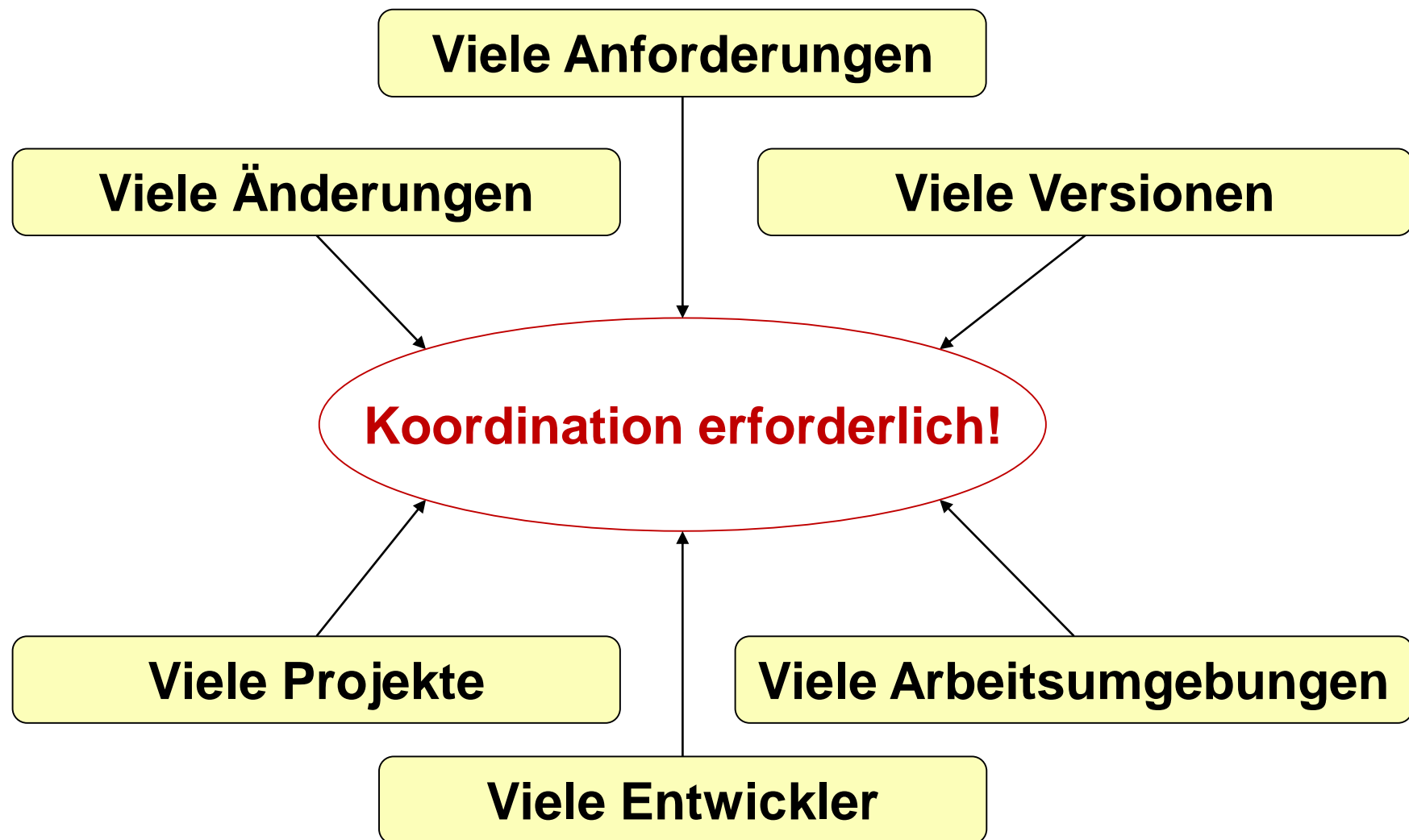


Kapitel 2

Software Configuration Management

Stand: 5.11.2015

Herausforderungen während der Softwareentwicklung



Warum Software Configuration Management?

Problem

- Software-Entwicklung ist nicht linear
 - ◆ Man macht Programmierfehler
 - ◆ Man trifft falsche Entwurfsentscheidungen
- Software-Entwicklung ist Teamarbeit
 - ◆ Sie und andere arbeiten parallel
 - ◆ ... auf vielen verschiedenen Dateien
- Verwaltung verschiedener Versionen
 - ◆ Kunde erhält stabile Version, während Entwicklung weitergeht
 - ◆ Fehlerkorrekturen müssen in alle Versionen integriert werden
 - ◆ Verschiedene Kunden erhalten verschiedene Varianten des Produkts

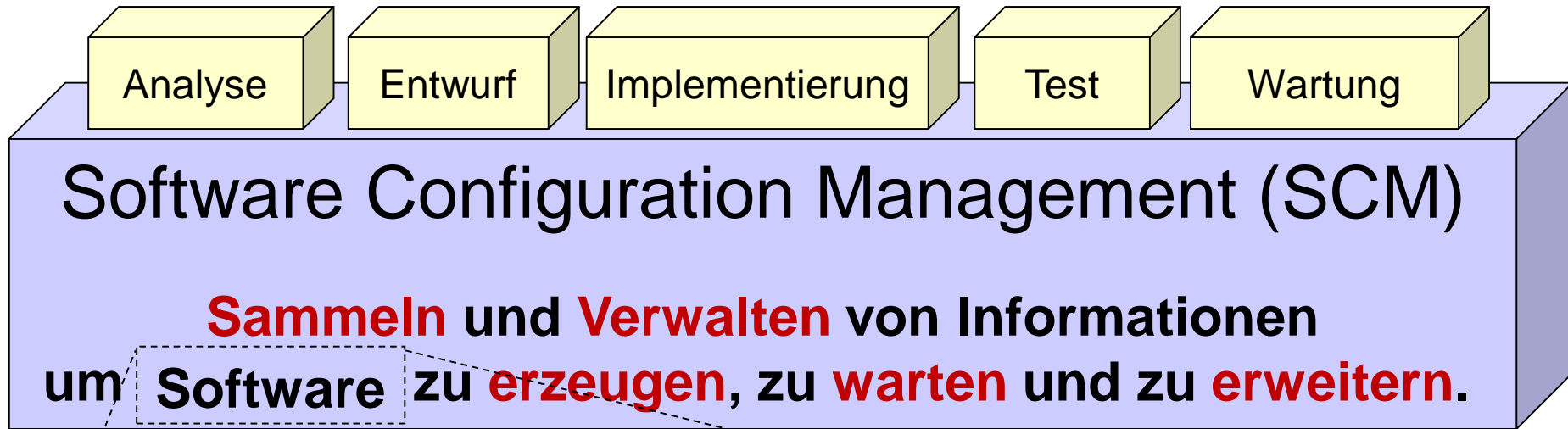
Sie möchten

→ wissen, was Sie wann warum getan haben
→ zu alter Version zurückgehen

→ wissen, wer was wann warum getan hat
→ Änderungen gemeinsam nutzen

→ parallele Entwicklung und Integration kontrollieren
→ Varianten kontrollieren

SCM: Das Fundament des Entwicklungsprozesses



- Quellcode
- Binärcode
- Build Skripte
- Konfigurationen
- Bitmaps & JPEGs
- HTML/XML Dateien
- CGI, Javascript
- CSS
- Anforderungen
- Entwürfe
- Test Skripte
- Dokumentation

SCM Funktionalitäten

- Verwaltet alle Komponenten eines Projekts in einem „Repository“
 - ◆ Keine redundanten Kopien
 - ◆ Sicher
- Macht Unterschiede zwischen Versionen sichtbar
 - ◆ „Was hat sich verglichen mit der gestrigen Version verändert?“
- Erlaubt Änderungen zu identifizieren, auszuwerten, zu diskutieren (!) und sie schließlich anzunehmen oder zu verwerfen.
- Verwaltet Metadaten: Wer hat was, wann, warum und wo getan?
 - ◆ „Wer hat was in Klasse X geändert?“
 - ◆ „Warum hat er es verändert?“
- Ermöglicht die Wiederherstellung voriger Zustände
 - ◆ Vollständig oder selektiv
- Erlaubt die Definition von Referenzversionen (Tags)
 - ◆ Release (Veröffentlichte Version)
 - ◆ Milestone (interne Version, Demo beim Kunden, ...)
 - ◆ Wichtiger Zwischenstand (fehlerfrei getestete Version, Stand vor ...)

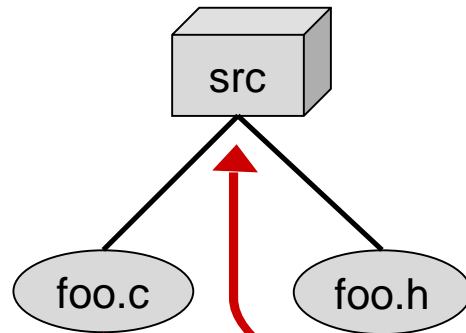
Grundlegende Ansätze

Vault Model
Virtual File System

Arbeitsumgebung und Repository

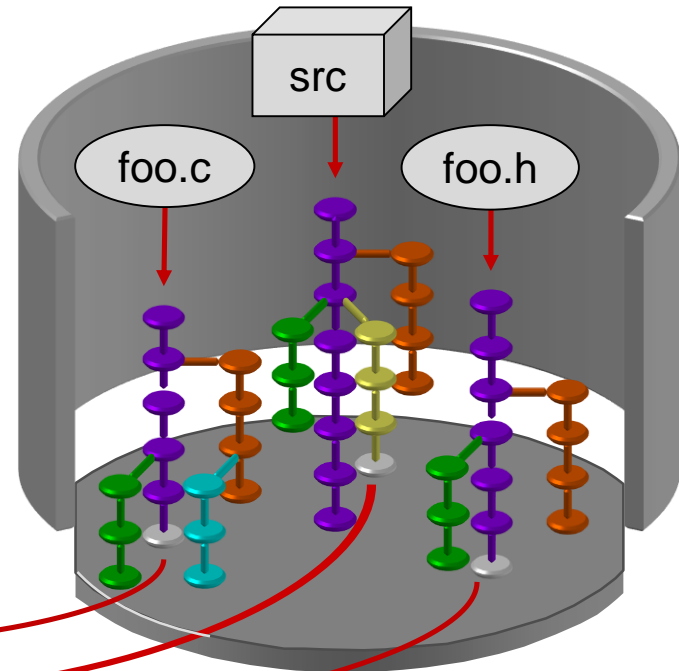
Arbeitsumgebung (“Sandbox”)

- Enthält nur die aktuell relevante Version eines jeden Artefaktes aus dem Repository
 - ◆ die Aktuellste
 - ◆ die letzte Funktionierende
 - ◆ ...

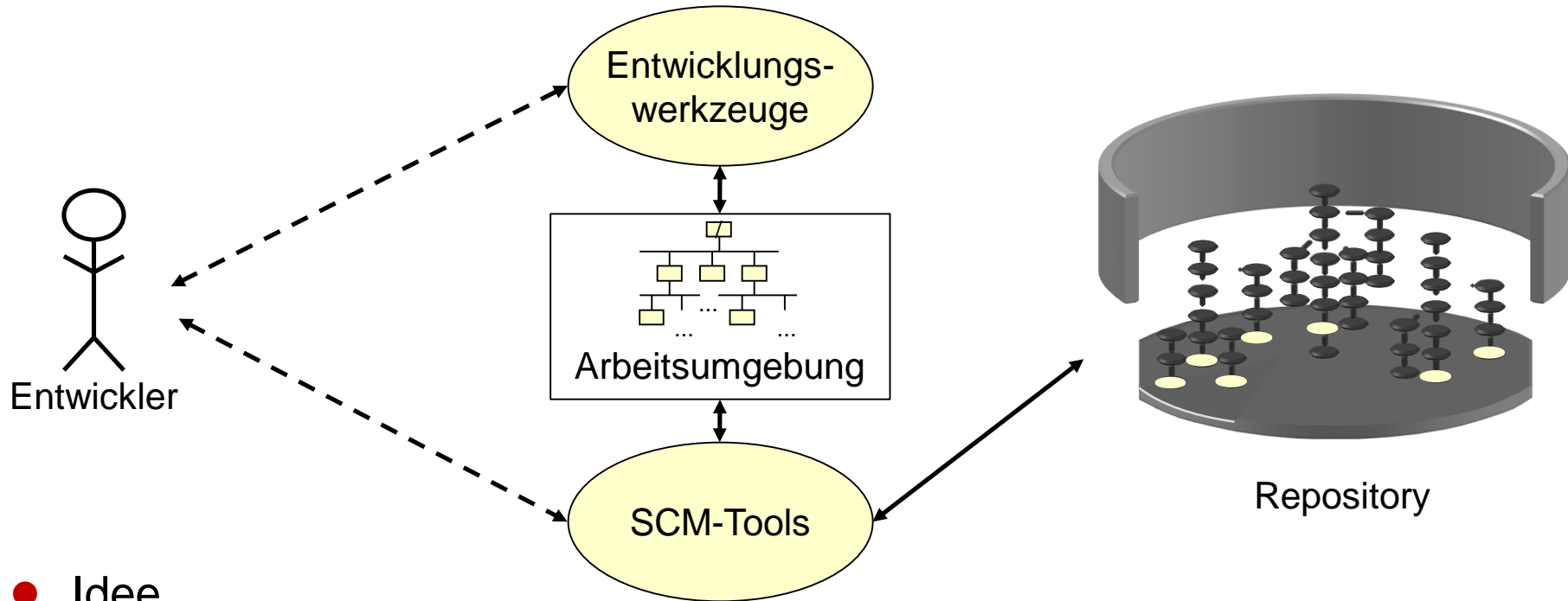


Repository

- Enthält alle Versionen (incl. “branches”) eines jeden Artefaktes, das unter SCM-Kontrolle stehen



SCM-Ansätze: „Vault Model“



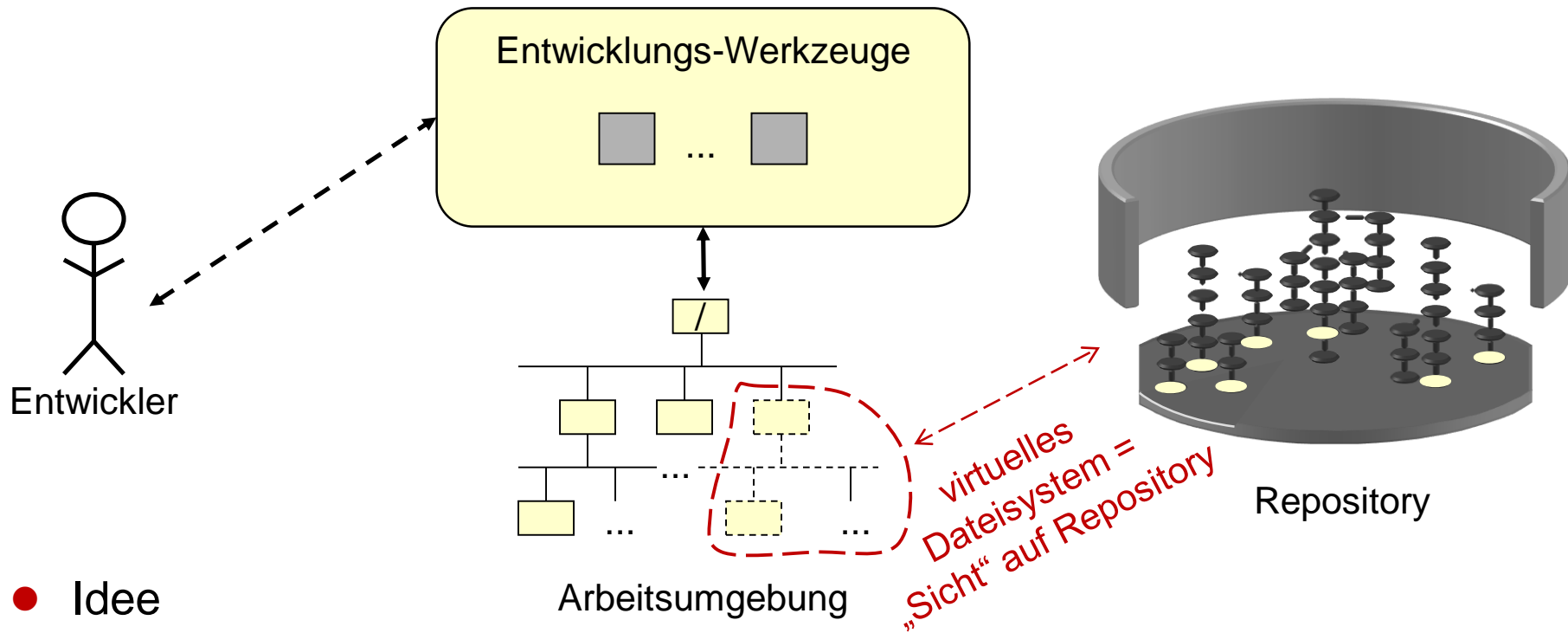
- Idee

- ◆ Dateien kopieren: Arbeitsumgebung $\leftarrow \rightarrow$ Repository

- Probleme

- ◆ Evtl. **viele private Kopien** in Arbeitsumgebungen außerhalb des Repository
- ◆ Entwickler arbeitet evtl. auf **veralteten** Dateien
- ◆ Möglicher Datenverlust durch **gleichzeitige** oder **abgebrochene** Updates

SCM-Ansätze: "Virtual File System" (VFS)



- Idee

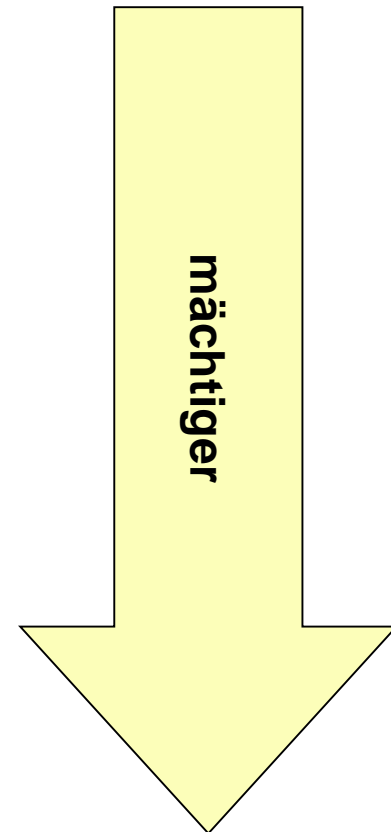
- ◆ Abfangen von I/O-Operationen des Betriebssystems (open, read, write) und Umleiten zum Repository

- Vorteile

- ◆ Transparenz, da das Repository als **normaler Verzeichnisbaum** erscheint
 - ⇒ nahtlose Integration bestehender Standardsoftware
 - ⇒ Benutzer kann wie gewohnt arbeiten

