GPS: Our Version Controlled Environment – Pros & Cons

Kamila Duniec, Principal Programmer, Novartis

Nigel Montgomery, Senior Principal Programmer, Novartis

Presentation Outline

1. GPS – Global Programming and Statistical environment
2. ClearCase – version control system
3. Pros and Cons of ClearCase/GPS and our approach to versioning
4. Summary
5. Industry trends

Novartis SCE - GPS

Global Programming and Statistical Environment (GPS)

- GPS is Novartis statistical computing environment (SCE), which is used by :

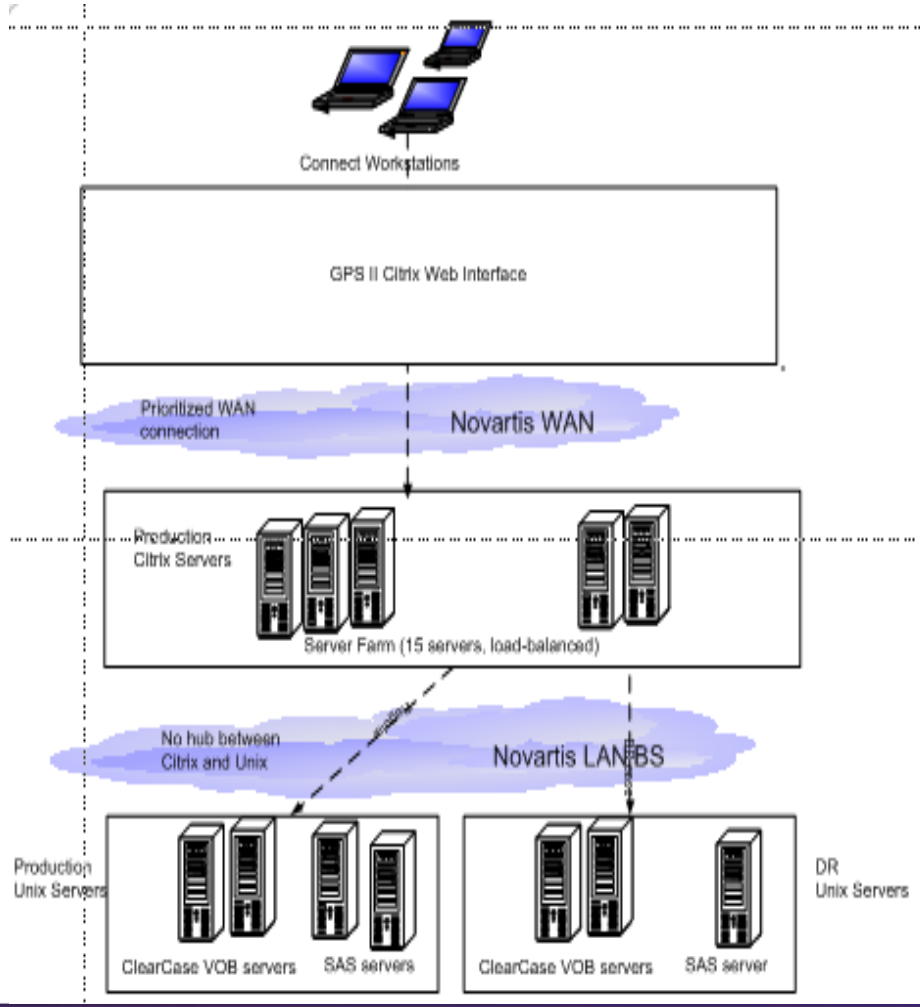
  Statistical Programmers and Statisticians, as well as users from Data Management, Health Economics, Marketing, etc.

