Classic version control tools are not prepared for pharma data

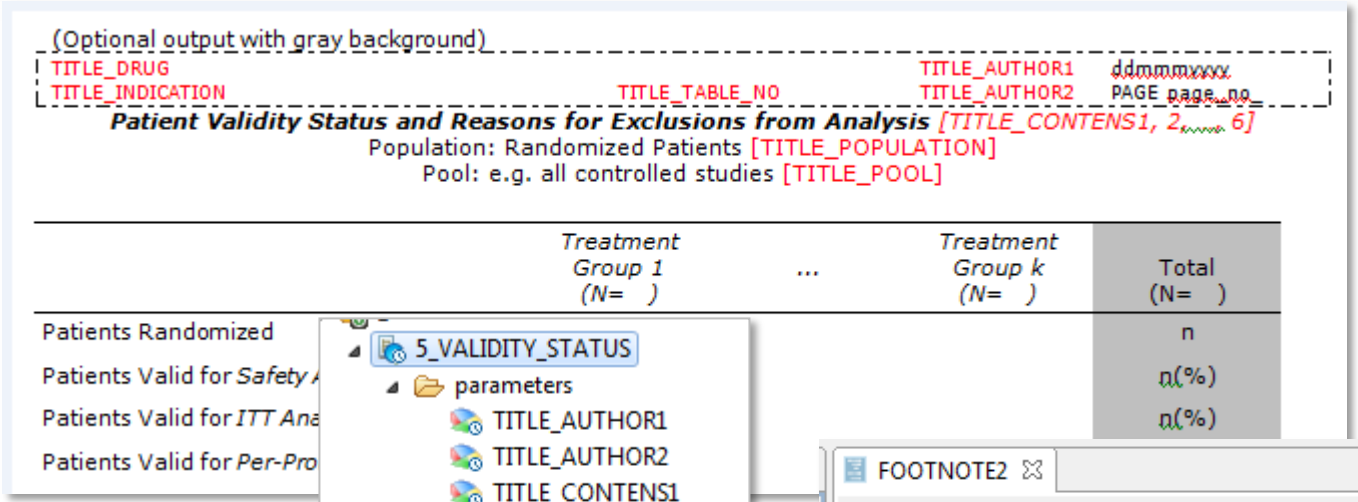
- **End-to-End Standards:**
  - Tabular models (e.g. CDASH, RAVE, SDTM, ADaM…)
  - Class- and concept-based models (e.g. BRIDG)
  - Non-tabular standards (e.g. TLF layouts, titles, footers)
  - Different metadata attributes per data model
  - Global, project, study levels