Verbreitete SCM-Tools die obige Ansätze umsetzen

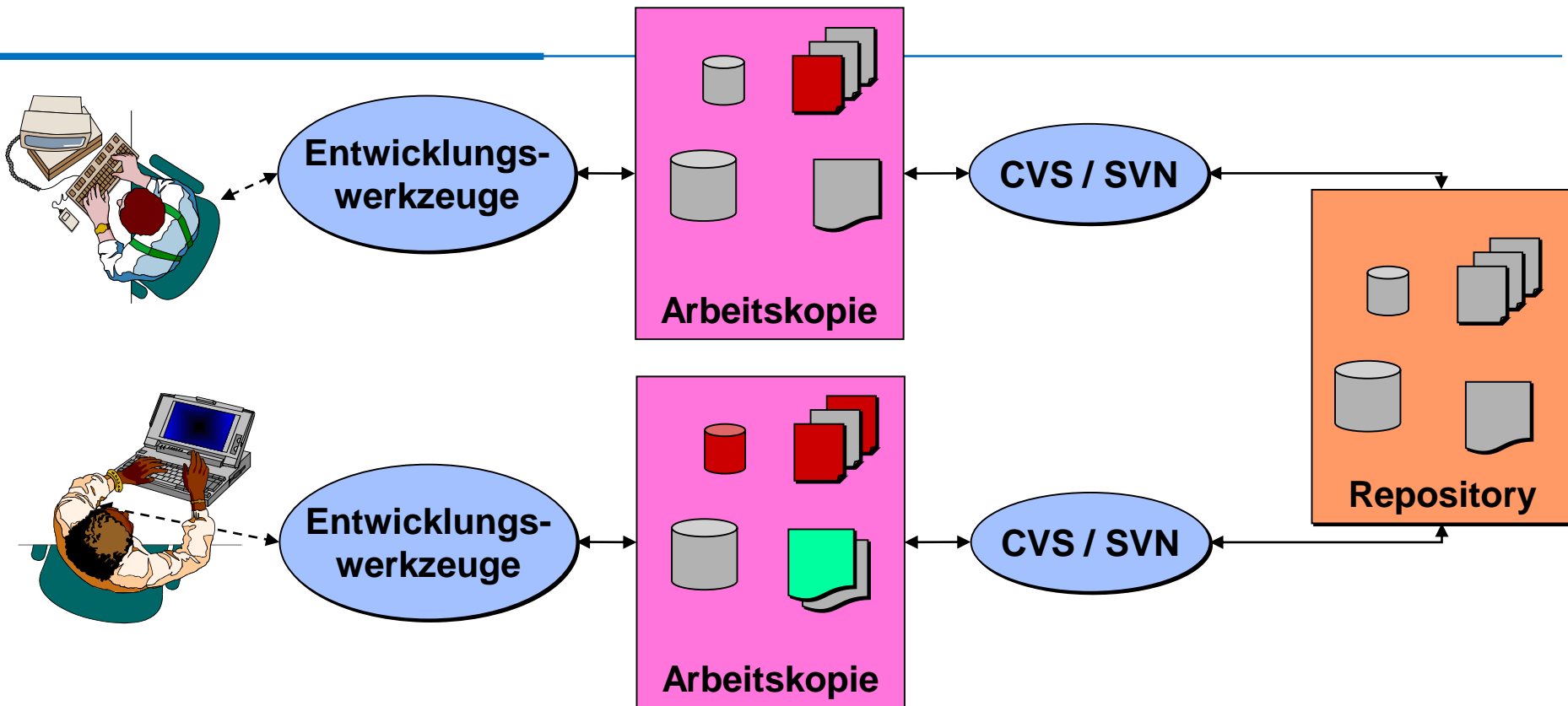
- ◆ CVS (simpel, kostenlos)
 - ⇒ **Repository = Dateisystem**
 - ⇒ Versionierung von Dateien
 - ⇒ Hauptsächlich Textdateien, Binärdateien nur eingeschränkt
- ◆ SVN (besser, kostenlos)
 - ⇒ **Repository = Datenbank** → Transaktionsunterstützung
 - ⇒ Versionierung von Dateien und Ordnern → Umbenennungen
 - ⇒ Text und Binärdateien
 - ⇒ Komfortable graphische Benutzeroberflächen
- ◆ ClearCase (high-end, kommerziell)
 - ⇒ **Virtuelles Dateisystem** basierend auf **Datenbank**
 - ⇒ Alle Eigenschaften von SVN plus...
 - ⇒ Dynamische Sichten auf Repository
 - ⇒ „Derived Object Sharing“
 - ⇒ Multiple Server, Replikation
 - ⇒ Prozessmodellierung



Arbeiten mit SVN (und CVS)

- Check-in und Check-out
- Commit und Update
- Synchronisierung
- Konfliktbeurteilung durch Dateivergleich
- Manuelle Konfliktauflösung

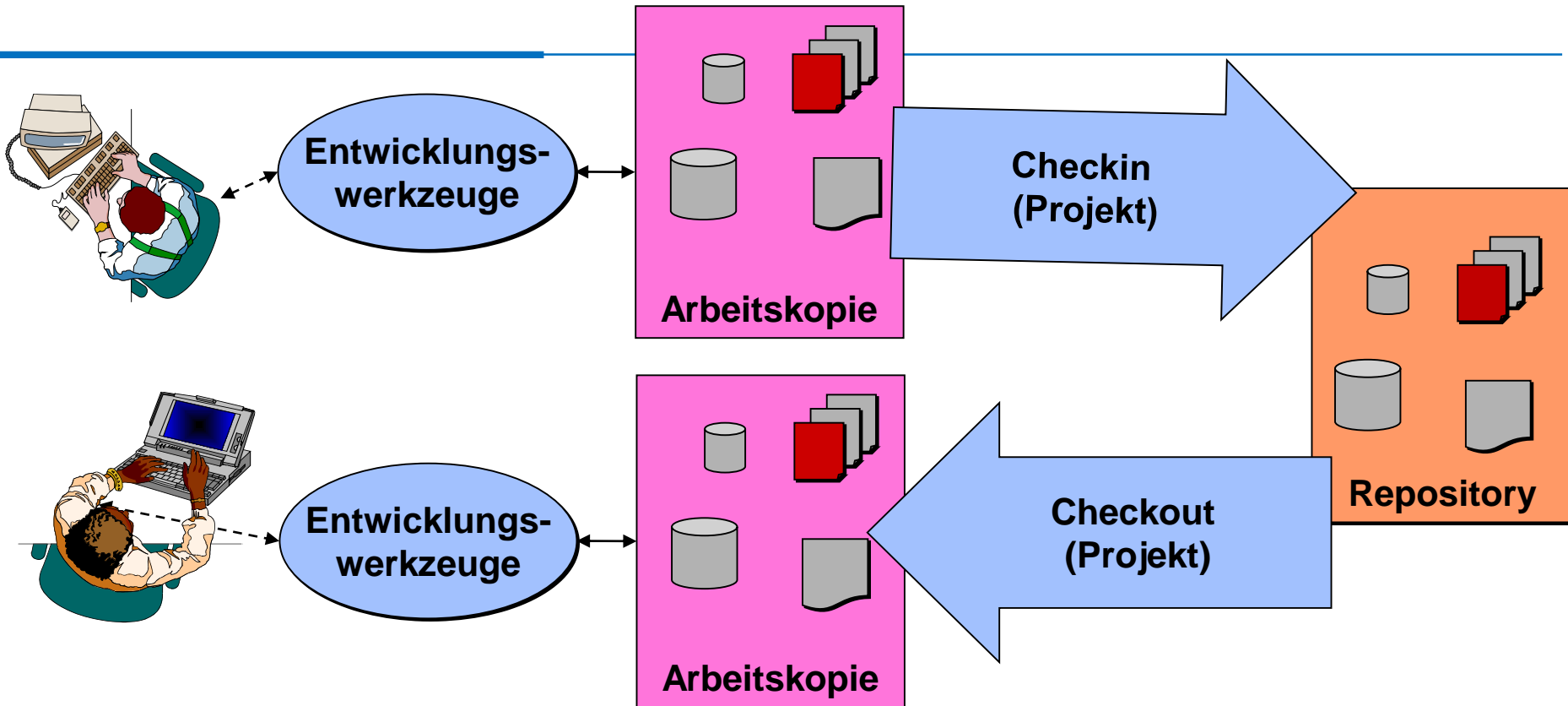
CVS und SVN: "Vault Model"



● Prinzip

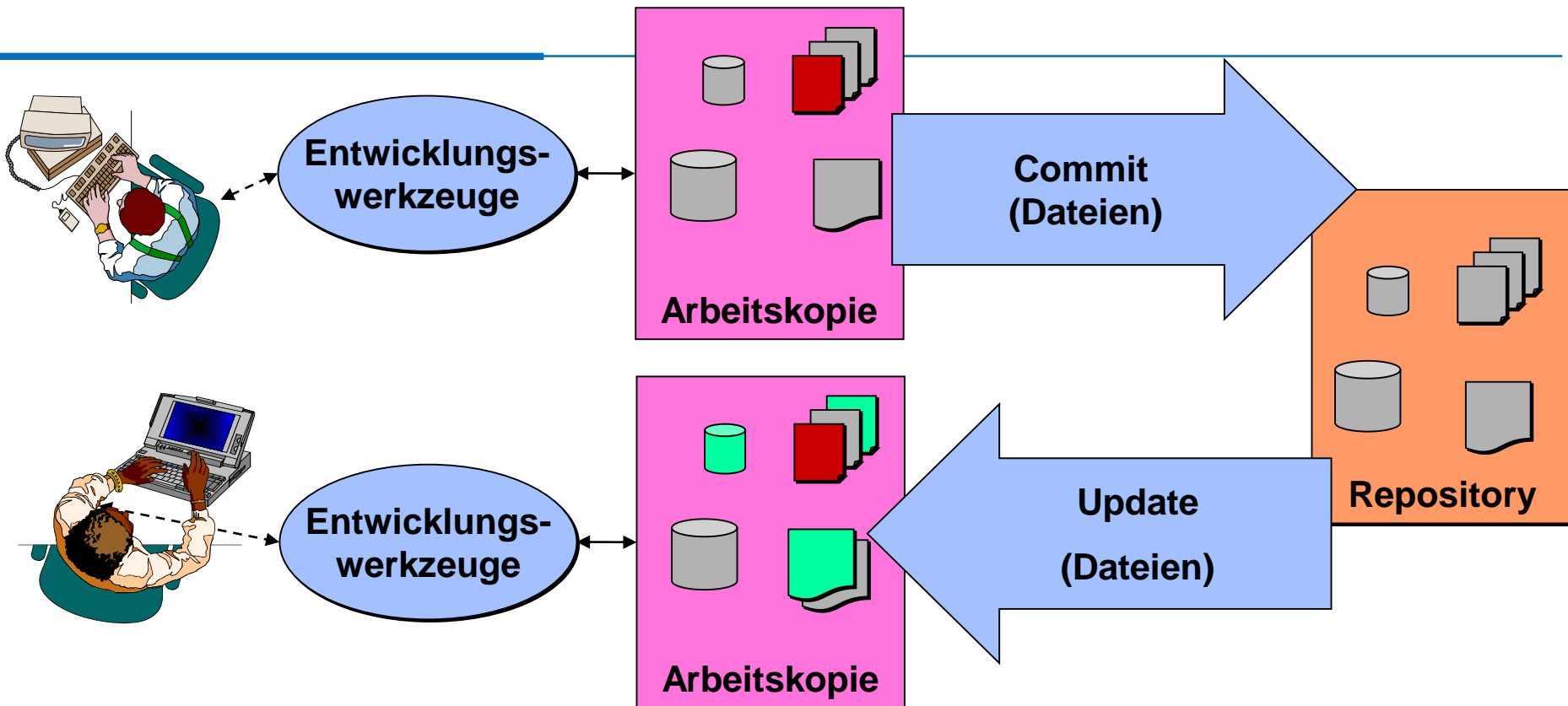
- ◆ Ein zentrales Sammelbecken („Repository“) aller relevanter Dateien
 - ⇒ Nur „offizielle“ Versionen
- ◆ Viele private Arbeitsumgebungen („Sandbox“) mit Kopien von Dateien
 - ⇒ Auch temporäre, inkonsistente, unfertige, ... Versionen

CVS und SVN: Checkin und Checkout



- Checkin
 - ◆ Fügt **Projekt** dem Repository hinzu
- Checkout
 - ◆ Erstellt eine Arbeitskopie des **Projekts** vom Repository

CVS und SVN: Commit und Update



- Commit

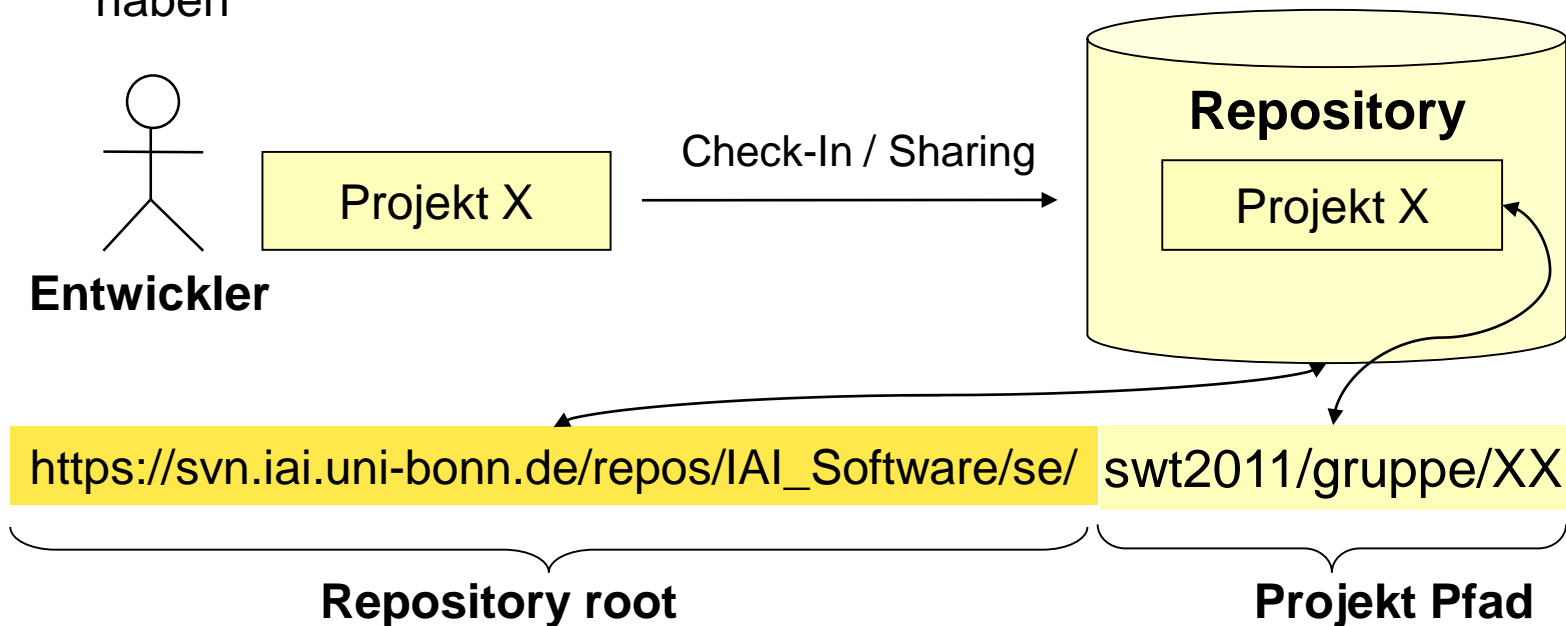
- ◆ Transferiert vom Programmierer geänderte **Dateien** in das Repository

- Update

- ◆ Transferiert geänderte **Dateien** vom Repository in die Arbeitskopie

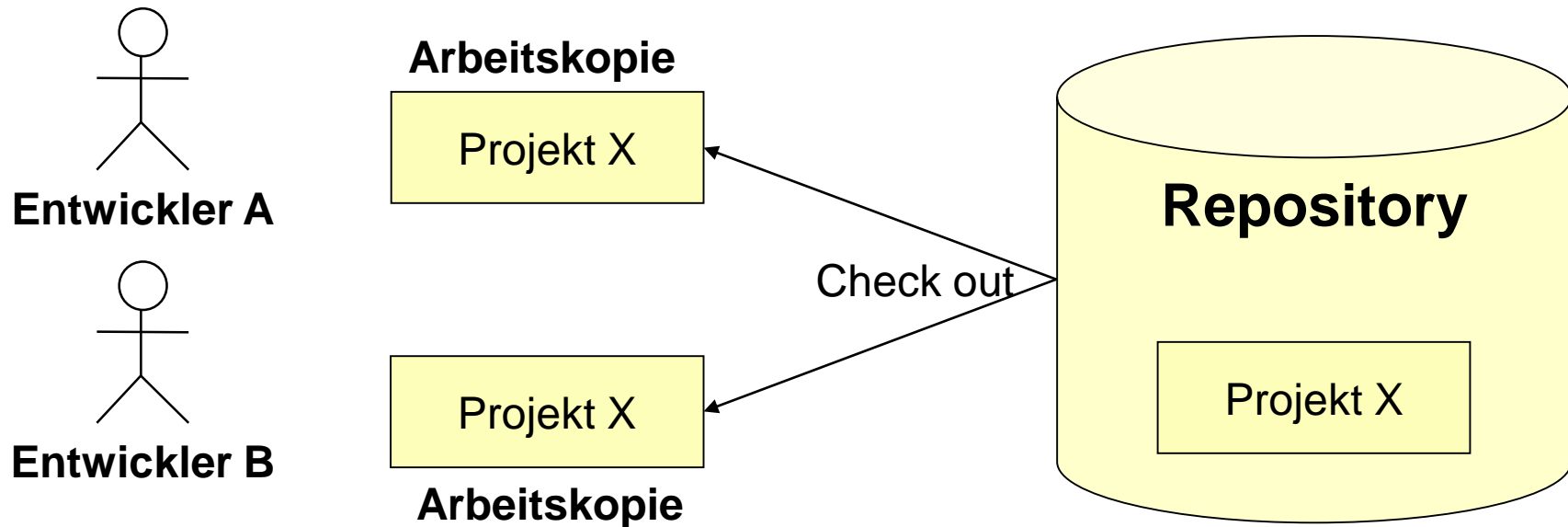
Check-In: Projekt unter Versionskontrolle stellen

- Ausgangssituation
 - ◆ Ein Repository existiert
 - ◆ Ein Projekt existiert, wird aber noch nicht gemeinsam genutzt.
- Check-In / Sharing / Promotion des Projektes
 - ◆ Das Projekt wird dem Repository hinzugefügt
 - ◆ Es ist nun für alle Entwickler verfügbar die Zugriff auf das Repository haben