- Current version of GPS dates back to 2004
- At any one time, there are 400-600 people actively using GPS across more than 8 geographical locations
- GPS is an environment providing access to applications for data analytics, coding, file transfer and Microsoft Office

Novartis SCE - GPS

Infrastructure and basics

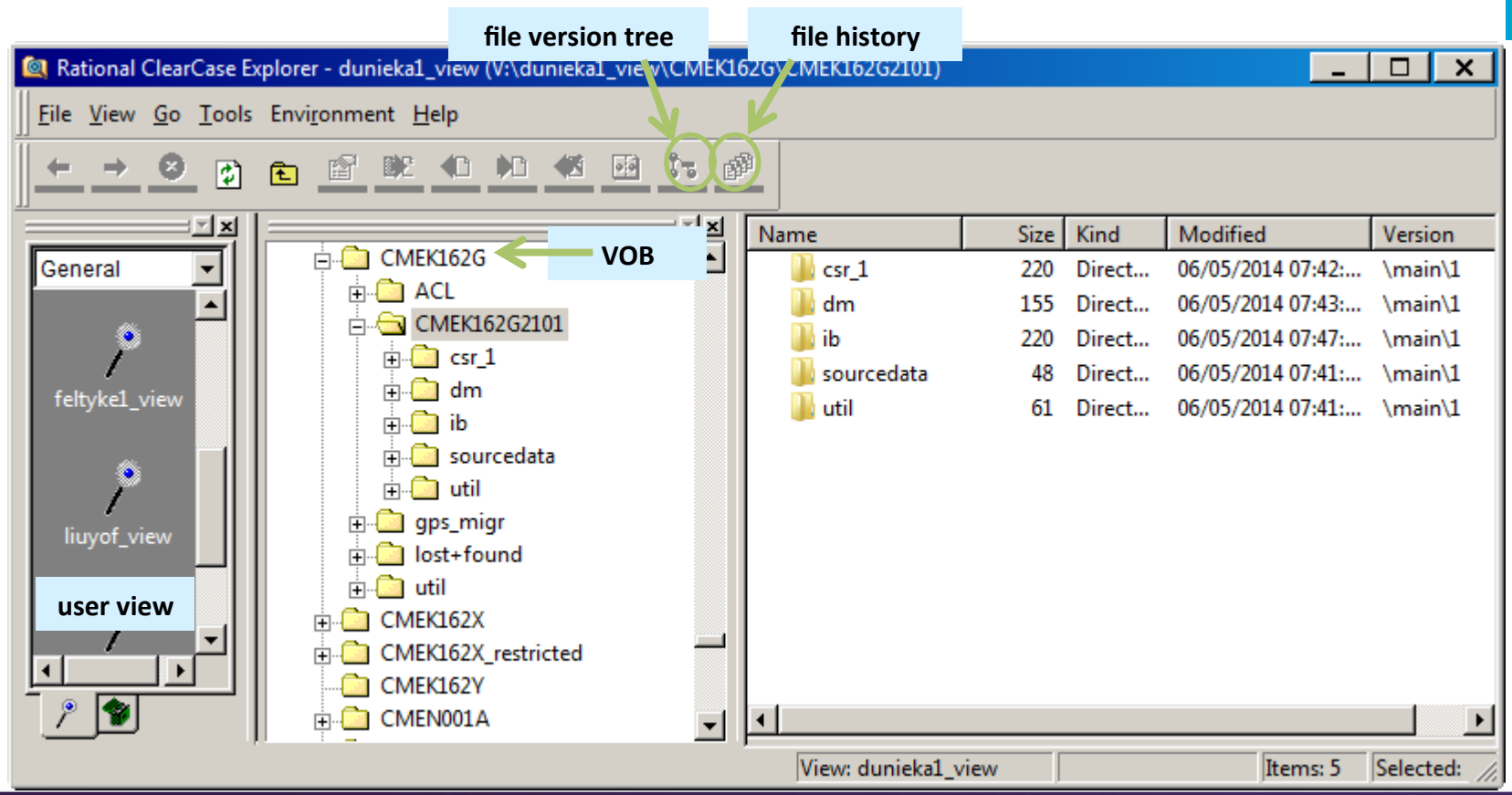
- GPS servers are centrally based in Basel and accessible from all sites through the Citrix Web interface
- Analysis is run in SAS 9.4 on Unix
- During development programs are executed in interactive SAS or via GPS MAKE Facility (GMF) on Clearcase
- For milestone runs (e.g. CSR, DMC) batch GMF script is executed
- Security/access controls are based on Unix groups and Access Control Lists