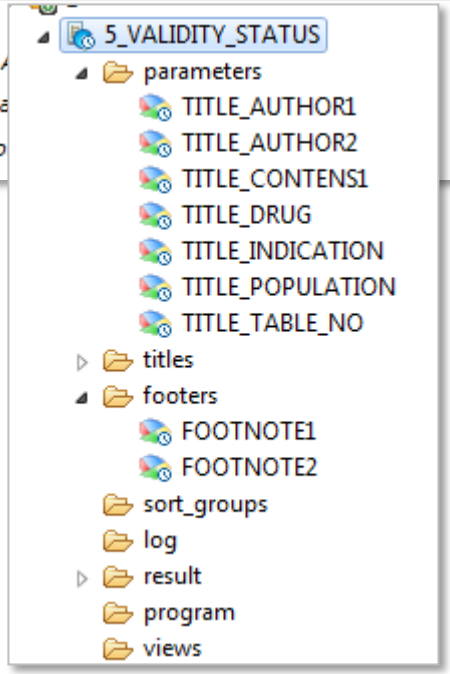
entimICE End-to-End Standards World

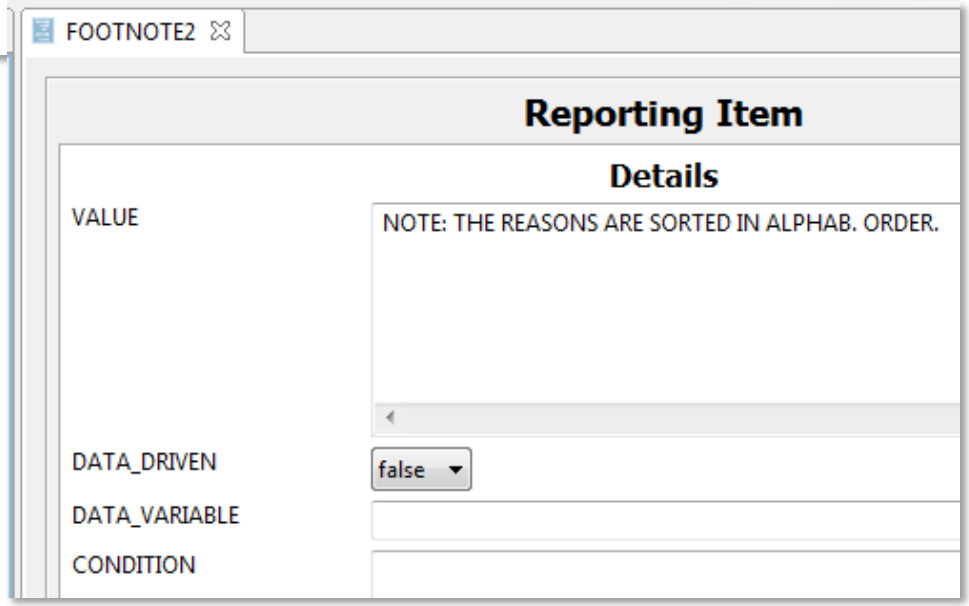


- **Challenges:**
  - Each level in hierarchical data models might have own versions in addition to model versions