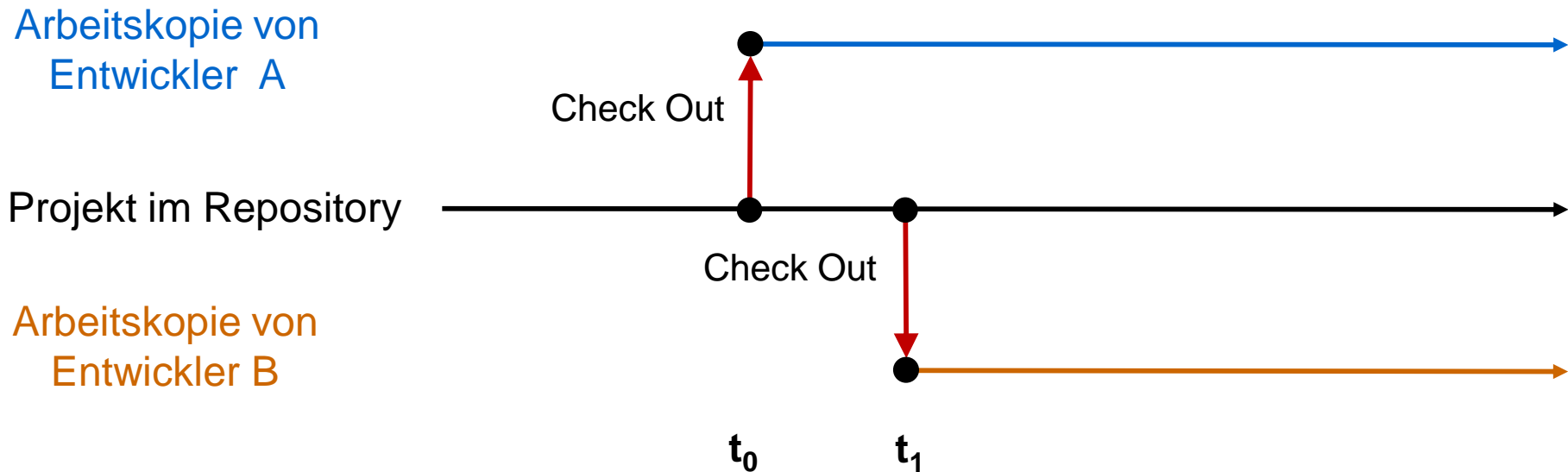
Check-Out: Initialer Projektdownload

- Check-out eines Projektes
 - ◆ Entwickler bekommen eine lokale Arbeitskopie
 - ◆ Von jetzt an können sie an dem Projekt arbeiten
- Auschecken wird nur einmal pro Projekt gemacht!
 - ◆ Neue Version aus Repository holen → siehe „Update“



Check-Out: Initialer Projektdownload

- Entwickler bekommt eine lokale Arbeitskopie eines Projektes
- Auschecken wird nur einmal pro Projekt gemacht!
 - ◆ Neue Version aus Repository holen → siehe „Update“



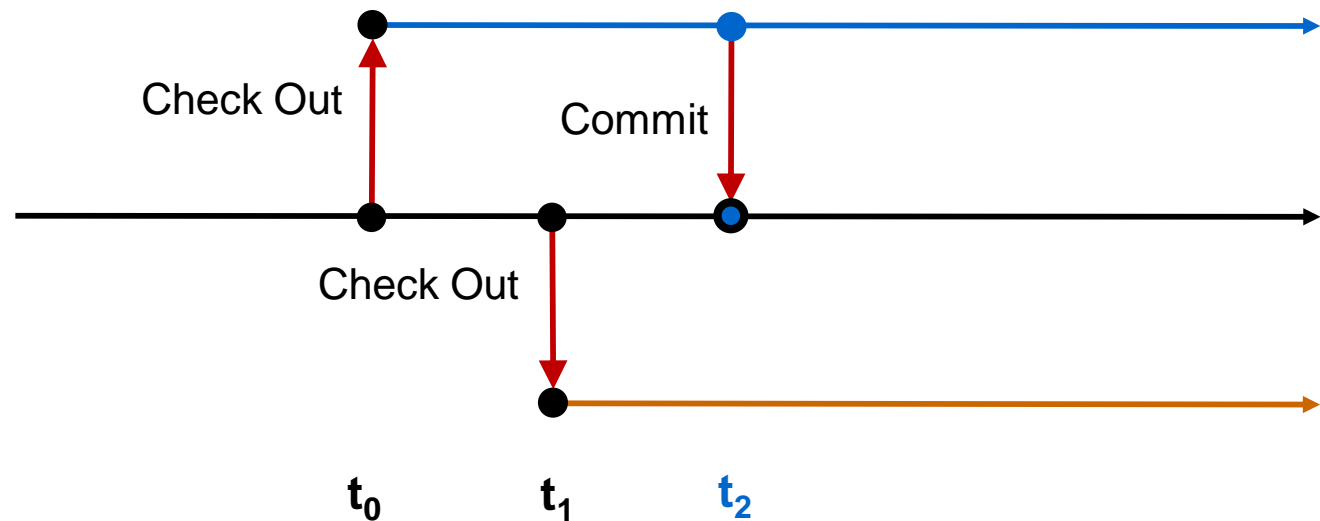
Commit: Änderungen in das Repository übertragen

- Ausgangslage: Entwickler A hat seine Arbeitskopie geändert
- Commit
 - ◆ Er fügt seine Änderungen dem Repository hinzu
 - ◆ Mit einem „Commit Kommentar“ teilt er dem Team mit was er warum geändert hat: „NullPointerException behoben, die auftrat wenn ...“

Arbeitskopie von
Entwickler A

Projekt im Repository

Arbeitskopie von
Entwickler B



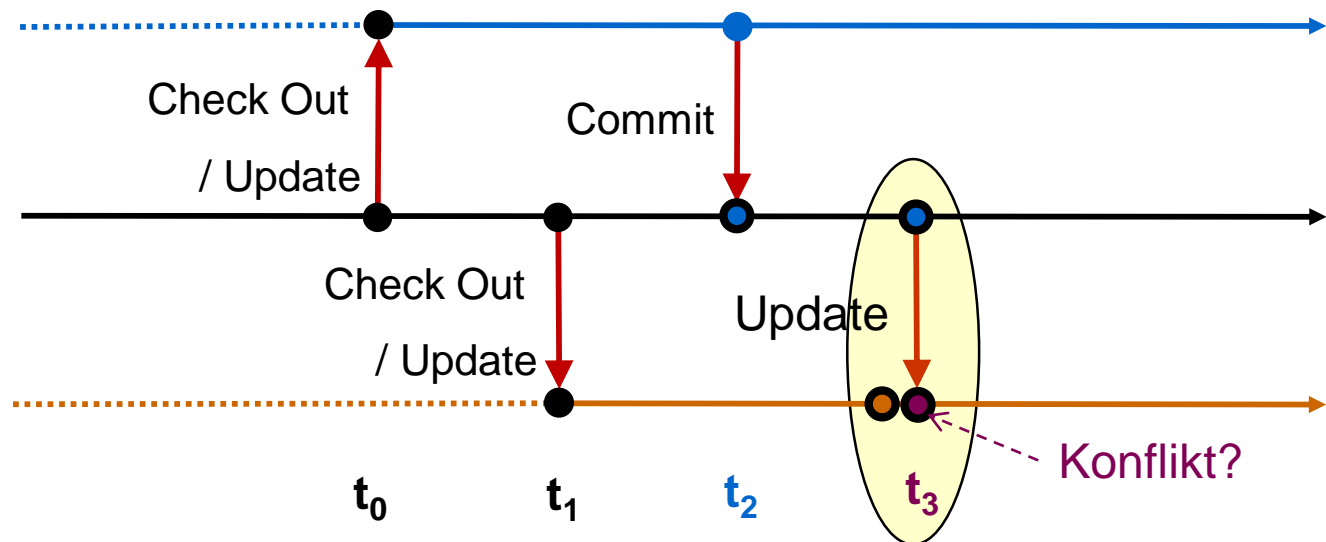
Update: Neueste Änderungen vom Repository übernehmen

- Ausgangslage: Repository hat sich geändert
 - ◆ Entwickler B ist sich **sicher(!)**, dass er die Änderungen übernehmen will
- Update
 - ◆ B aktualisiert seine Arbeitskopie mit dem aktuellen Zustand des Repository
- Problem
 - ◆ Bei Konflikten, wird automatisch ein „merge“ durchgeführt – evtl. unerwünscht
 - ◆ Besser: Zuerst „Synchronize“ benutzen und selbst entscheiden!

Arbeitskopie von
Entwickler A

Projekt im Repository

Arbeitskopie von
Entwickler B



Synchronisation: Versionsvergleich

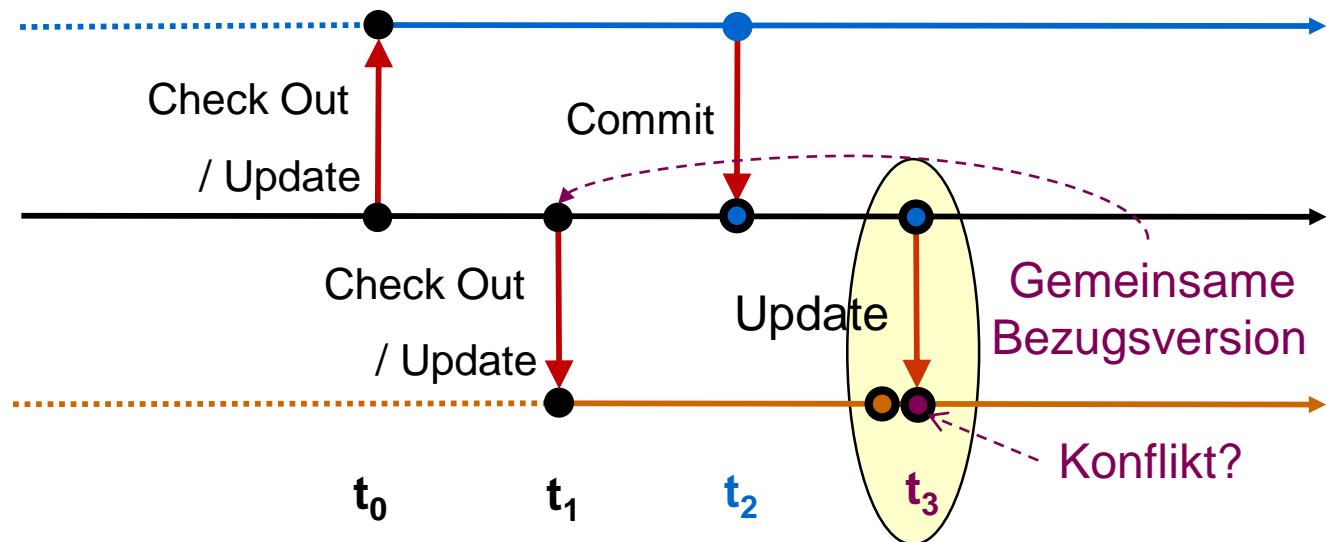
- Vergleich des Projektes in Arbeitskopie mit Repository bezogen auf Stand beim letzten Abgleich (check-in / commit / check-out / update)
 - ◆ Automatisierter Vergleich **aller Dateien** im Projekt
 - ◆ Automatisierter Vergleich **einzelner Dateiinhalte**
 - ◆ Der Entwickler entscheidet selbst was aktuell ist
 - ◆ ... und führt Updates oder Commits explizit durch (eventuell selektiv)



Arbeitskopie von
Entwickler A

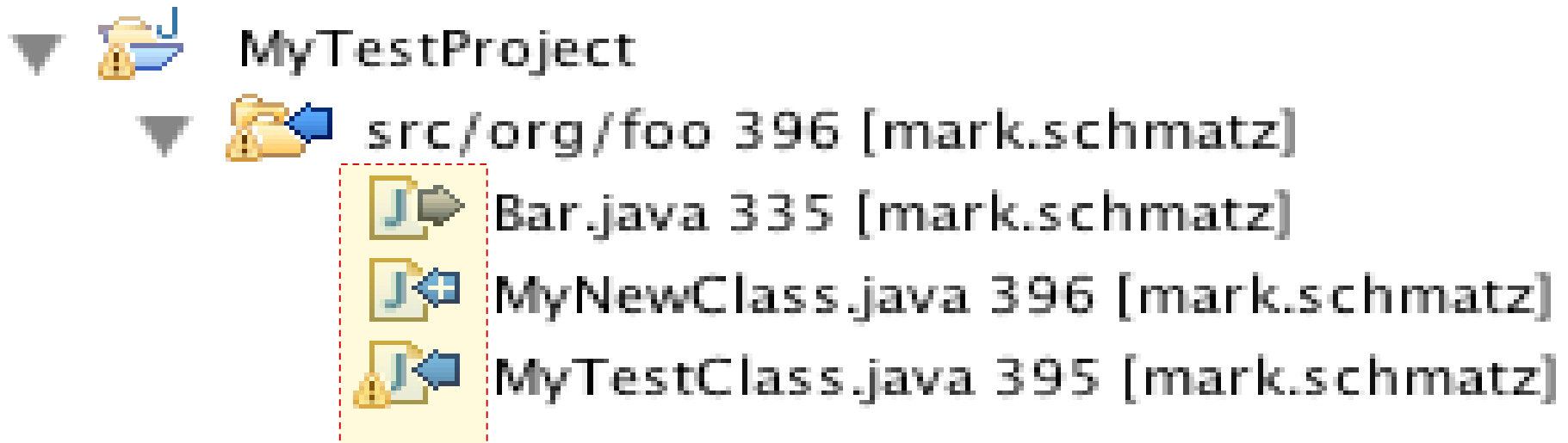
Projekt im Repository

Arbeitskopie von
Entwickler B

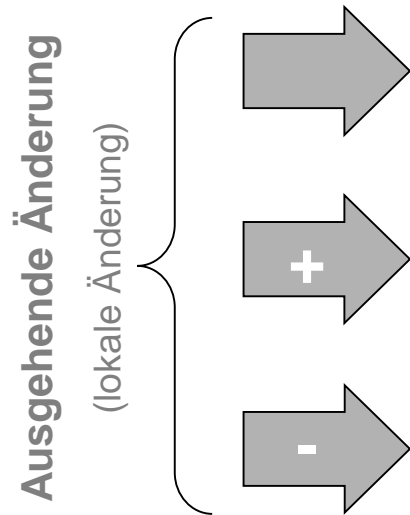


"Synchronize View": Gesamtübersicht

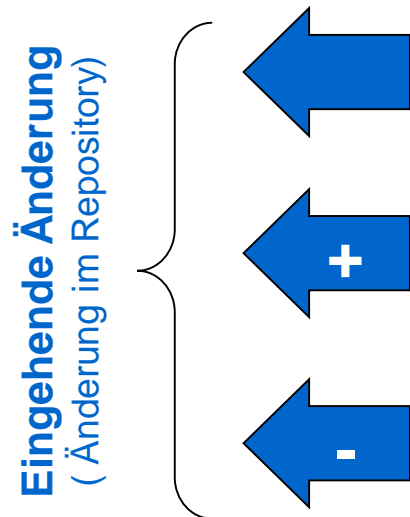
- Automatisierte Gesamtübersicht
 - ◆ Vergleich auf Projektebene: Alle Dateien des Projektes (oder des selektierten Ordners) werden verglichen
 - ◆ Zeigt pro Datei **ein- und ausgehende Änderungen sowie Konflikte** durch entsprechende Symbole



Symbole im „Synchronize View“



- Lokale Datei ist neuer als ihre Version im Repository
→ Überschreibe Version im Repository mit lokaler Version
- Neue lokale Datei existiert nicht im Repository
→ Füge Datei zum Repository hinzu
- Datei aus Repository wurde lokal gelöscht
→ Datei aus (nächster Version in) dem Repository löschen

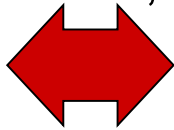


- Datei im Repository ist neuer als ihre lokale Version
← Überschreibe lokale Kopie mit der aus dem Repository
- Neue Datei aus Repository existiert nicht lokal
← Füge die Datei der lokalen Arbeitskopie hinzu
- Lokale Datei wurde im Repository gelöscht
← Lösche die Datei aus der lokalen Arbeitskopie

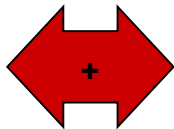
Bedeutung der Symbole im „Synchronize View“ (Fortsetzung)

- Die vorherige Folie stellt die Fälle dar, wenn eine Datei gegenüber dem Stand des letzten Abgleichs mit dem Repository nur auf einer Seite verändert wurde
 - ◆ Änderung nur lokal → Übernahme ins Repository (ausgehende Änderung)
 - ◆ Änderung nur in Repository → Übernahme in lokale Kopie (eingehende Änderung)
- Wenn eine Datei auf beiden Seiten neuer als der Stand beim letzten Abgleich ist, wird ein **Konflikt** gemeldet:

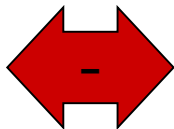
Konflikt



- Dateiinhalt wurde lokal und im Repository verändert
↔ Manuelle Konfliktlösung notwendig



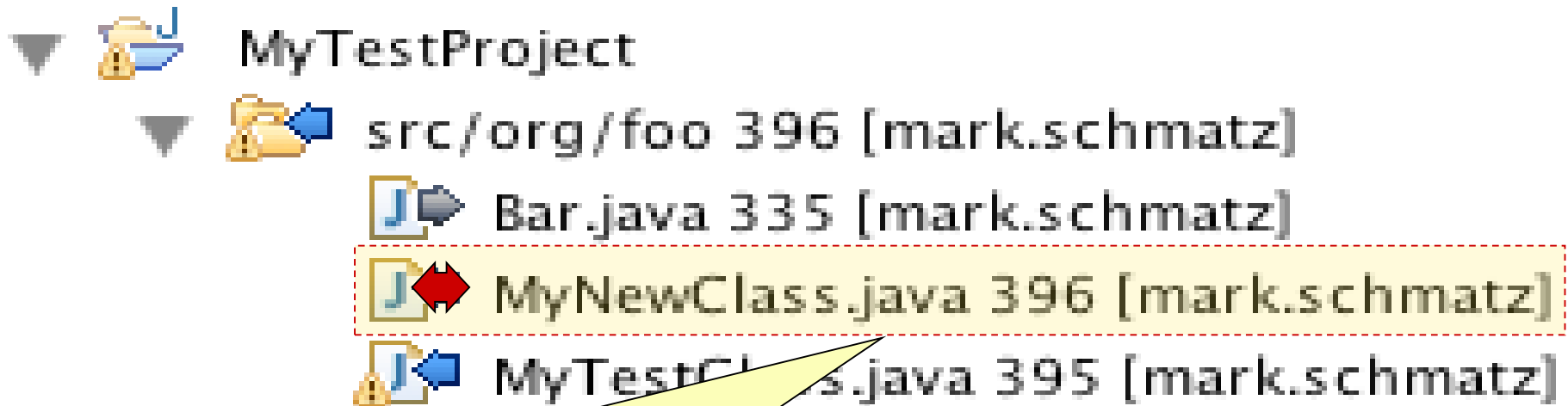
- Lokale veränderte Datei wurde im Repository gelöscht
↔ Manuelle Konfliktlösung notwendig



- Lokale gelöschte Datei wurde im Repository verändert
↔ Manuelle Konfliktlösung notwendig

Synchronisieren: Konfliktauflösung

- Konfliktauflösung erfordert Inhaltsvergleich der Dateien
 - ◆ Angezeigt in “Side-by-side View” / “Compare Editor” von Eclipse
 - ◆ Zeigt Versionsvergleich von Dateien („diff“) übersichtlich an