Novartis SCE - GPS

Version Control

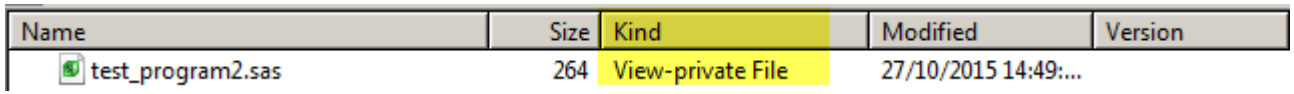
Project directory setup, version control, and tracking are done in Rational Clearcase



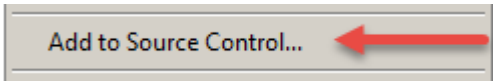
Novartis SCE - GPS

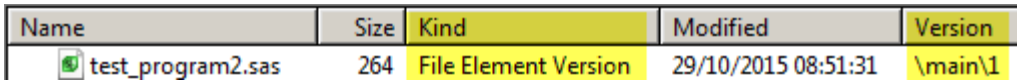
Version Control

1. Programs and outputs are initially developed in user's **private view**

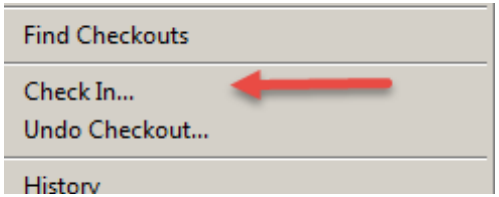


2. Once files are added to **Source Control** they become visible in other user views and are permanently stored





3. Files can be modified by checking out the file, modifying, and then checking back in



ClearCase History

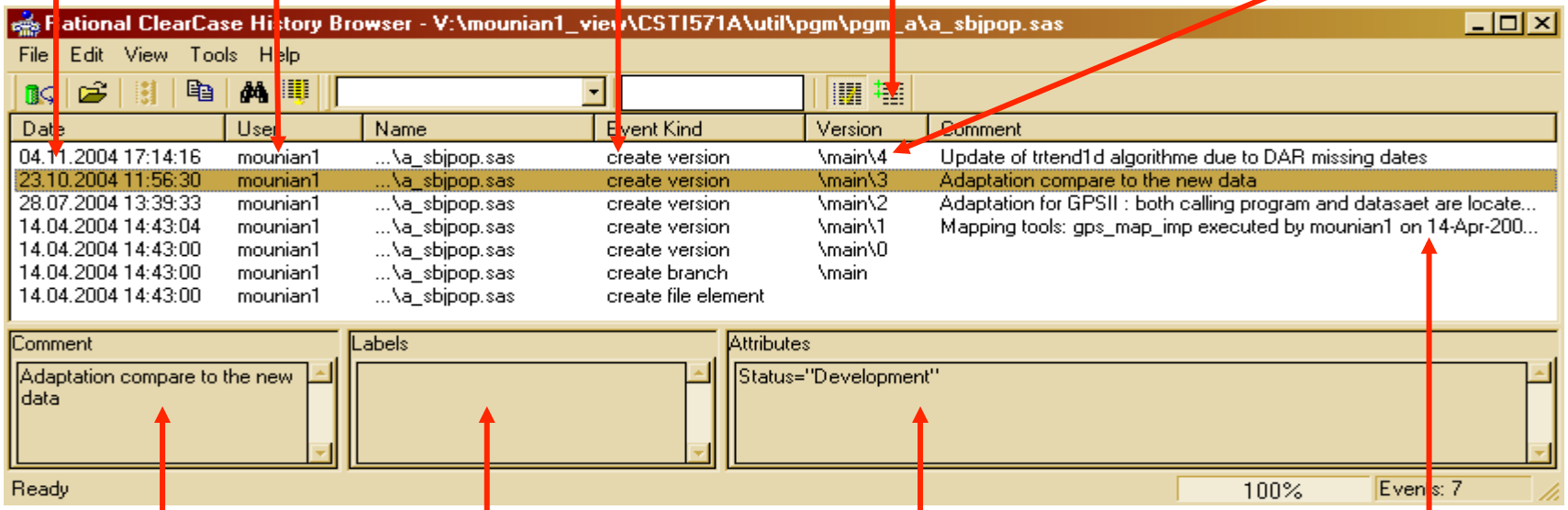
Date of change

Who made the change

What the change was

Click here to see minor events

Which version was created



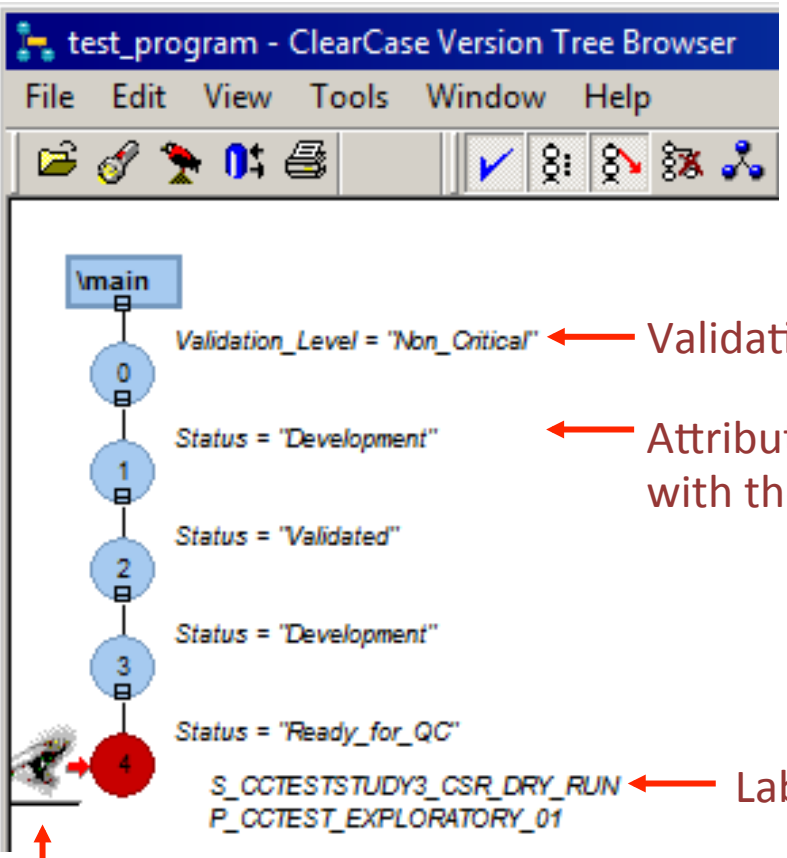
Comments about the change

Any label created

Attribute associated with the version

Comments about the change

ClearCase Version Tree



Validation level applied

Attribute associated with the version

Label applied

Which version is currently seen

- Comments about the change
- Date of change
- Who made the change

```
\main\4
   View selected version
   Created By : Duniec