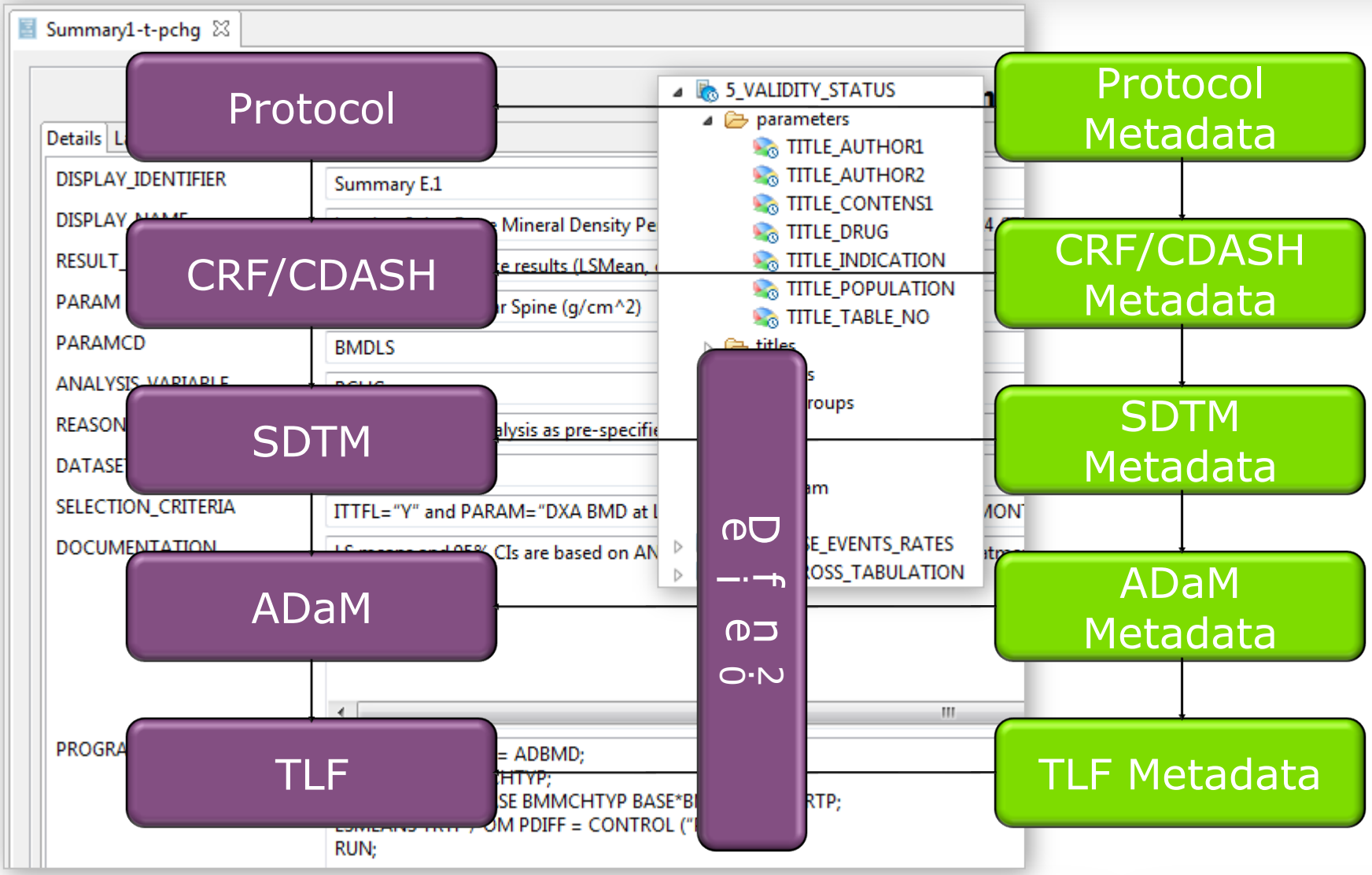



(Optional output with gray background)
TITLE_DRUG
TITLE_INDICATION
TITLE_TABLE_NO
TITLE_AUTHOR1
TITLE_AUTHOR2
ddmmmyyyy
PAGE page_no
Patient Validity Status and Reasons for Exclusions from Analysis [TITLE_CONTENS1, 2, ... 6]
Population: Randomized Patients [TITLE_POPULATION]
Pool: e.g. all controlled studies [TITLE_POOL]
Treatment Group 1 (N=)
...
Treatment Group k (N=)
Total (N=)
Patients Randomized
Patients Valid for Safety
Patients Valid for ITT Ana
Patients Valid for Per-Pro
n
n(%)
n(%)


5_VALIDITY_STATUS
parameters
TITLE_AUTHOR1
TITLE_AUTHOR2
TITLE_CONTENS1
TITLE_DRUG
TITLE_INDICATION
TITLE_POPULATION
TITLE_TABLE_NO
titles
footers
FOOTNOTE1
FOOTNOTE2
sort_groups
log
result
program
views

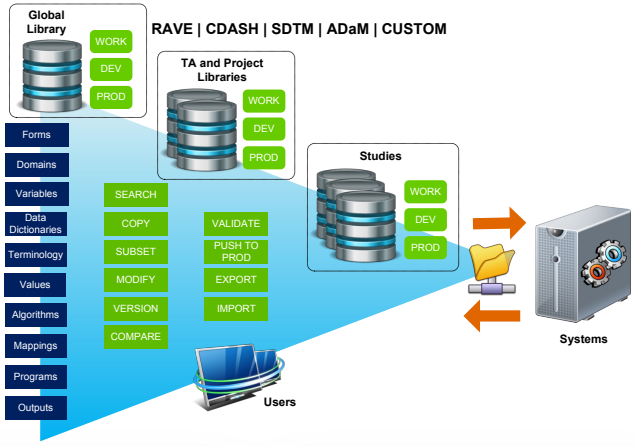

FOOTNOTE2
Reporting Item
Details
VALUE
NOTE: THE REASONS ARE SORTED IN ALPHAB. ORDER.
DATA_DRIVEN
false
DATA_VARIABLE
CONDITION

Metadata Dependency



- **Programming Artifacts:**
  - Program code versions
  - Run parameters
  - Macros and versions
  - Input data versions
  - Outputs (e.g. ADaM data, TLF)
  - Logs…
- **Challenges:**
  - Traceability and reproducibility requirement goes beyond source code control!