**Beginne Detailvergleich durch Doppelklick
auf die entsprechende Datei ...**

Synchronisieren: Inhaltsvergleich im „Compare Editor“ Fenster

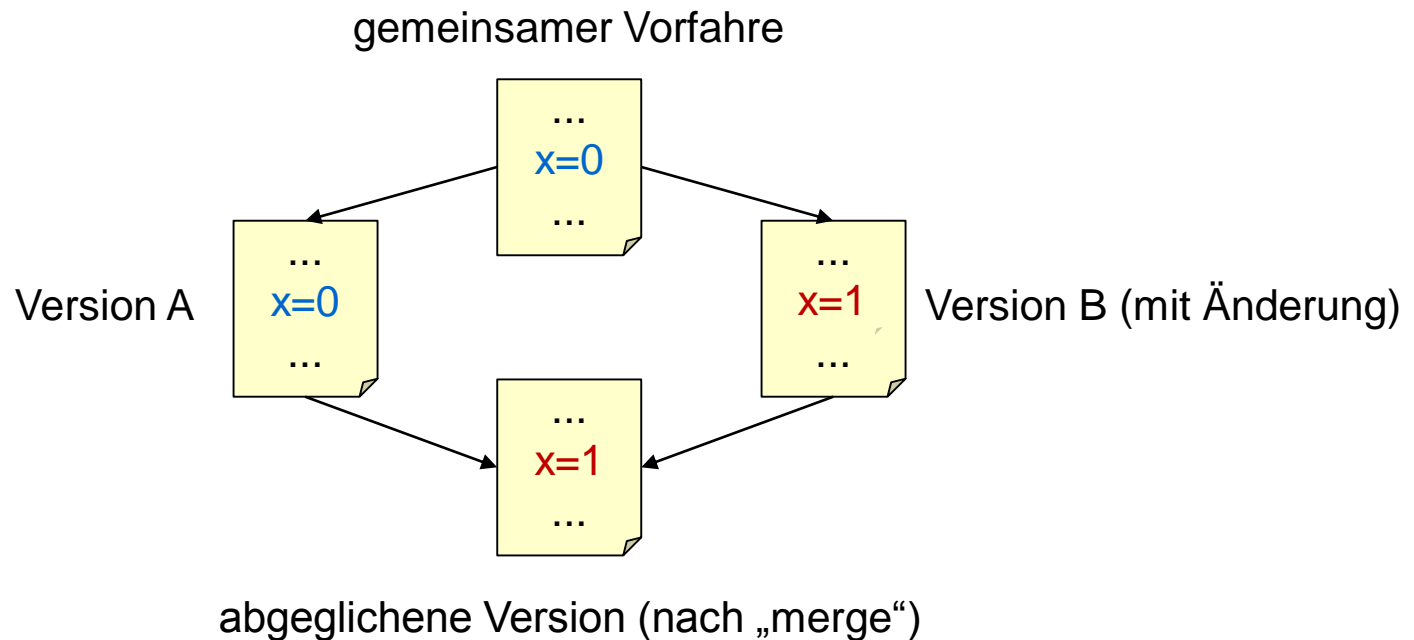
- Hier die Anzeige eines Konfliktes

Local File		Remote File (394 [mark.schmatz])
<pre>public void myTestMethod() { log("Test method entered."); boolean keepOnRunning = true; int i=0; while(keepOnRunning) { int r1 = doSomething1(); int r2 = doSomething3(); if(r1+r2 > MAX_THRESHOLD) { log("Threshold exceeded."); log("Iterations: " + i); keepOnRunning = false; } } }</pre>		<pre>public void myTestMethod() { log("Test method entered."); boolean keepOnRunning = true; int i=0; while(keepOnRunning) { int r1 = doSomething1(); int r2 = doSomethingElse(); if(r1+r2 > MAX_THRESHOLD) { log("Threshold exceeded."); log("Iterations: " + i); keepOnRunning = false; } } }</pre>

- Bei Bedarf kann auch die gemeinsame Bezugsversion angezeigt werden (“Show Ancestor Pane”).
 - ◆ So kann man selbst entscheiden, welches die relevante Änderung ist

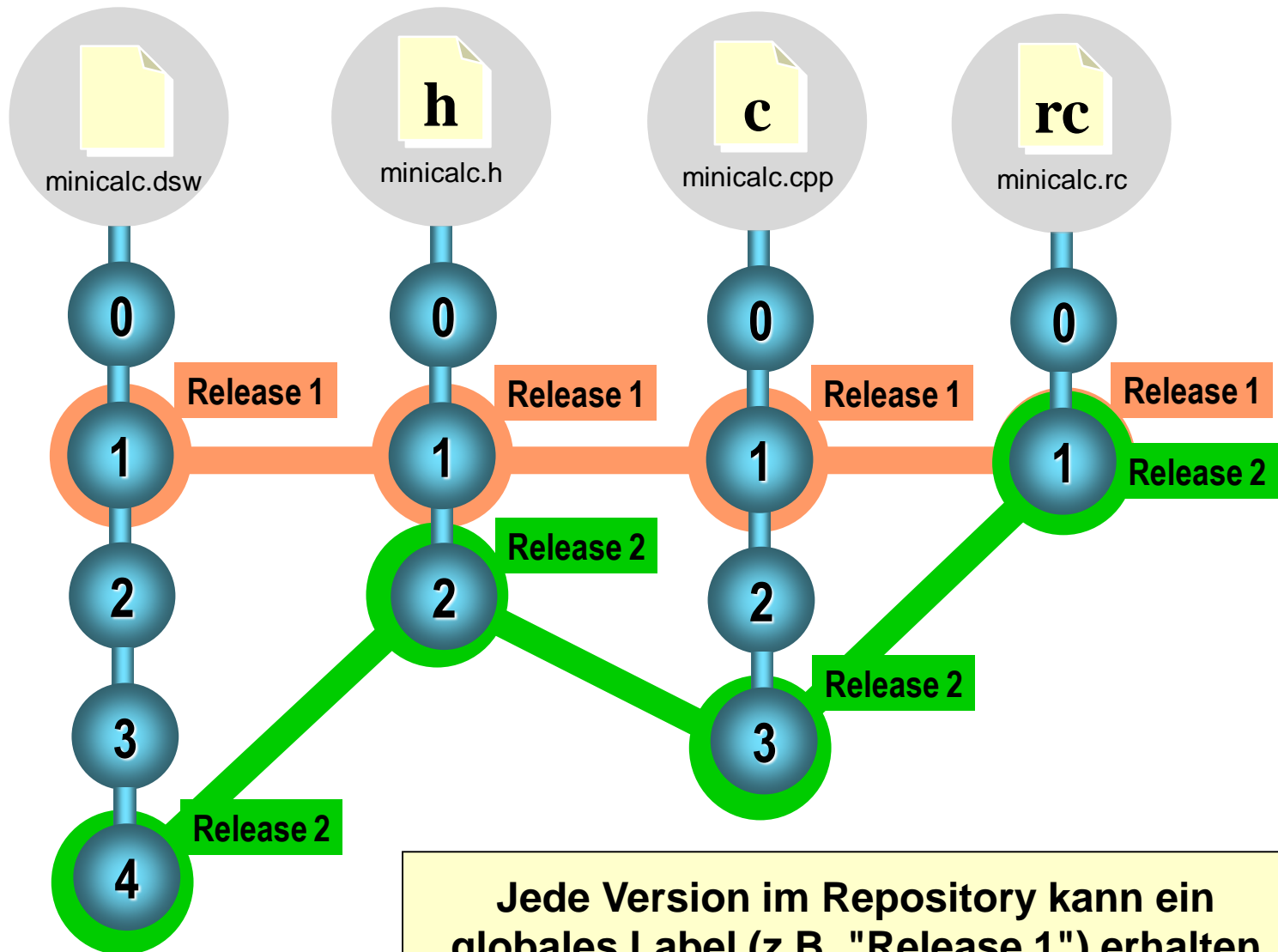
3-Wege-Konfliktauflösung mit Hilfe der „Common Ancestor Pane“

- Anhand des gemeinsamen Vorfahren feststellen was sich geändert hat
- Die Änderung übernehmen



- Mit SVN ist dies ein manueller Prozess: Sie selbst entscheiden
 - ◆ für jede Datei mit Konflikten
 - ◆ für jeden Konflikt in der Datei

Tagging (= Baselining)



Jede Version im Repository kann ein globales Label (z.B. "Release 1") erhalten

Vergleich von Subversion (SVN) und CVS

Versionierung von Verzeichnissen

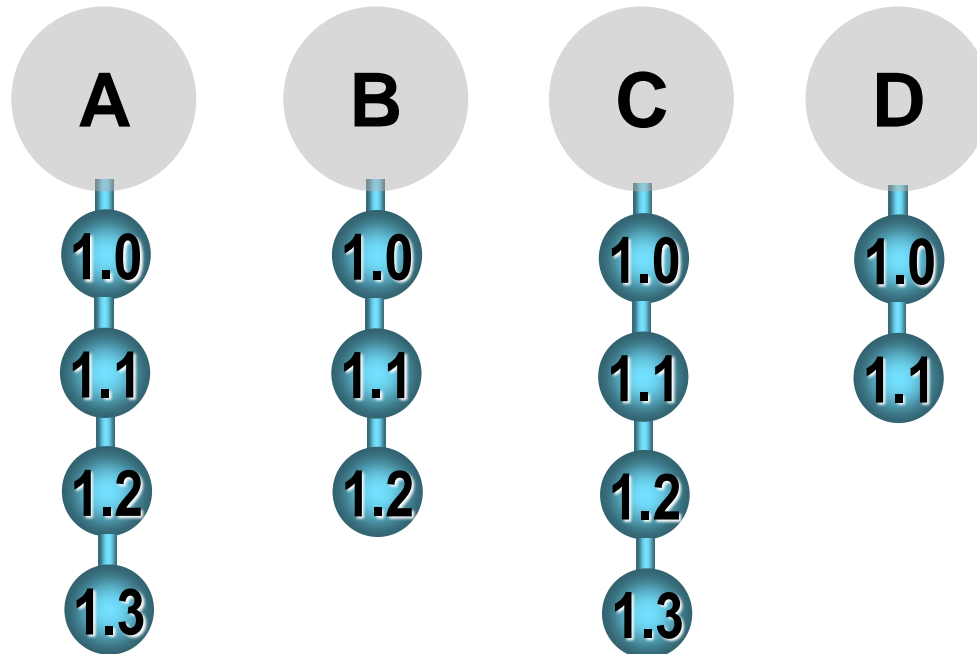
Globale Commit-Nummerierung

Atomare Commit-Aktionen

Subversion kann mehr als CVS:

1. Globale Versions-Nummerierung

- CVS weist jeder Datei ihre **eigenen** Versions-Nummern zu

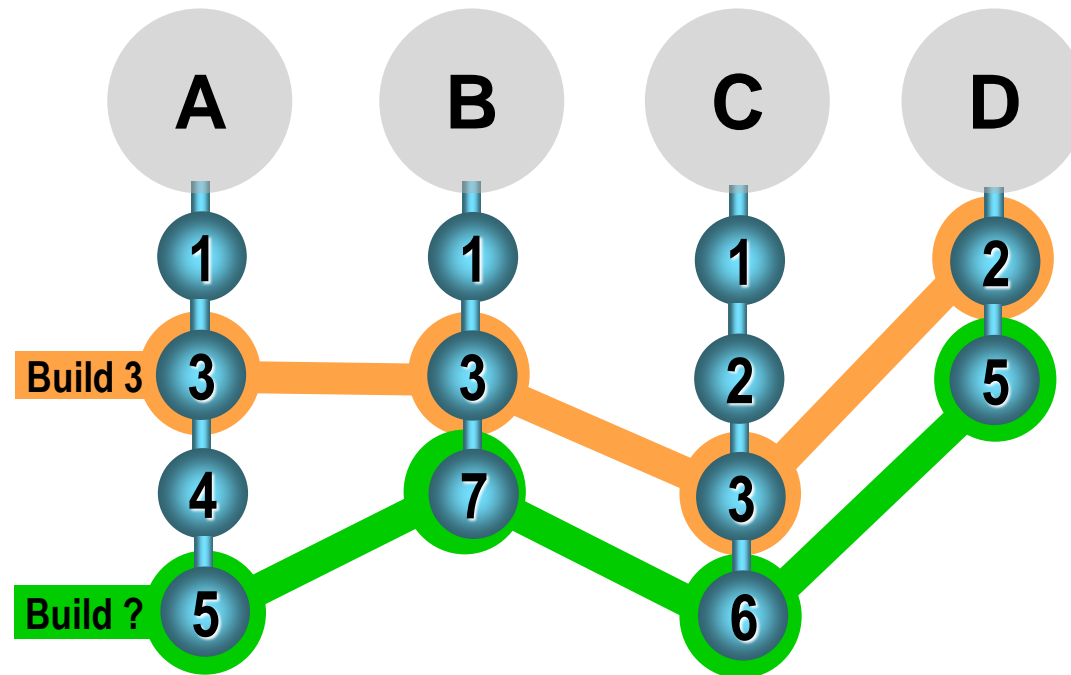


- Problem: Es ist nicht einfach zusehen welche Versionen verschiedener Dateien zusammengehören
 - ◆ „Gehört Version 1.0 von *Datei A* wirklich zu Version 1.1 von *Datei B*?“

Subversion kann mehr als CVS:

1. Globale Versions-Nummerierung

- SVN weist jeder Datei die bei einem Commit verändert wurde die selbe Versions-Nummer zu.

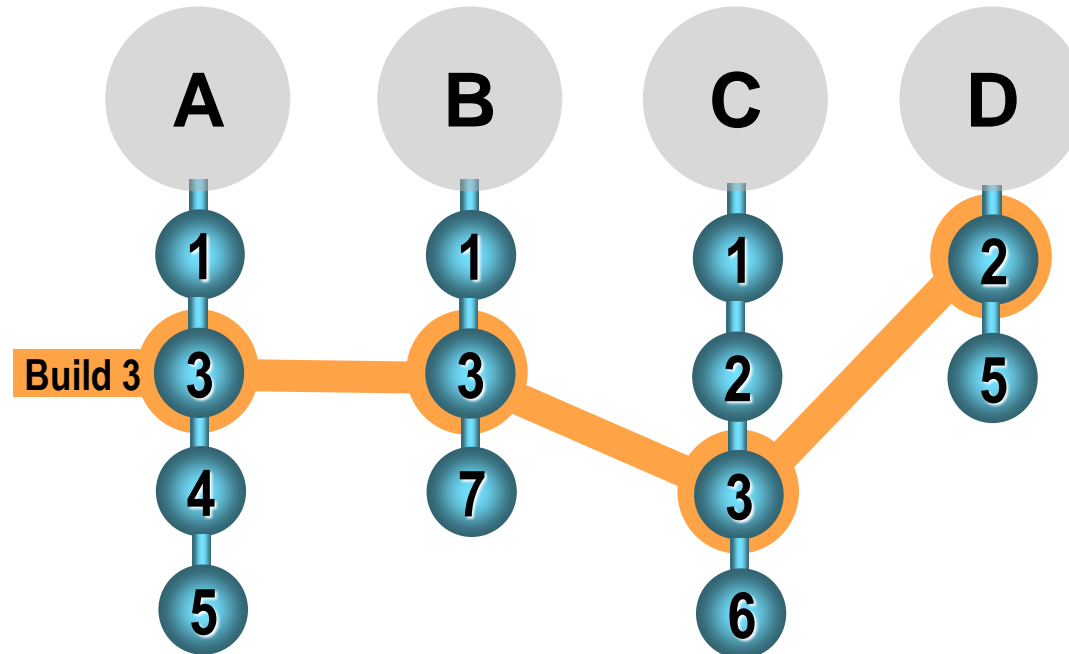


- Vorteil: Dateiversionen die zusammengehören sind auf einen Blick zu erkennen: Zu einer Projektversion V gehören alle Elemente mit Versionsnummer V oder kleiner als V

Subversion kann mehr als CVS:

1. Globale Versions-Nummerierung

- Ein „Roll back“ auf eine alte Version ist einfach, selbst ohne „Tags“
 - ◆ „Gehe zurück zu Version 3“



- Tagging sollte für wichtige Zwischenstände trotzdem genutzt werden
 - ◆ „Release 1.0 alpha“ ist leichter zu merken als Version 1093

Subversion kann mehr als CVS:

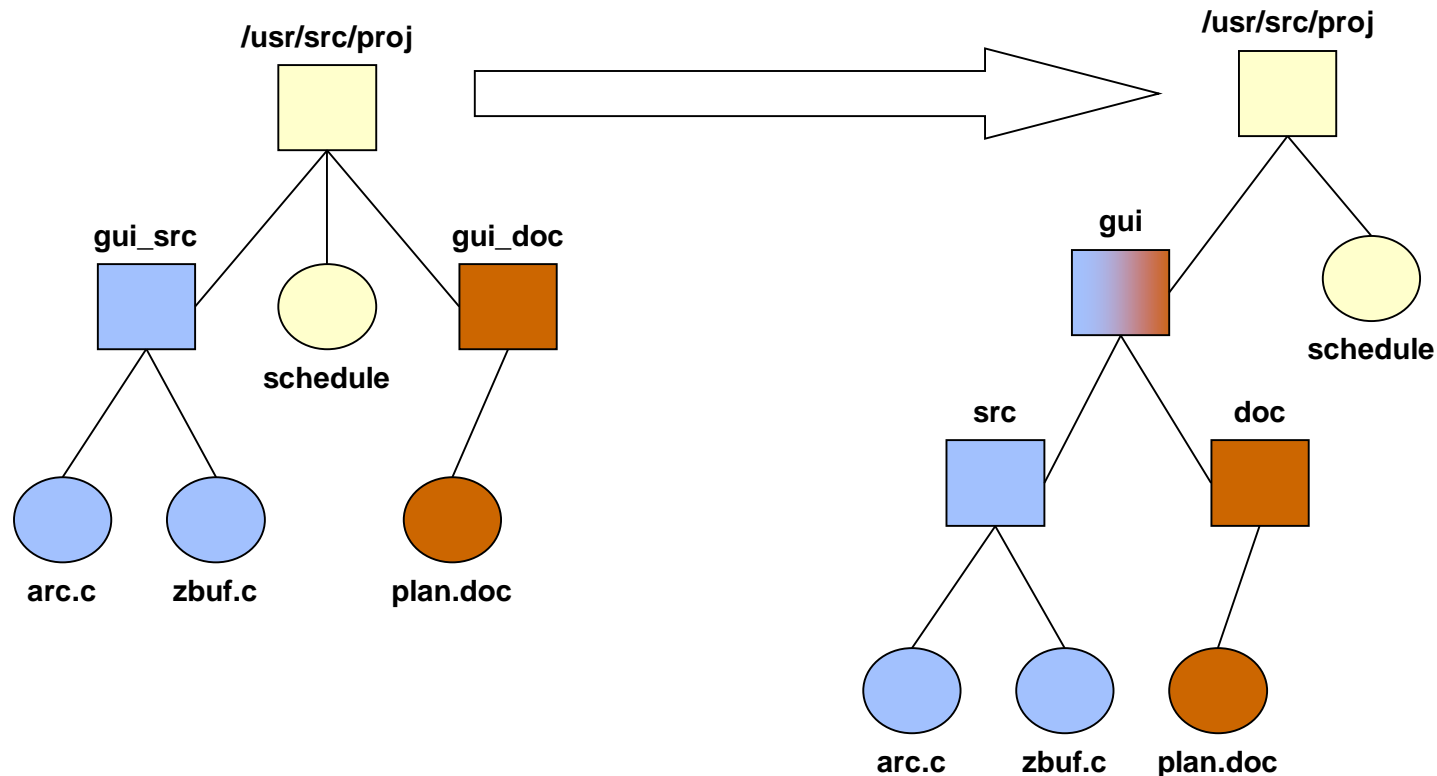
2. Atomare commits

- Netzwerkfehler oder andere Probleme können zu unvollständigen Commits führen (nicht alle Dateien wurden „commitet“)
- Wenn dies passiert, hinterlässt CVS das Repository in einem inkonsistenten Zustand.
- SVN führt bei unvollständigen Commits einen „Roll back“ durch.
 - ◆ SVN benutzt ein Datenbanksystem um das Repository zu speichern und kann daher Commits als atomare Transaktionen implementieren!