   Created On : 17 June 2015 13:56:13
Description : updated after demotion from Validated
=== Ready_for_qc_comment/17-Jun-15.14:02:11/dunieka1/Non_Critical/========
Promote to Ready_for_QC
=======================================================
```

Novartis SCE - GPS

Version Control – Pros & Cons

- Traceability (audit trail) & Reproducibility
  - Full information on history and status of each file
  - When it was modified, by whom and what was the change
  - Re-create the same outputs as created in the past
- Support and Maintenance
  - A big and complex system requires sufficient IT support
  - System maintenance and enhancements are hard

Novartis SCE - GPS

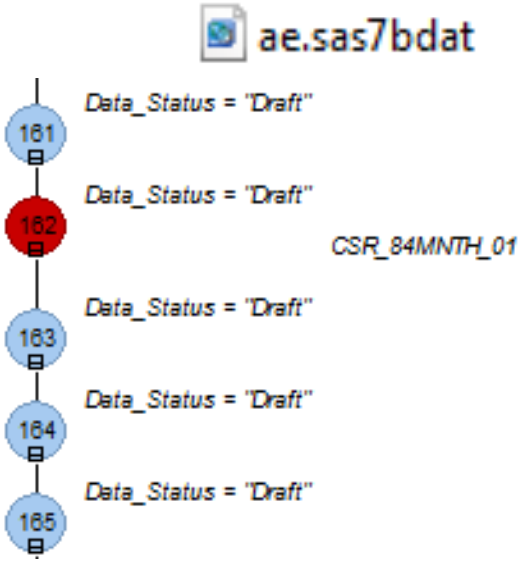
Version Control – Pros & Cons

- Availability of all files
  - All versions of each data set, all versions of each output, all versions of each program and other files are available in case investigations are needed, ad-hoc analysis is requested, which needs a historical version, etc.
- Performance & Stability
  - System performance has declined with a growing number of users and projects
  - Various problems (e.g. login issues) have lead to user perception of lower stability