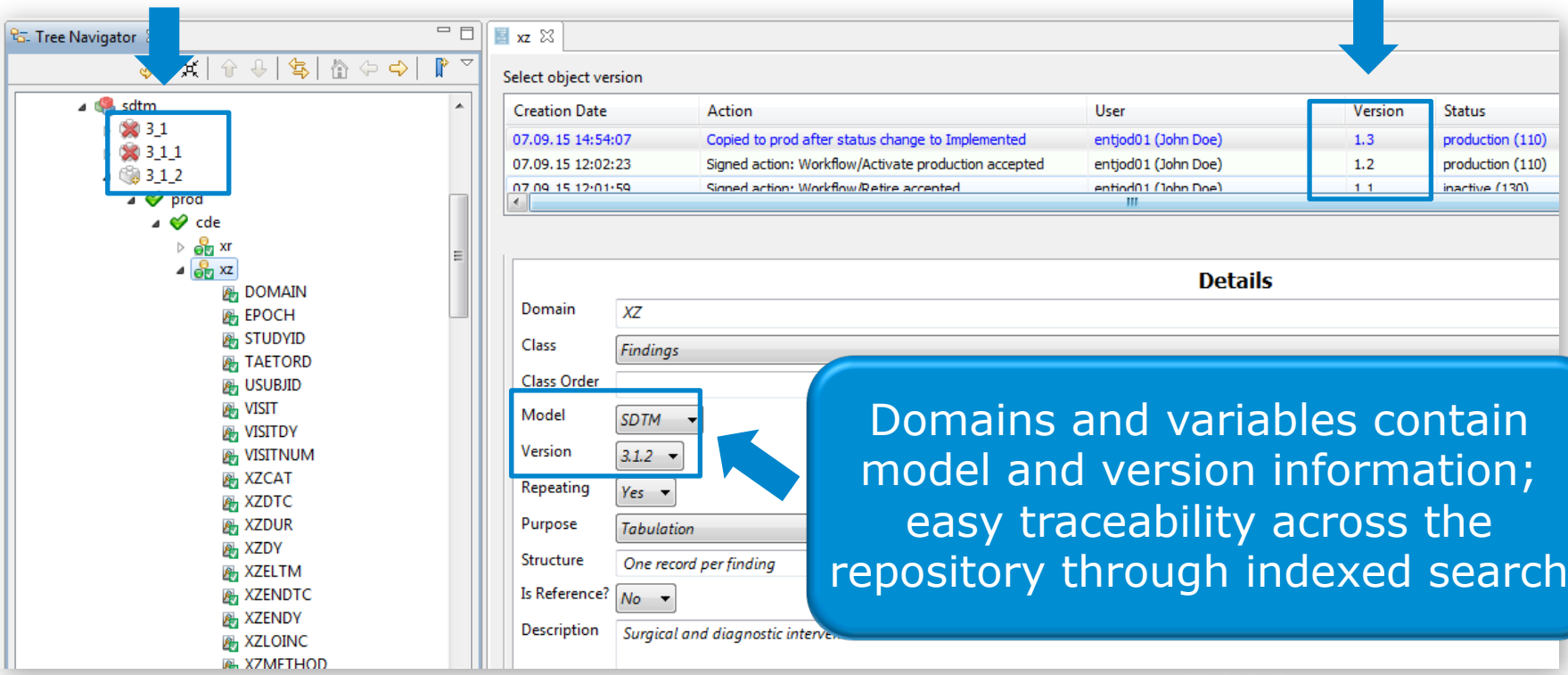
- **Differentiate Work, Val, Prod areas**
  - Supports parallel work without locks
  - No direct changes in prod – lifecycle goes always via val areas
  - Different access levels for work/val/prod
  - Commits – manual in val/work, auto in prod
  - Area locks when lifecycle is finished