Subversion kann mehr als CVS:

3. Versionierung von Verzeichnissen

- Dateien und Verzeichnisse zu löschen, umzubenennen und zu verschieben entspricht einer neuen Version des umgebenden Verzeichnisses
- Hier eine Versionsänderung von `/usr/src/proj`



Subversion kann mehr als CVS:

3. Versionierung von Verzeichnissen

- Es ist in Subversion möglich Dateien und Verzeichnisse zu löschen, umzubenennen und zu verschieben

Warnung 1: Umbenennung und Verschieben

- Man muss diese Operationen mit einem „Subversion Client“ durchführen! → Siehe Folie „Wichtige Links“
- Führt man es selbst auf Systemebene durch wird Subversion nichts vom Umbenennen oder Verschieben erfahren. Es wird annehmen eine Datei wurde gelöscht und eine andere hinzugefügt!

Warnung 2: .svn-Ordner

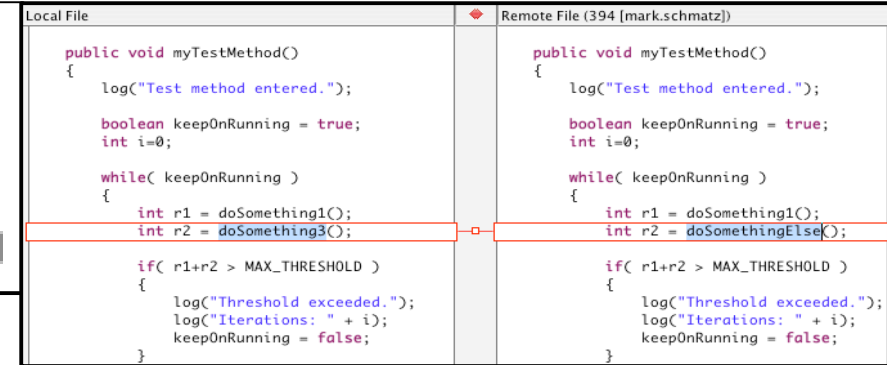
- Die „.svn“ Ordner innerhalb der Arbeitskopie nicht verändern oder löschen! Sie enthalten die Meta-Informationen für SVN über die seit dem letzten Abgleichen mit dem Repository stattgefundenen Änderungen

Von zentraler zu verteilter Versionskontrolle

- ▶ **Motivation:** Nachteile zentralisierter Ansätze
- ▶ **Prinzip:** Was bedeutet verteilte Versionskontrolle?
- ▶ **Lokal arbeiten:** add, commit, diff, checkout, merge, reset
- ▶ **Verteilt arbeiten:** clone, fetch, pull, push
- ▶ **Gesamtüberblick:** Wie alles zusammenspielt

Austausch halbfertiger Zwischenstände

SHOW



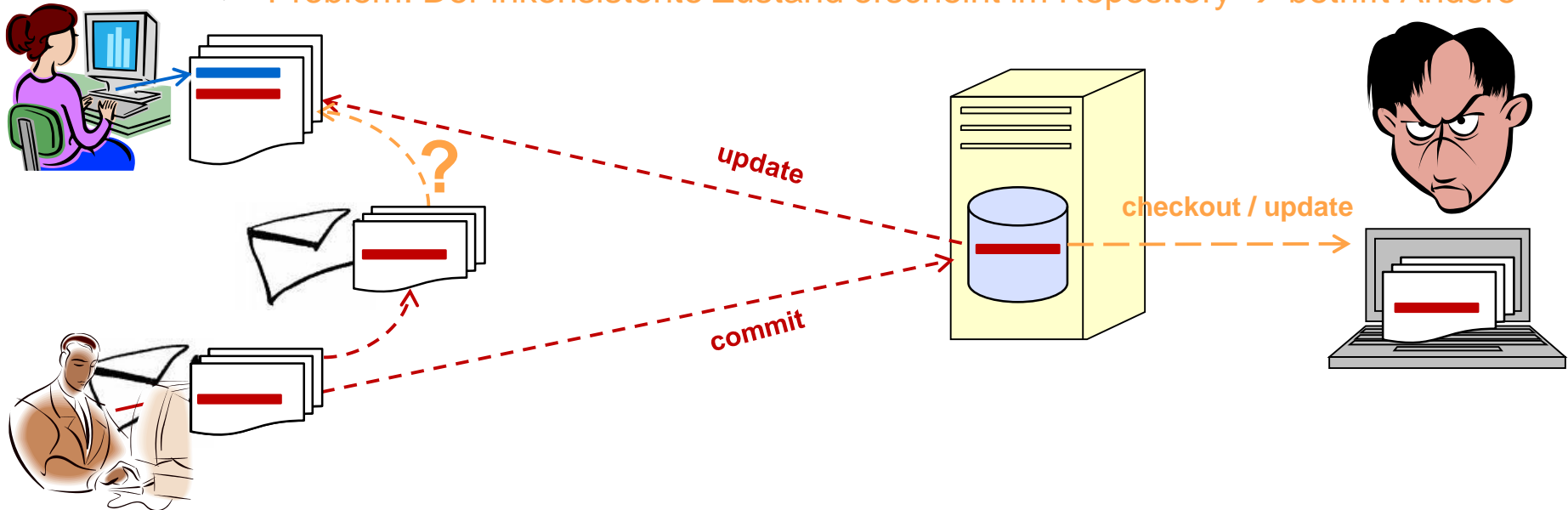
geben, damit sie sie integriert

◆ Ohne SCM: Via e-mail, etc.

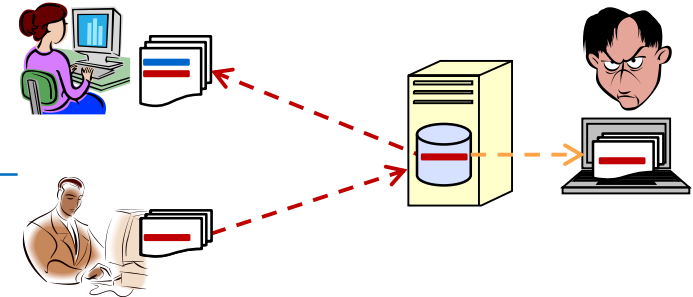
⇒ Problem: Keine Unterstützung für Abgleich (Synchronize & Merge)

◆ Mit SCM: Via commit und update

⇒ Problem: Der inkonsistente Zustand erscheint im Repository → betrifft Andere



Probleme



- Falls Commit von inkonsistenten Zuständen
 - ◆ inkonsistenter Zustand betrifft Andere ☹
 - ◆ Regel: Kein commit von inkonsistenten Zustände! Niemals!!!
- Commit inkonsistenter Zustände verhindern durch spezielle „committer“ Rolle
 - ◆ Nur sehr erfahrene Entwickler haben das Recht zum „Commit“
 - ◆ Sie erhalten jeden Änderungsvorschlag als „patch“ den sie erst auf Ihrer eigenen Arbeitskopie testen und gegebenenfalls ablehnen
 - ◆ Folgeproblem: Engpass
- Falls kein commit inkonsistenter Zustände
 - ◆ Synchronize & Merge nicht für Abgleich von Zwischenzuständen nutzbar
 - ⇒ Fehleranfälliger manueller Abgleich von Zwischenzuständen (Peter → Lisa)
 - ◆ Commit & Revert nicht für Zwischenzustände nutzbar
 - ⇒ Es sammeln sich umfangreiche Änderungen an
 - ⇒ Rücksetzen auf Repository-Stand entsprechend verlustreich und aufwendig

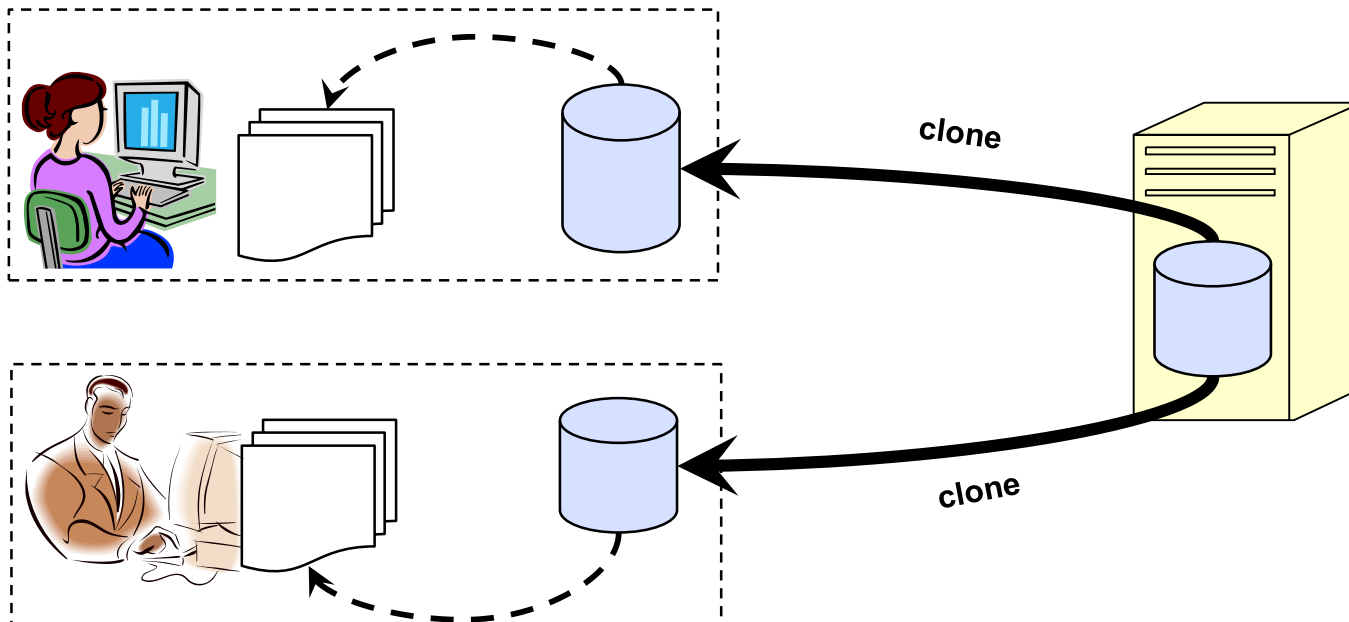
Grundlagen des verteilten Arbeitens

- ▶ Operationen
- ▶ Remotes
- ▶ Zugriffskontrolle
- ▶ Organisatorische Zentralisierung

Verteilte Versionskontrolle ►

Multiple, verteilte Repositories

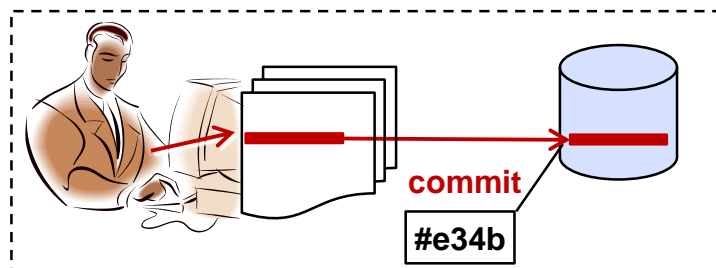
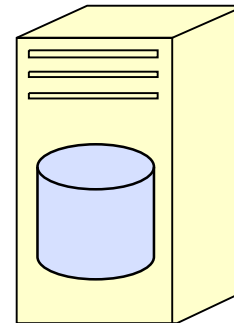
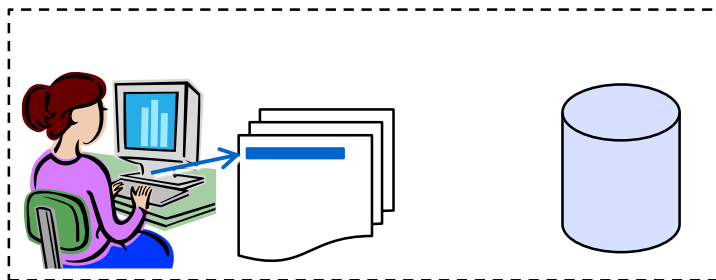
- Jede(r) hat ein oder mehrere lokale Repositories
 - ◆ Mehrere → Verschieden Projekte oder verschiedene Aufgaben im Projekt
- Repository kann geklont werden → **inkl. der gesamten Versionshistorie!**
 - ◆ Ab jetzt kann man lokal weiterarbeiten
 - ◆ Mit allen Funktionalitäten (commit, branch, switch brach, merge, ...)
 - ◆ Man kann diese Funktionen lokal auch für „Unfertiges“ verwenden!



Verteilte Versionskontrolle ►

Multiple, verteilte Repositories

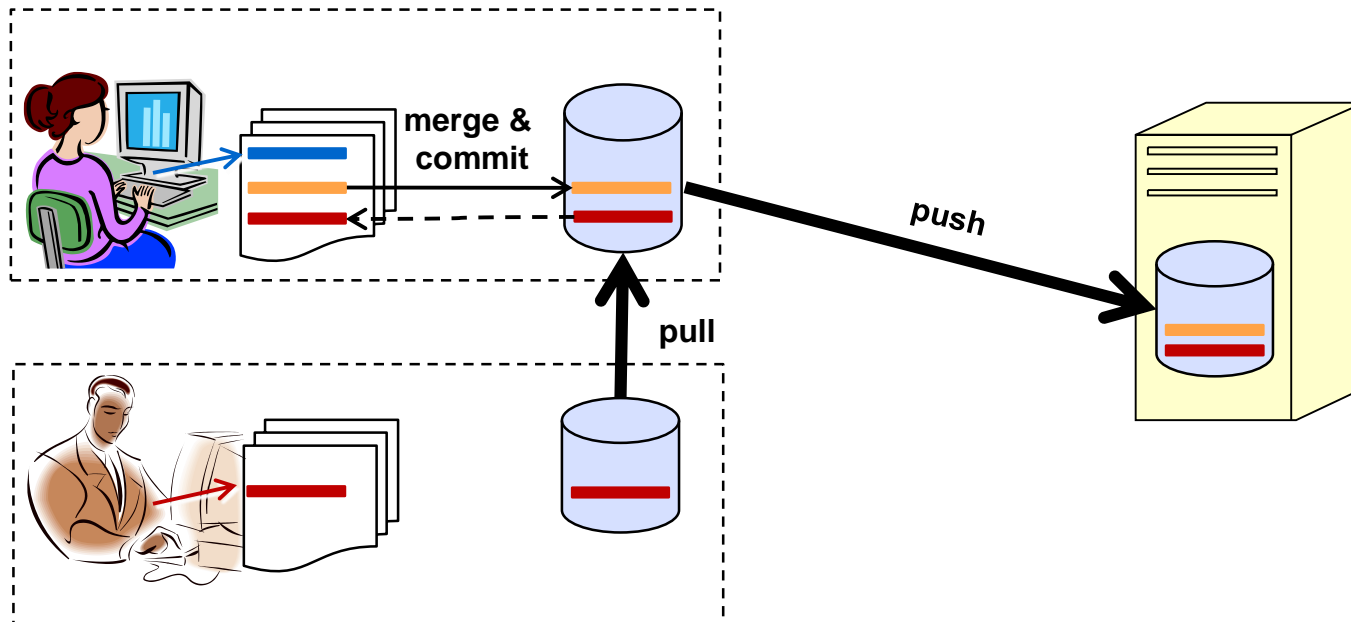
- Checkout, Commit, Branch, Switch Branch, ... geschehen lokal
 - ◆ Sehr schnell
 - ◆ Unabhängig von Netzverbindung und Server
- Fokus auf Gesamt-Repository statt einzelnen Artefakten
 - ◆ Commit = Menge von Änderungen („Change set“ / „Patch“)
 - ◆ Jedes Commit hat eine eindeutige ID
 - ◆ Repository-Zustand = ID einer Änderung = Zustand nach dieser Änderung



Verteilte Versionskontrolle ►

Multiple, verteilte Repositories

- Repositories können miteinander abgeglichen werden
 - ◆ Pull: Holen von Änderungen aus anderem Repository
 - ◆ Push: Übertragen von Änderungen in ein anderes Repository
 - ◆ Pull beinhaltet bei Bedarf automatisches Merge
 - ⇒ Merge wird als eigene Änderung verwaltet (Hat eigene ID)
 - ⇒ In der Abb. nehmen wir an es sei kein Merge nötig gewesen

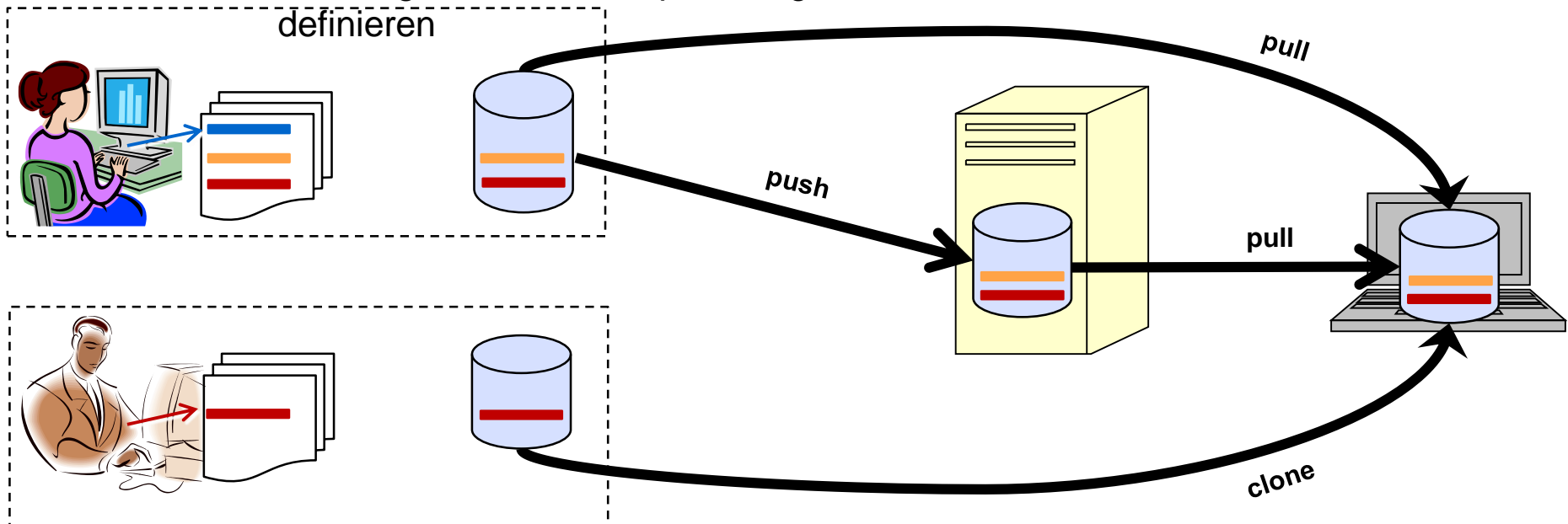


Verteilte Versionskontrolle ►

Multiple, verteilte Repositories

- Technische Gleichberechtigung

- ◆ Jedes Repository kann geklont werden
- ◆ Zu jedem kann „push“ und von jedem kann „pull“ durchgeführt werden
- ◆ Voraussetzungen
 - ⇒ „remote“ setzen = anderes repository dem eigenen bekannt machen
 - ⇒ „fetch“ = Entfernte „branches“ holen / bekannt machen
 - ⇒ „tracking“ setzen = Entsprechungen von lokalen und entfernten „branches“ definieren