Novartis SCE - GPS

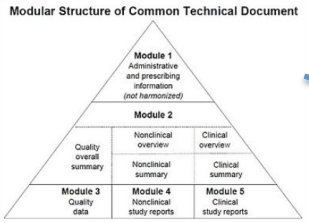
Version Control – Pros & Cons

- Point and Pick
  - Version selection by version number, label, date:time stamp
  - Rule applied to all files or a file type (e.g. *.sas)

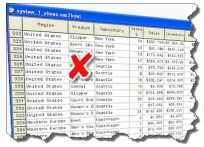


- Surgical re-runs

Surgical re-run example

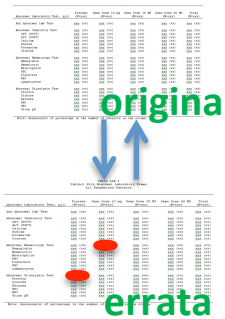


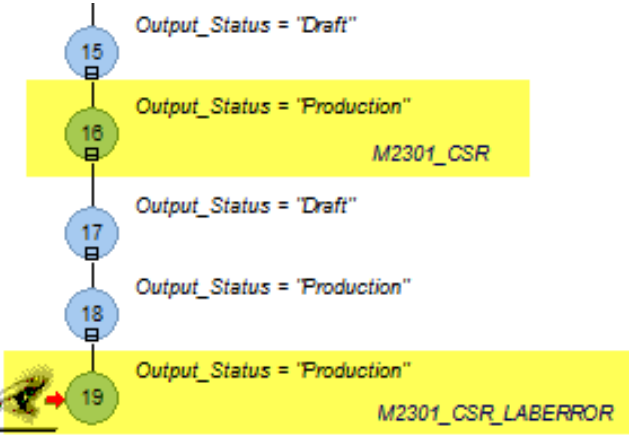
submission



error

compare and label





ClearCase allows us to re-generate only what is impacted by the change

Novartis SCE - GPS

Version Control – Pros & Cons

- Flexibility of the system leading to
  - Heavy business processes
  - Extensive training and onboarding needs

Summary

Pros	Cons
Traceability & Reproducibility	Support and Maintenance
Availability of all files	Performance & System stability
Point and pick & Surgical re-runs	Heavy business processes & Extensive training needs

Industry Trends

- PhUSE is driving an industry-wide discussion via the SCE project
  - working group: Emerging Trends & Technologies
  - project: Experiences with Statistical Computing Environments
- At Novartis we are building BRS/URS for our future system, and collaborating with our IT to look into HPCE technologies, etc.

Thank you to

- Attendees
- PhUSE
- Actelion
- GPS System Support
- GPS System Owner
- GPS Process Owner


phuse
The premier community for people working in the biometric area
@PhUSETwitta
PhUSE
www.phusewiki.org

GPS: Our Version Controlled Environment – Pros & Cons

Authors:
Kamila Duniec
Nigel Montgomery

Abstract

Global Programming and Statistical Environment (GPS) is the Novartis system providing tools for data analysis and reporting. There are 400-600 people using it at any time across more than 8 geographical locations.

GPS is accessible via Citrix and consists of ClearCase, SAS and other applications for code, data and output viewing and editing. Rational ClearCase is an IBM product, which provides a graphical user interface to file management, version control of the files, record of file history, as well as information on file dependencies.

At Novartis we are versioning all the files (SAS programs and macros, source and analysis data, and reports). The advantages of applying version control to all the files includes full traceability of every file under version control, which is related to GCP compliance (21 CFR Part 11), as well as ability to reproduce any version of any output from any version of a SAS program and any version of the data.

Such flexibility and capacity, however, come at a cost. With over 3000 GPS accounts and hundreds of projects we are experiencing performance issues. Furthermore, complexity of the system lead to a heavy business process, and a need for a greater training and support infrastructure and resources.

There is no industry standard with regards to the scope of what needs to be version controlled. At Novartis we are exploring the alternatives and system structures, which reduce the need to version control. Also, an industry-wide discussion has begun in order to understand the regulatory and business requirements for the most suitable reporting system.