- **Combine physical and logical versions!**

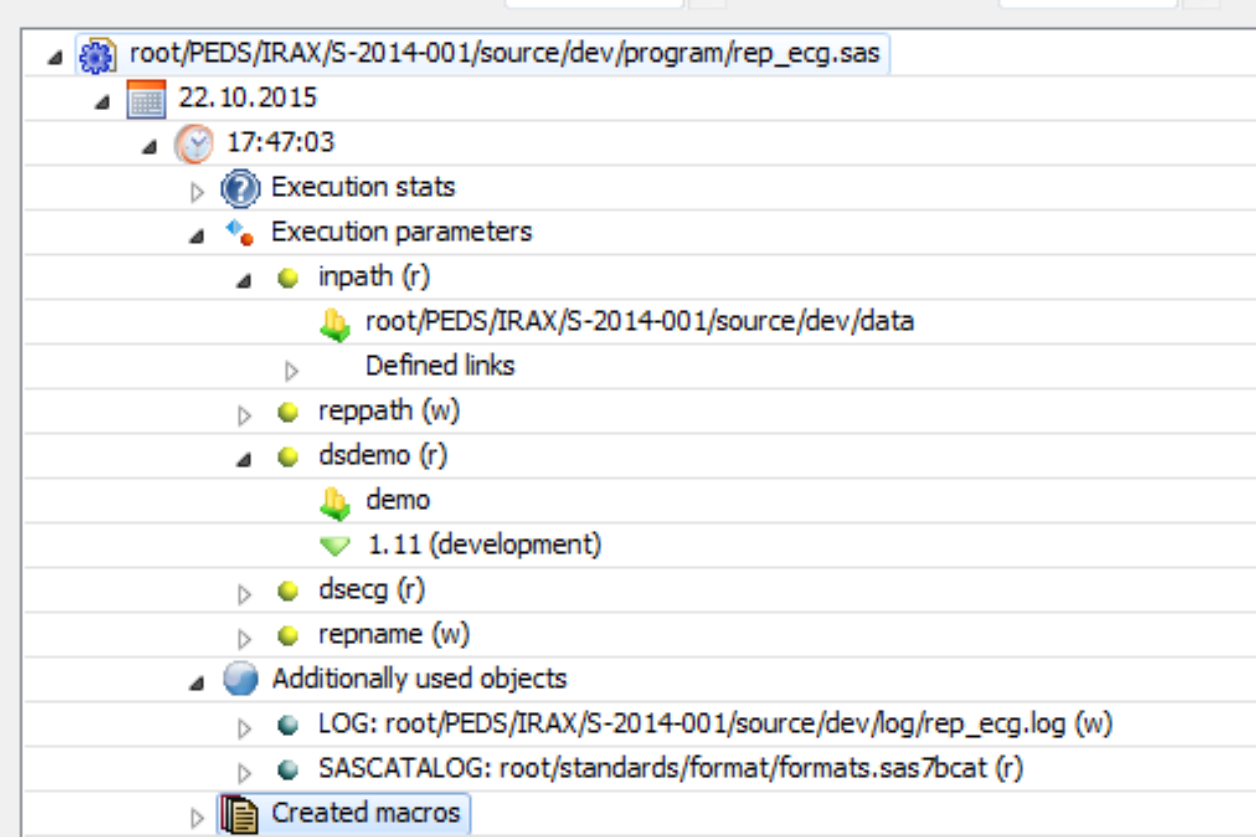
Logical versioning per data model with lifecycle

Physical versioning per object

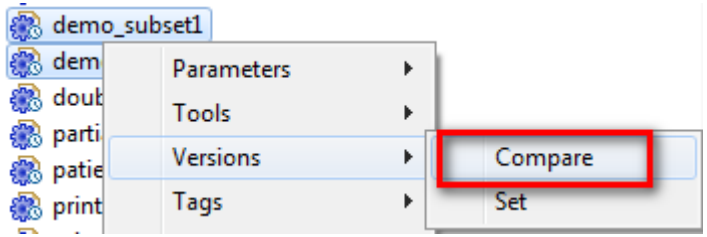


Domains and variables contain model and version information; easy traceability across the repository through indexed search