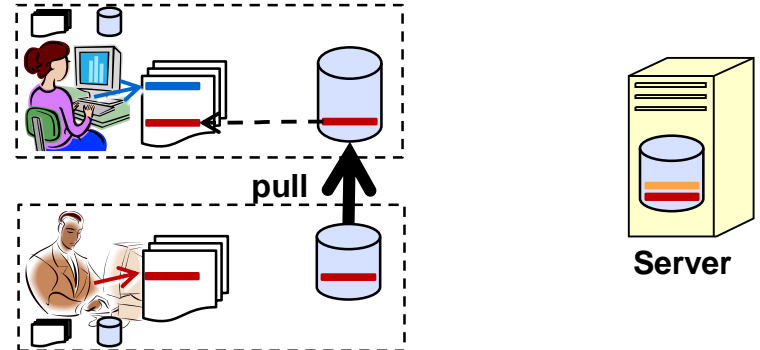
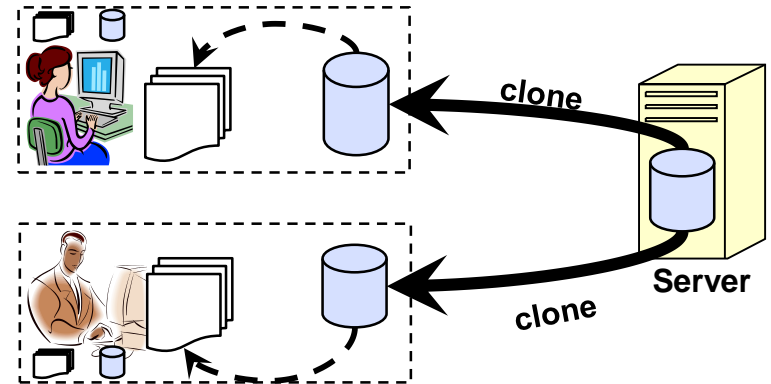


Verteilungsspezifische Git Terminologie

- **remote** = anderes bekanntes repository
 - ◆ Spezifiziert durch URL oder Pfad im lokalen Dateisystem
- **origin** = Repository von dem dieses geklont worden ist
 - ◆ ausgezeichnete Rolle unter den „remotes“
- **tracking** = Beziehung zwischen lokalen und entfernten „branches“
 - ◆ Push / pull bezieht sich immer auf die branches zu denen ein tracking definiert ist
 - ◆ Push / pull kann für **einen** lokalen branch und seinen tracking branch in einem remote durchgeführt werden
 - ◆ ... oder für **alle** lokalen branches und ihre jeweiligen tracking branches in dem aktuell ausgewählten remote.

Remotes im vorherigen Beispiel

1. Nach dem clonen haben die lokale Repos von Lisa und Peter den Server als einzigen remote mit Standardnamen „origin“
`git remote add ... <name> <url>`
2. Um das „pull“ von Peter durchzuführen muss Lisa erst sein Repo als zusätzlichen remote eintragen
`git remote add ... <name> <url>`
3. ... und den Namen des neuen remotes beim „pull“ als Quelle angeben
`git pull ... <name>`



Details siehe <https://git-scm.com/docs/git-remote>,
[.../git-pull](https://git-scm.com/docs/git-pull) und [.../git-fetch](https://git-scm.com/docs/git-fetch)

Zugriffskontrolle

Wie kontrolliert Peter wer auf sein Repository zugreifen darf?

- GIT hat keine eigene Zugriffskontrolle
- Es wird stattdessen der Mechanismus genutzt, der gerade verfügbar ist:

Zugriff auf	über	Zugriffskontrolle
Lokales Repository	Betriebssystem	Zugriffskontrolle des BS
Remote Repository	FTP	Username + Passwort
	HTTPS	Username + Passwort
	SSH	Private Key (lokal) + Public Key (remote)
	GIT	Read only

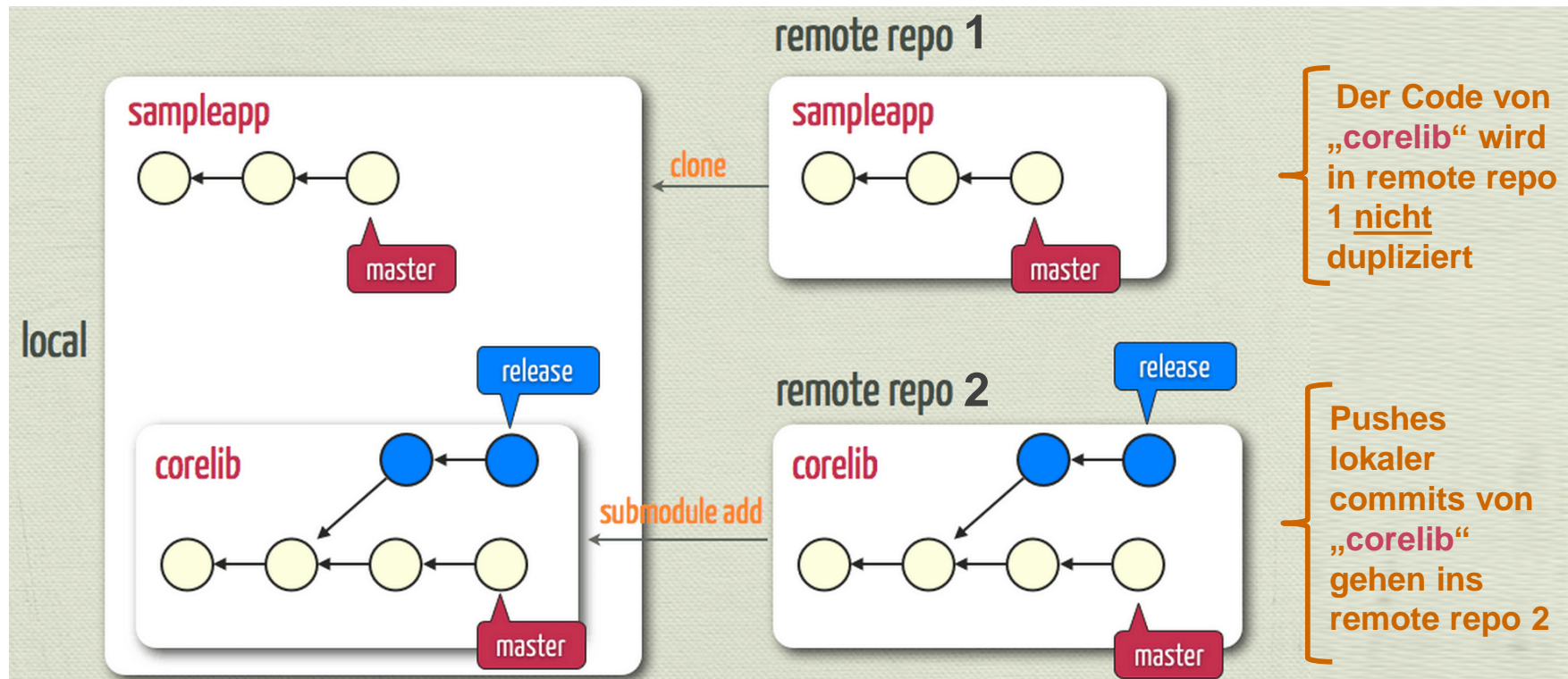
Zentralisierung trotz Dezentralisierung

Zentrales Referenz-Repository ist organisatorisch möglich

- Team vereinbart, welcher „remote“ die „offizielle“ Bezugsversion ist
- So muss man nicht alle remotes kennen, um dennoch an den aktuellsten Stand zu kommen
- Beliebiger skalierbar
 - ◆ Jedes Team hat seinen eigenen Bezugsserver
 - ⇒ Z.B. für Linux Filesystem, Prozessverwaltung, ...
 - ◆ Integration kann mehrstufig erfolgen bis hin zum Server der alles enthält
- Submodul-Konzept hilfreich
 - ◆ Eigenständiges Git-Repository, das lokal als Teil eines übergeordneten Repositories verwaltet wird
 - ◆ So können in einem Repository verschiedene Teil-Repositories für verschiedene Teilsysteme oder Libraries von Drittanbietern existieren
 - ◆ Grundidee: → Nächste Folie
 - ◆ Details: → <https://git-scm.com/book/en/v2/Git-Tools-Submodules>

Submodule ► einheitliche lokale Gesamtsicht auf dezentrale Repositories

- Submodul = Eigenständiges Git-Repository, das lokal als Teil eines übergeordneten Repositories verwaltet wird
- Beispiel ► „corelib“ duplikationsfrei ins eigene Projekt „sampleapp“ integriert:



Grundlagen des lokalen Arbeitens

- ▶ **Speicherbereiche:** Arbeitsbereich, Stage / Index, Repository
- ▶ **Der Weg ins Repository:** add, commit
- ▶ **Der Weg aus dem Repository:** checkout, reset

Ehre wem Ehre gebührt

Danksagung



Der Inhalt der folgenden Folien ist zum Teil angelehnt an

<https://marklodato.github.io/visual-git-guide/index-de.html>

Unser Dank gilt Mark Lodato für die ursprüngliche Website und insbesondere all seine tollen Grafiken und an Martin Funk für die deutsche Übersetzung

Acknowledgement

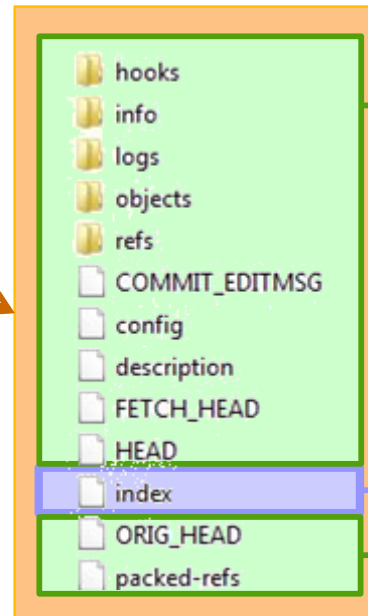
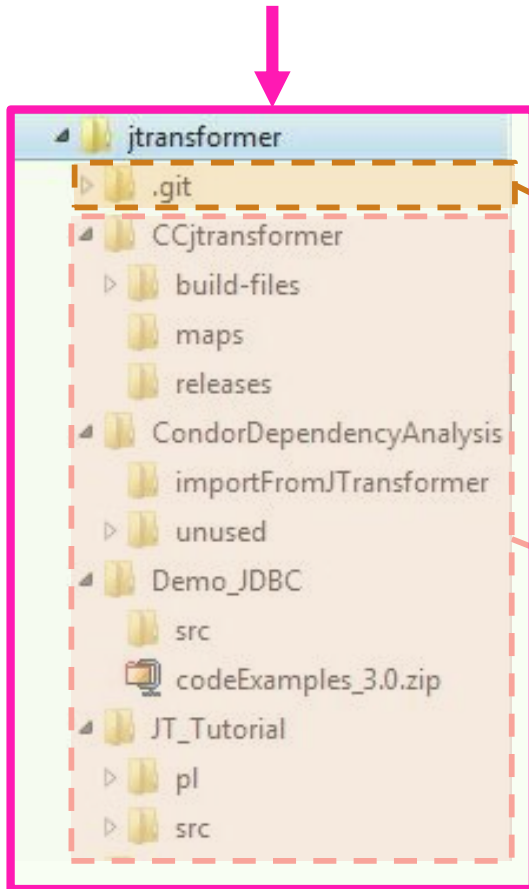


The following slides are inspired by

Our thanks go to Mark Lodato for the original website and especially for all his great graphics and to Martin Funk for the German translation

Der GIT-Ordner und der .git Ordner

GIT-Ordner = Ordner der das GIT-Repository und die Arbeitskopie enthält:



Konzeptuelle lokale Arbeitsbereiche für git-Operationen:

History
(repository)

Index
(stage)

Sandbox
(working directory)

Lokale Arbeitsbereiche

History (repository) = Kopien aller unter Versionskontrolle stehenden Dateien in all ihren Versionen und komplette Historie der durchgeführten Veränderungen, Branches, Commits, Tags, ...

Index (Stage) = Kopien aller seit letztem „commit“ durchgeführten und zur Übernahme ins Repository „vorgemerkten“ Änderungen.
Das nächste „commit“ wird sie ins Repo übertragen und die Stage leeren.

Sandbox (Working directory) = Dem Entwickler aktuell zugängliche Dateien, d.h, noch nicht unter Versionskontrolle stehende Dateien und aus dem Repo ausgecheckte (eventuell veränderte) Dateien. Ihre Veränderungen können im Index gespeichert sein oder auch nicht (= noch nicht zu committen).

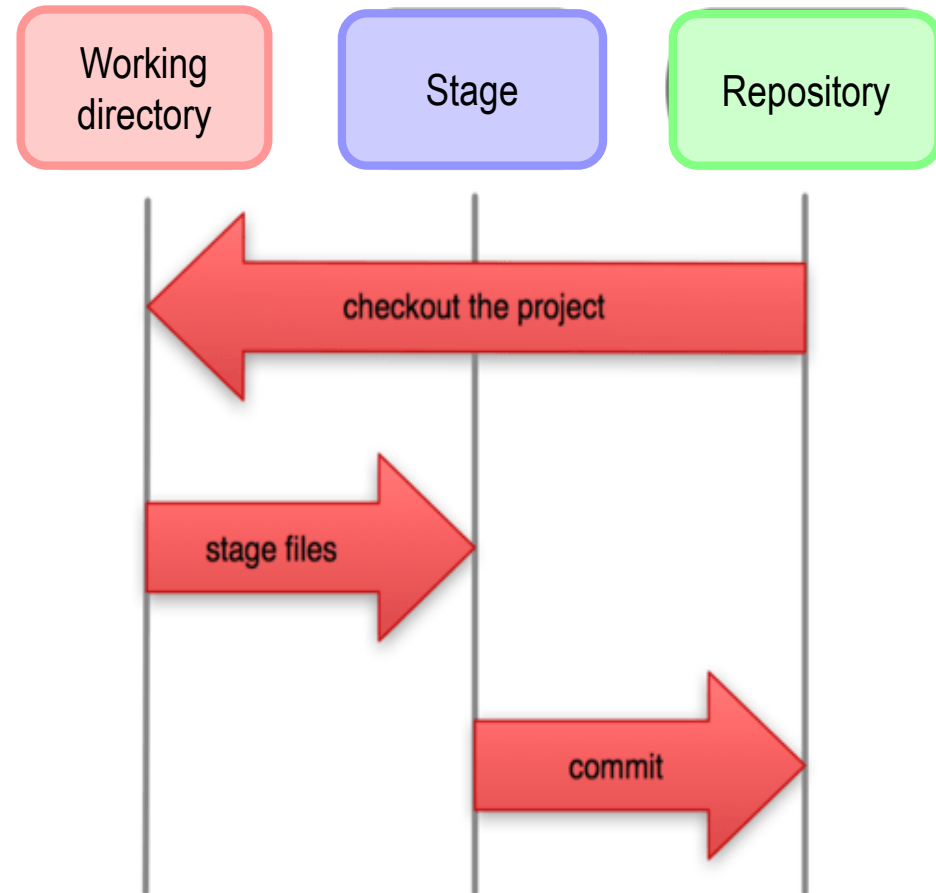
Konzeptuelle lokale Arbeitsbereiche
für git-Operationen

History
(repository)

Index
(stage)

Sandbox
(working directory)

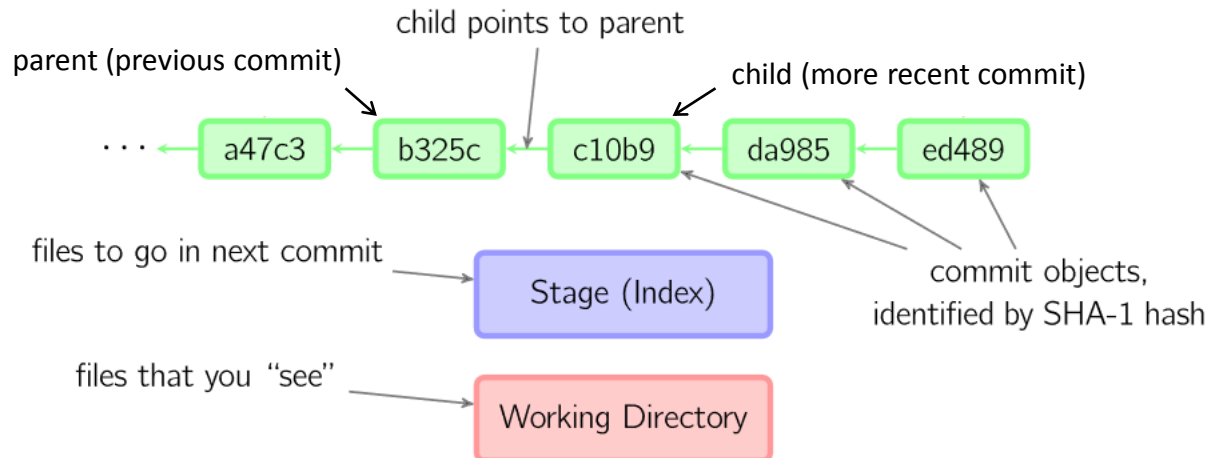
- Checkout
 - ◆ Überschreibt lokale Arbeitskopie
- Staging
 - ◆ erlaubt den Entwicklern ihre Commits so zusammenzubauen wie sie es wünschen.
 - ◆ Dazu werden in der „**Staging area**“ (= dem „**Index**“) die ausgewählten geänderten Dateien zwischengespeichert
- Commit
 - ◆ Inhalt der Stage ins repository übernehmen



Interne Darstellung und Operationen im Detail

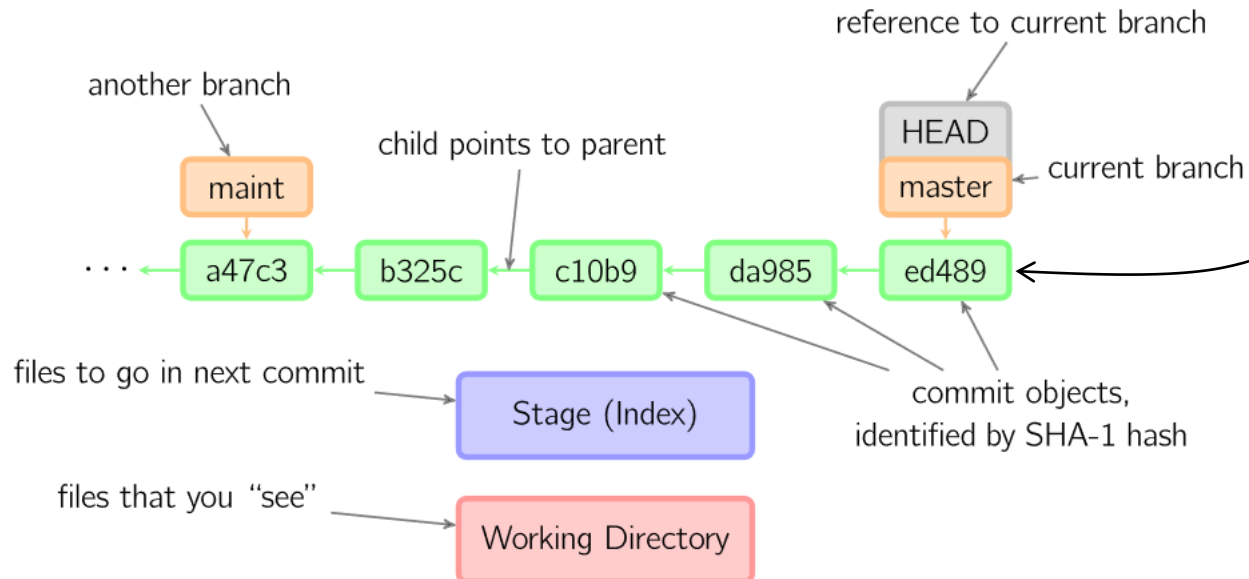
Interne Darstellung ► Commits

- Ein Commit ist eine Momentaufnahme („snapshot“) der unter Versionskontrolle stehenden Dateien zu einem bestimmten Zeitpunkt
 - ◆ Commits werden im Folgenden grün mit 5-Buchstaben-IDs dargestellt
- Jeder Commit verweist auf seine Eltern (Vorhänger)
 - ◆ Normalfall → Genau ein Vorgänger
 - ◆ Commit für eine Merge-Operation → Zwei Vorgänger



Interne Darstellung ► Branches

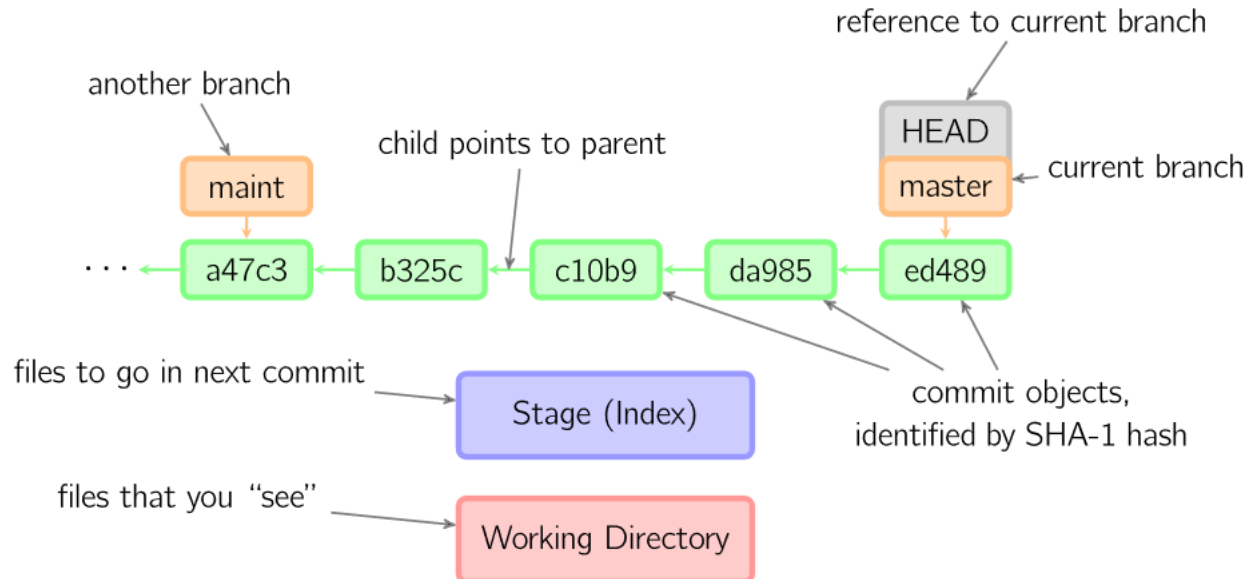
- Ein Branch ist konzeptuell eine lineare Folge von Commits
- Ein Branch ist intern ein Zeiger auf das neueste Commit in der Folge
 - ◆ in orange dargestellt
- HEAD = der aktuelle Branch
 - ◆ meint oft implizit den aktuellsten Commit des aktuellen Branches
- HEAD~n = n commits zurück im aktuellen Branch



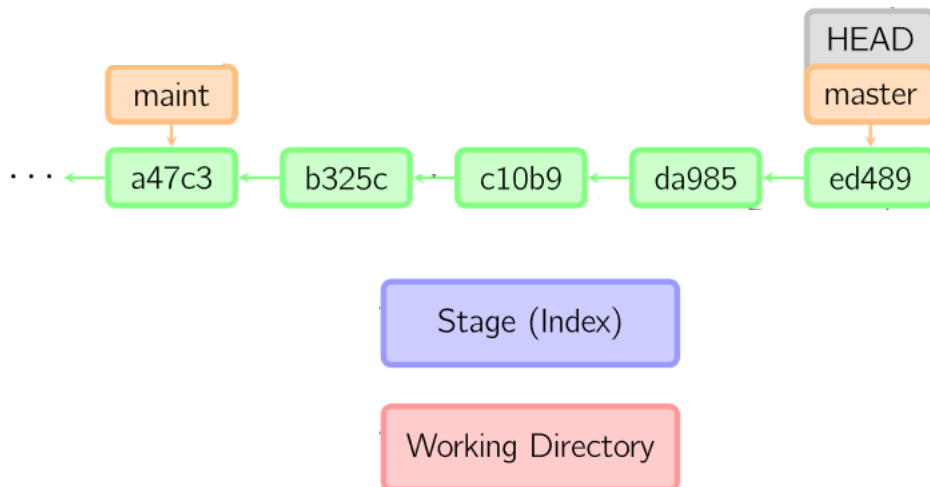
Interne Darstellung ► Branches

In diesem Bild werden die fünf aktuellsten Commits gezeigt, wobei ed489 der jüngste ist.

- **master** zeigt auf genau diesen Commit
- **maint** zeigt auf den 4-ten Vorfahren von master
- **HEAD** (=der aktuelle Branch) ist **master**.



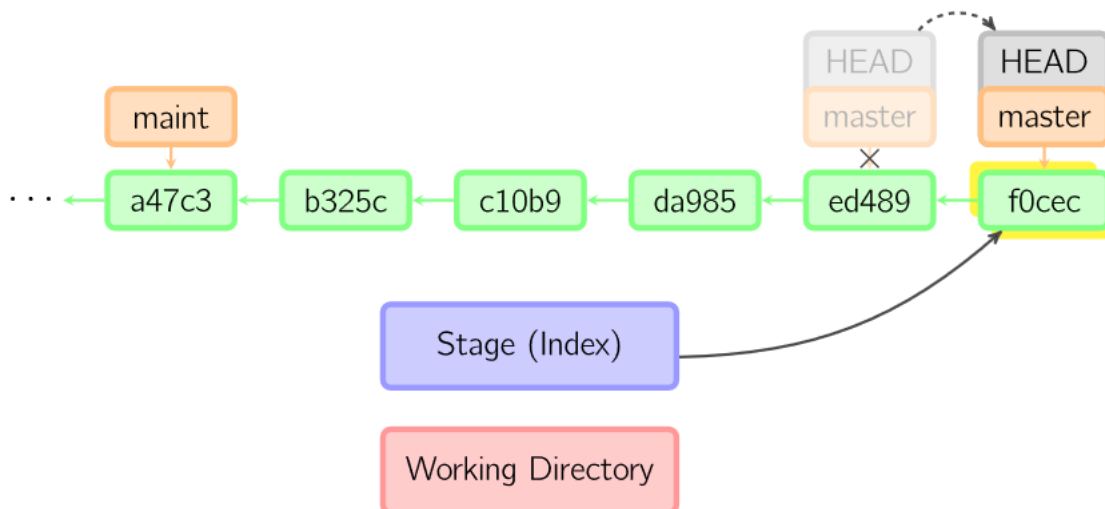
Interne Darstellung ► Branches



- Erzeugt ein neues Commit Objekt mit den Dateien des Indexes und setzt den aktuellen Commit als Vorfahre ein.
- Verschiebt den Zeiger des aktuellen Branches auf den neuen Commit.

In dem folgenden Bild ist der aktuelle Branch *master*:

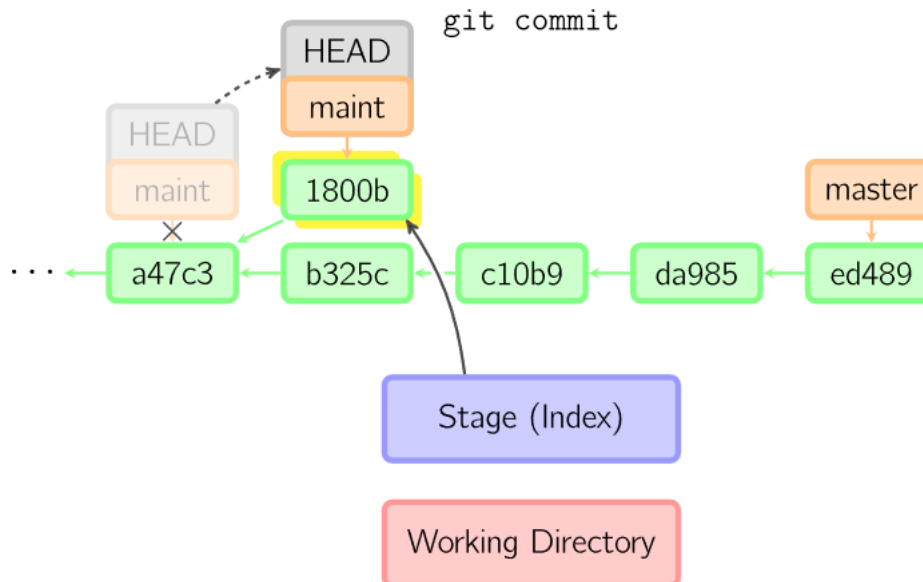
- ◆ Vor dem commit zeigte *master* auf *ed489*
- ◆ Danach wurde ein neuer Commit *f0cec* mit dem Vorläufer *ed489* erstellt
- ◆ ... und die Branch Referenz *master* auf den neuen Commit verschoben.



git commit

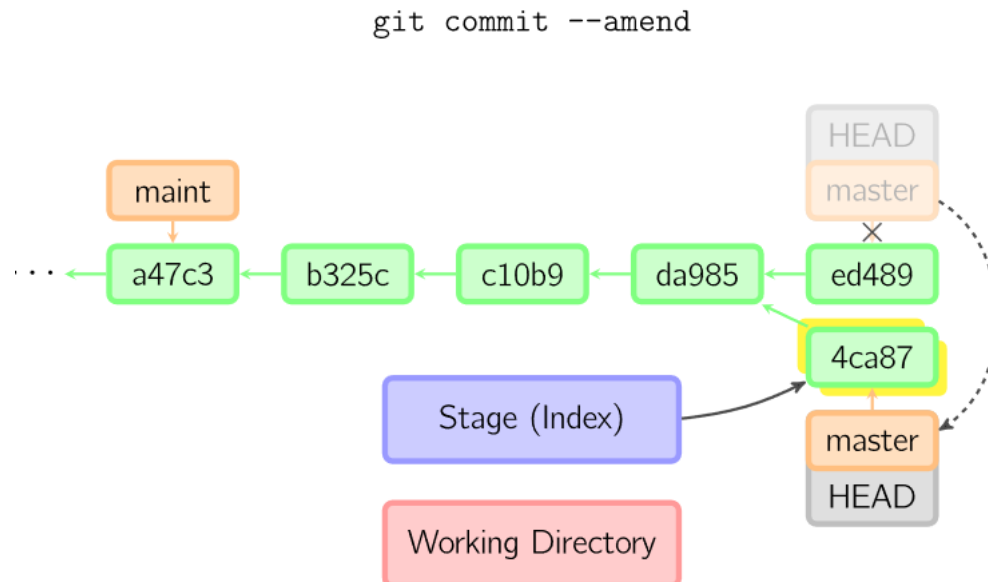
Es funktioniert auch wenn die Spitze des aktuellen Branches Kinder in einem anderen Branch hat:

- Beispiel: Commit **a47c3** ist Spitze von **maint** und hat **b325c** als Kind
 - ◆ Nun commit **1800b** in den **maint** Branch einfügen
 - ◆ Danach divergieren **maint** und **master**.
 - ◆ Um beide zusammen zu führen ist ein **merge** (oder **rebase**) notwendig.



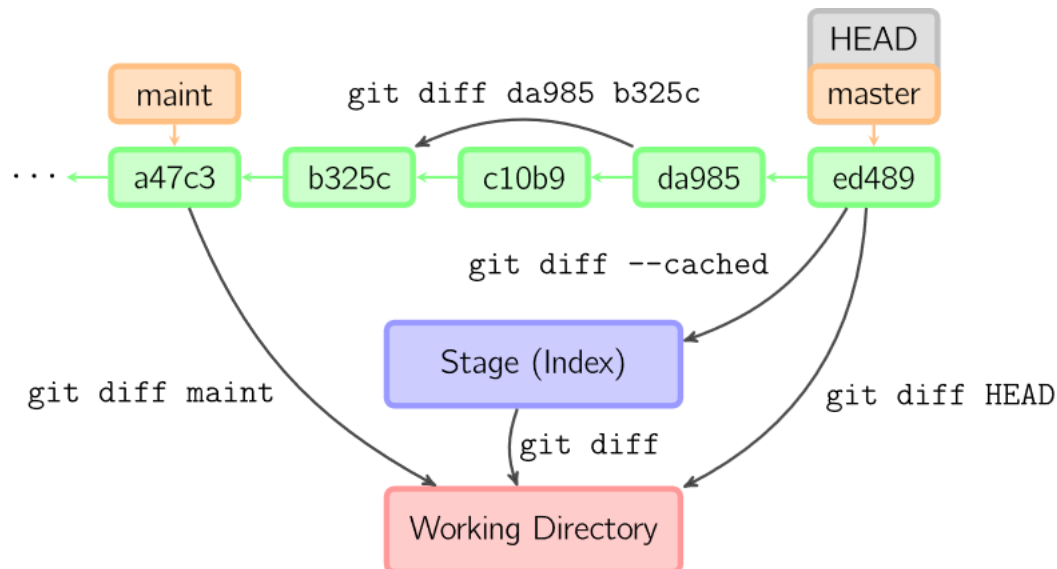
git commit -- amend

- Die `-- amend` Option ersetzt den vorherigen Commit
 - ◆ weil man zu viel oder zu wenig committed hat ...
 - ◆ oder um den Commit-Kommentar zu korrigieren
- Git erstellt einen neuen Commit der den selben Vorgänger hat wie der aktuelle Commit:



git diff

- **git diff** erlaubt es Unterschiede zwischen Commits zu betrachten
 - ◆ implizit: Stage versus Arbeitsbereich – **git diff**
 - ◆ explizit: Head versus Arbeitsbereich – **git diff HEAD**
 - ◆ explizit: Head versus Stage – **git diff --cached**
 - ◆ explizit: Branch versus Arbeitsbereich – **git diff maint**
 - ◆ explizit: zwei commits – **git diff da985 b325c**
 - ◆ Es können explizit Dateien als Argument angehängt werden.
Dies beschränkt die Operation auf diese Dateien.



Commits ohne vorhergehendes Staging

- **git commit -a**

- ◆ Automatisches staging aller Dateien unter Versionskontrolle vor dem commit → Committed alle Änderungen im Arbeitsbereich
- ◆ Oft nicht was man will! ☹️

- **git commit files**

- ◆ Stage und commit der Dateien **files** aus Arbeitsbereich

Arbeiten mit „Branches“

- ▶ **Erzeugen:** branch
- ▶ **Umschalten:** checkout
- ▶ **Abgleichen:** merge, cherry picking, rebase

Kurzüberblick – Branch, Checkout, Merge

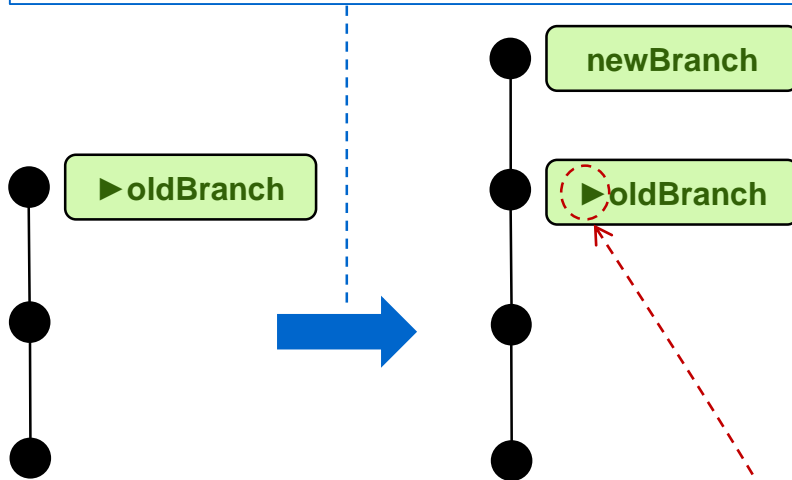
Neben der dezentralen Struktur bietet Git noch eine besondere Stärke: Es ist sehr leicht neue Entwicklungslinien („branches“) zu erzeugen und wieder zusammen zu führen („merge“).