- **Combine versions and process tracing!**
  - Program execution metadata as evidence



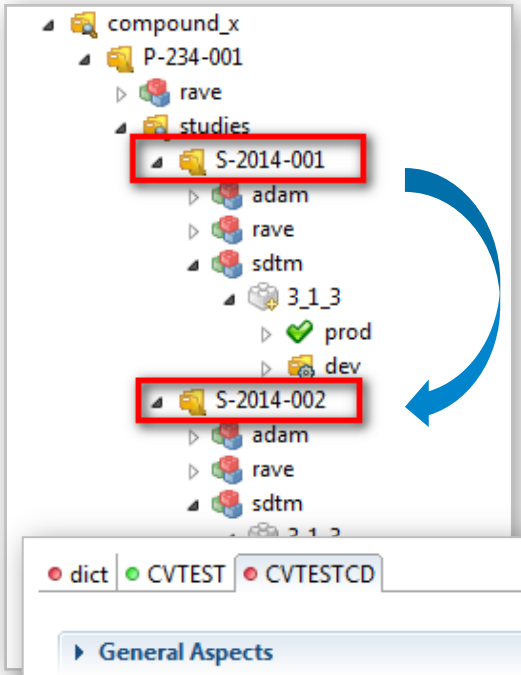
- **Show differences for line-based content …**




```
/*************************************************************
*************************************************************
***
*** Mapping/Pooling Program demo_subset1
***
*** Generated on 09DEC2013:16:30:12
*** Using version 1.2
***
*** Description:
***    Filter patients with id<20
***
*** Consists of steps:
***    1: step1 (Quick-Join Task)
***
*** Uses external datasets:
***    S.&S_DEMO
***    S.&S_DEMO_OUT
***
*************************************************************
***********

/*COMMENT1*/

%macro poolexec;
```

```
/*************************************************************
*************************************************************
***
*** Mapping/Pooling Program demo_subset2
***
*** Generated on 09DEC2013:16:31:45
*** Using version 1.1
***
*** Description:
***    age > 30
***
*** Consists of steps:
***    1: step1 (Quick-Join Task)
***
*** Uses external datasets:
***    S.&S_DEMO
***    S.&S_DEMO_OUT
***
*************************************************************
*************************************************************/

%macro poolexec;

OPTIONS nofmterr;
```

- **… and for binary content**



Use case: Compare all metadata in study S-2014-001 and in S-2014-001, similar compare global and study levels

Terminology

dict | CVTEST | CVTESTCD

General Aspects

No Severe Problems

Compare Keys and Attributes Source vs. Compare

Problem_Category	Result	Problem
CDISC_Synonyms (Source) eq CDISC_Synonyms (Co...	ERROR	Source key(s)[code = C119245] values [Source: Cardiac Index XXX Compare: Cardiac Index]
Length (Source) eq Length (Compare) VIOLATED	ERROR	Source key(s)[code = C119245] values [Source: 3 Compare: 8]

- **Search capabilities shall include version-related information!**
- **Indexed Content:**
  All textual Files (text and log files, also SAS programs...)
  PDF files, Excel sheets, XML files...
  SAS datasets, ClinRep datasets, CSV files as tables...
  entimICE object metadata (e.g. variables, study metadata)
  Version information and links
  Process metadata (e.g. mappings and exec traces)
- **Search examples:**
  fda AND cdisc
  node #version in (1.7, 1.10)
  node #version_history in (1.7, 1.10)
  node #ext = sas7bdat AND oncology
  node #mod (today –5 weeks, today -4 weeks)

- Classic version control tools and concepts are not sufficient for pharma processes!
- **Enhanced version control:**
  - Copy – modify – merge - lock
  - Management of logical and physical versions in parallel required
  - Version control for files and non-file content
- **Enhanced tools:**
  - Tagging for all content types
  - Diff tools for all content types
  - Links, dependencies and process tracing
  - Indexed search for all content types

Manageable programming process

Questions and Answers