- Typische Vorgehensweise:
 1. Branch erzeugen
 2. Implementieren und Testen (Bugfix, neue Funktionalität, etc.)
 3. Merge
- „Merge“ kann zwischen beliebigen branches stattfinden. Ablauf:
 - ◆ Wechsel in den Ziel-Branch: „**checkout** *Zielbranch*“
 - ◆ Merge des Quellbranches in aktuellen branch: „**merge** *Quellbranch*“
- „Merge“ findet automatisch statt als Teil eines „pull“. Dabei wird ein „tracking branch“ des Remote Repositories (von dem das „pull“ durchgeführt wird) in den entsprechenden („getrackten“) branch des lokalen Repositories ge-merged.
- Merge kann zu Konflikt(en) führen

Branch erzeugen und wechseln

Add Branch

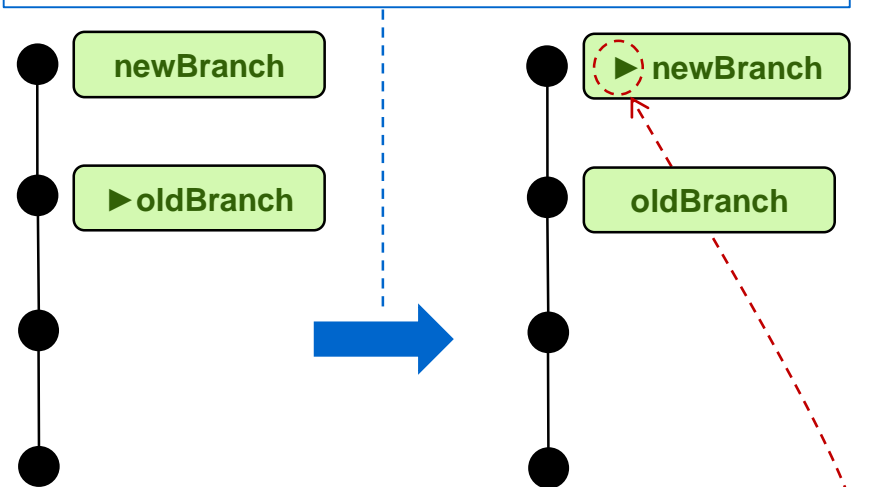
```
git branch newBranch
```



- „newBranch“ erstellt
- Weiterarbeiten in „oldBranch“
 - Nächster „commit“ geht in „oldBranch“

Checkout Branch

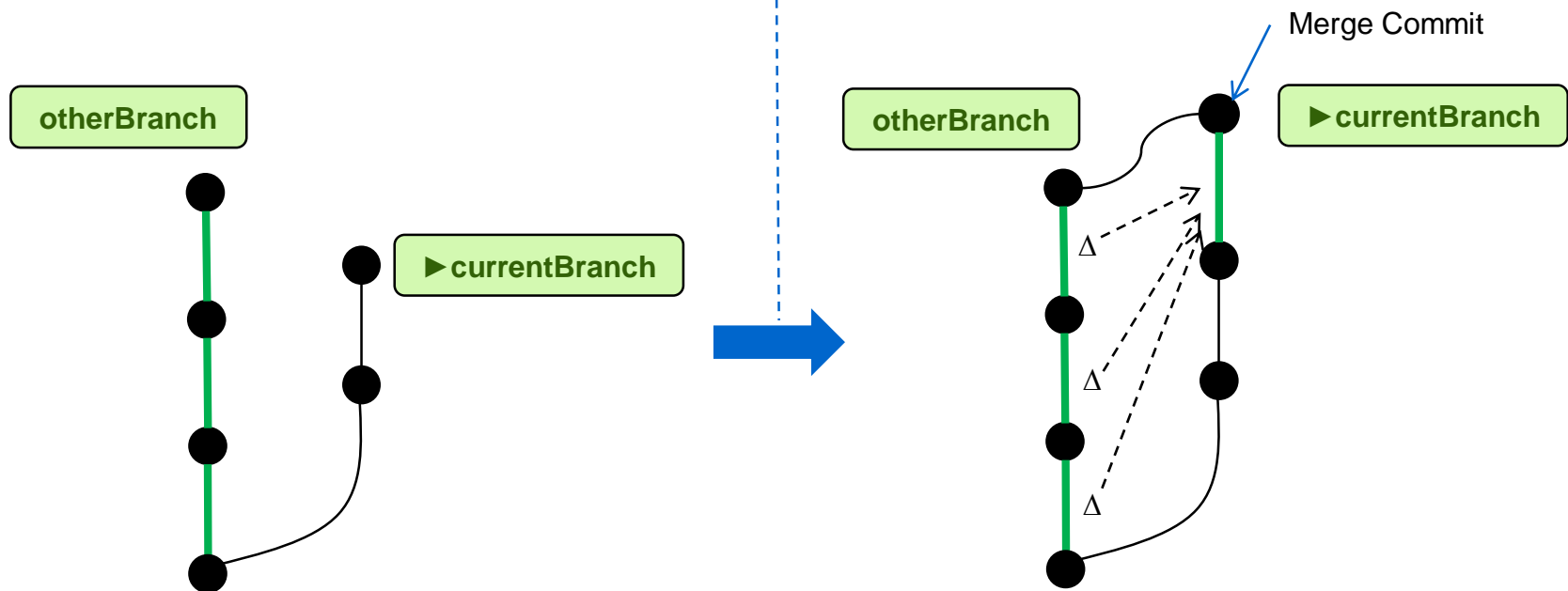
```
git checkout newBranch
```



- Wechsel in „newBranch“
 - Nächster „commit“ geht in „newBranch“

Merge: Branches Synchronisieren

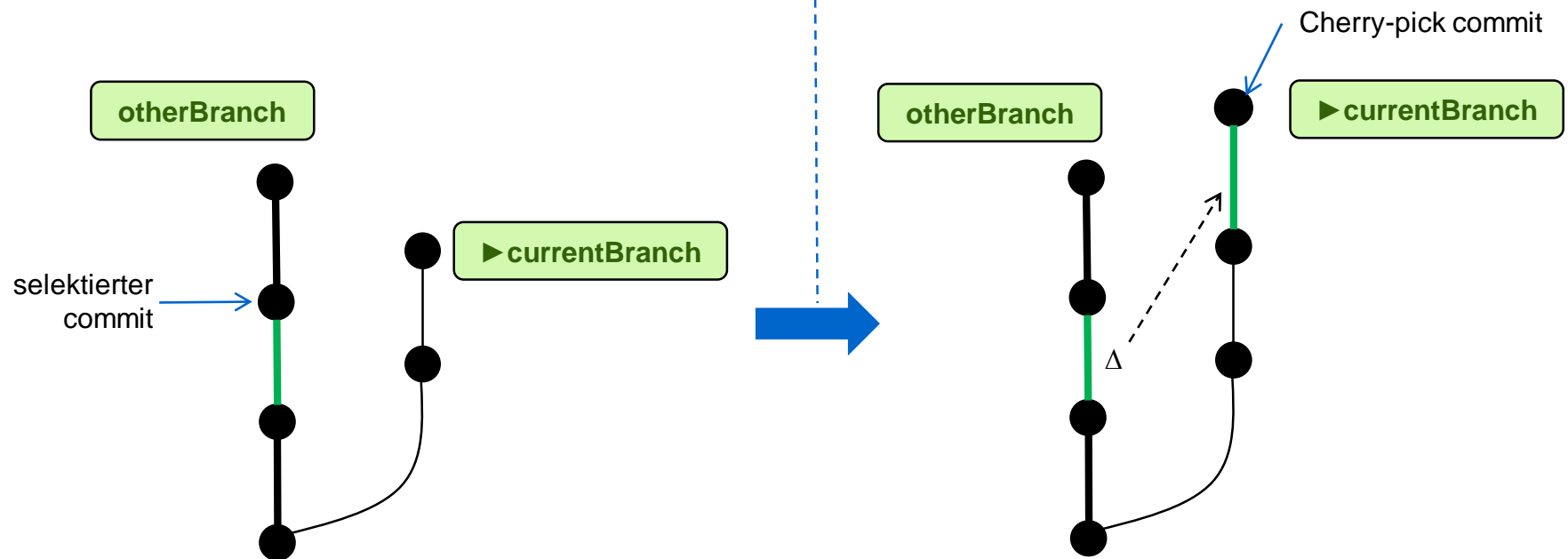
```
git checkout currentBranch  
git merge otherBranch
```



Alle Änderungen vom Verzweigungspunkt der beiden branches bis zum neuesten Commit des Quellbranches werden in den aktuellen Branch übernommen

Chery Pick: Einzelnes Commit übernehmen

```
git checkout currentBranch  
git cherry-pick commit
```



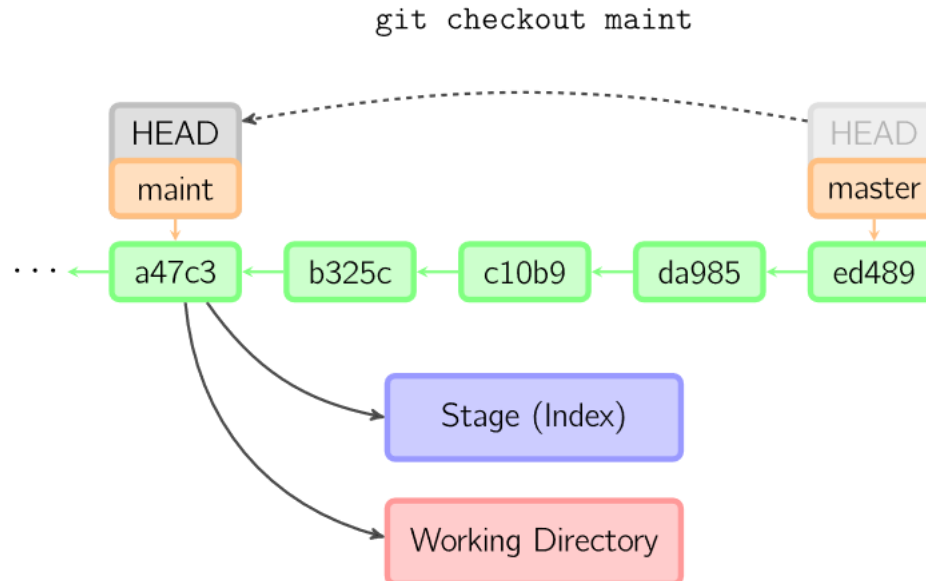
Nur Änderungen des selektierten Commits aus einem anderen Branch werden in den aktuellen Branch übernommen (z.B. um Bugfix selektiv zu übernehmen)

git checkout **branch**

- ▶ **Hier**: Nutzung von checkout zum Wechsel in einen anderen Branch
- ▶ **Später**: Nutzung von checkout zur Wiederherstellung alter Dateiversionen

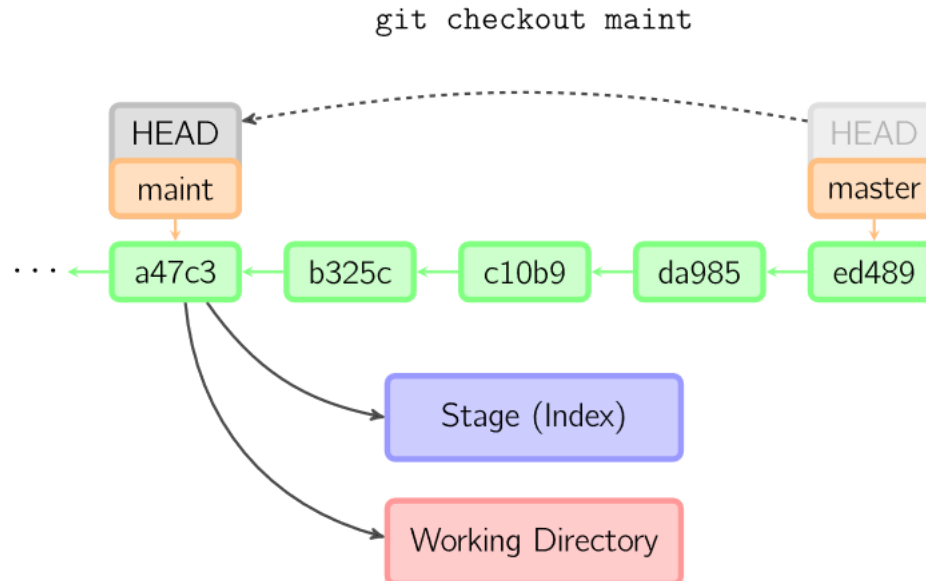
git checkout **branchName**

- Ändern des Branches
 - ◆ Checkout mit Name eines (lokalen) Branches
 - ◆ HEAD wird auf diesen Branch gesetzt
 - ◆ Daraufhin werden die Dateien im Index und im Arbeitsverzeichnis mit denen aus dem neuen Head überschreiben
- Beispiel: **git checkout maint**



git checkout **branchName**

- Jede Datei die im neuen Commit (a47c3 s.u.) existiert wird kopiert.
- Jede die im alten Commit (ed489) existiert, aber nicht im neuen, wird im Index und Arbeitsverzeichnis gelöscht.
- Alle weiteren werden nicht angefasst.
 - ◆ Dateien, die nicht unter Versionskontrolle stehen bleiben unberührt



git merge

Fast Forward Merge

3-Wege-Merge

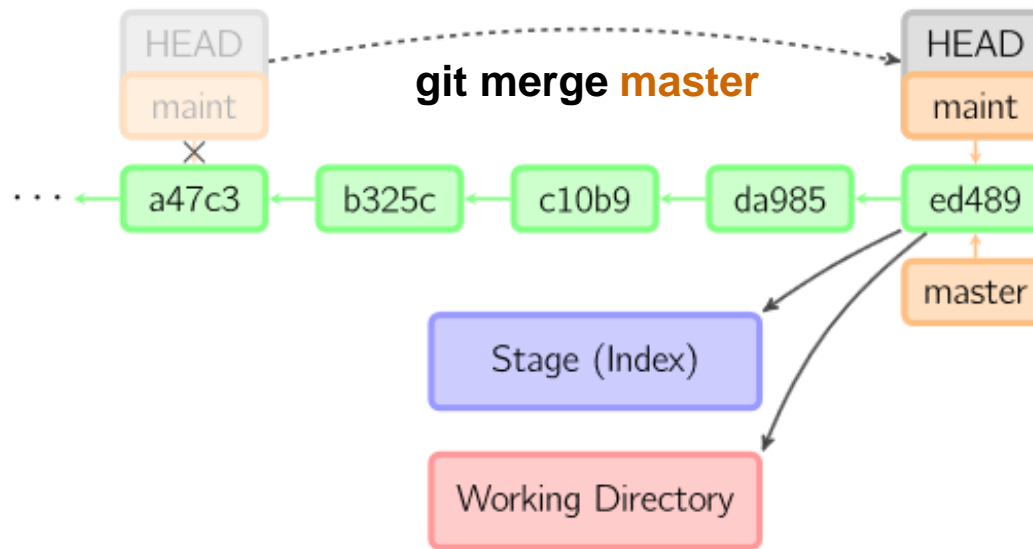
git merge **branchName**

Merge – allgemein

- im aktuellen branch neuen Commit erstellen, der Änderungen von commits aus einem anderen branch übernimmt

Merge – fast forward

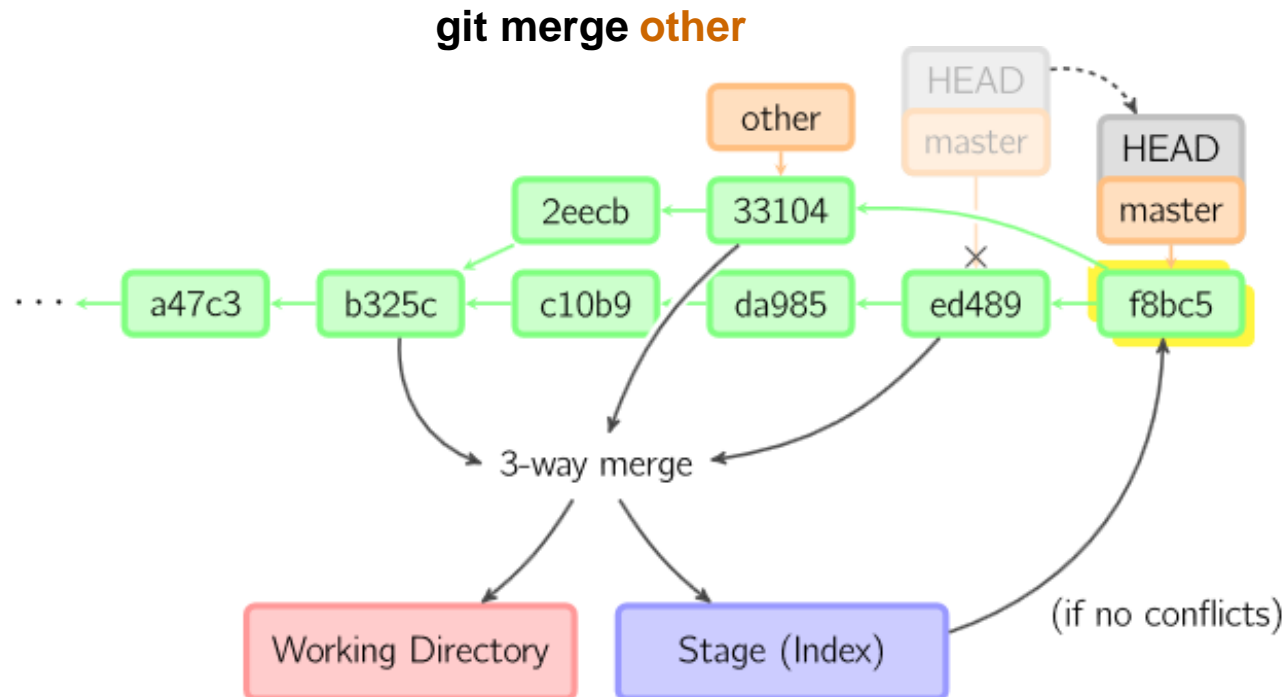
- falls eigener branch (hier **maint**) Vorläufer des anderen ist, einfach eigenen branch-Zeiger weiter setzen:



git merge **branchName**

Ansonsten wird ein **rekursiver 3-Wege-Merge** durchgeführt auf Basis

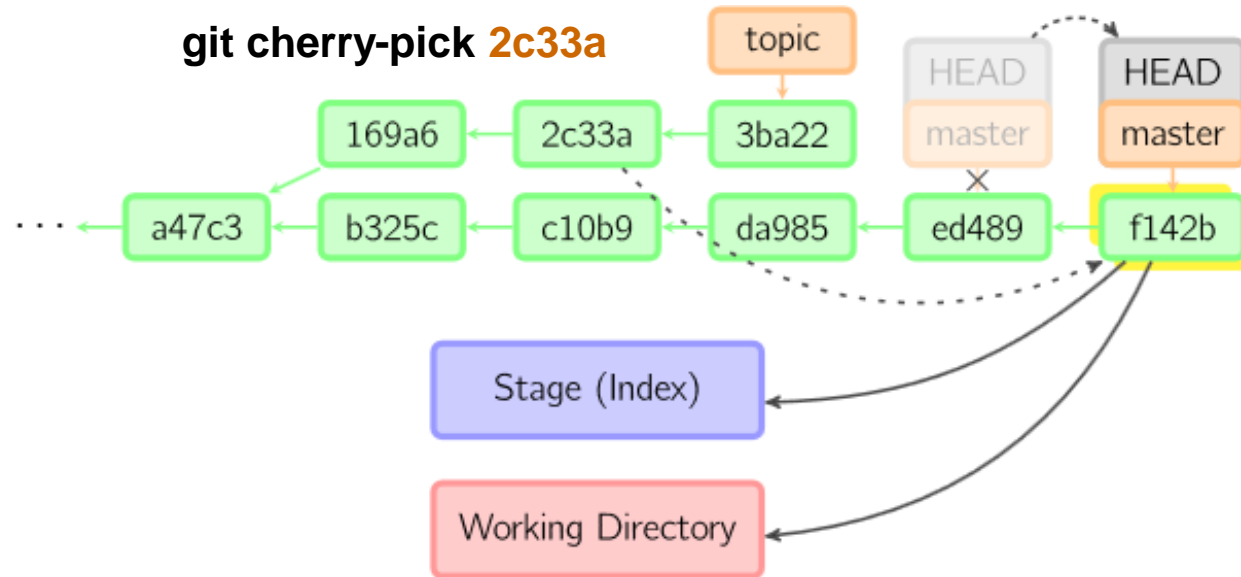
- des aktuellsten Commit im eigenen branch (**ed489**),
- des aktuellsten Commit im anderen Branch (**33104**) und
- ihres gemeinsamen Vorläufers (**b325c**)



cherry picking

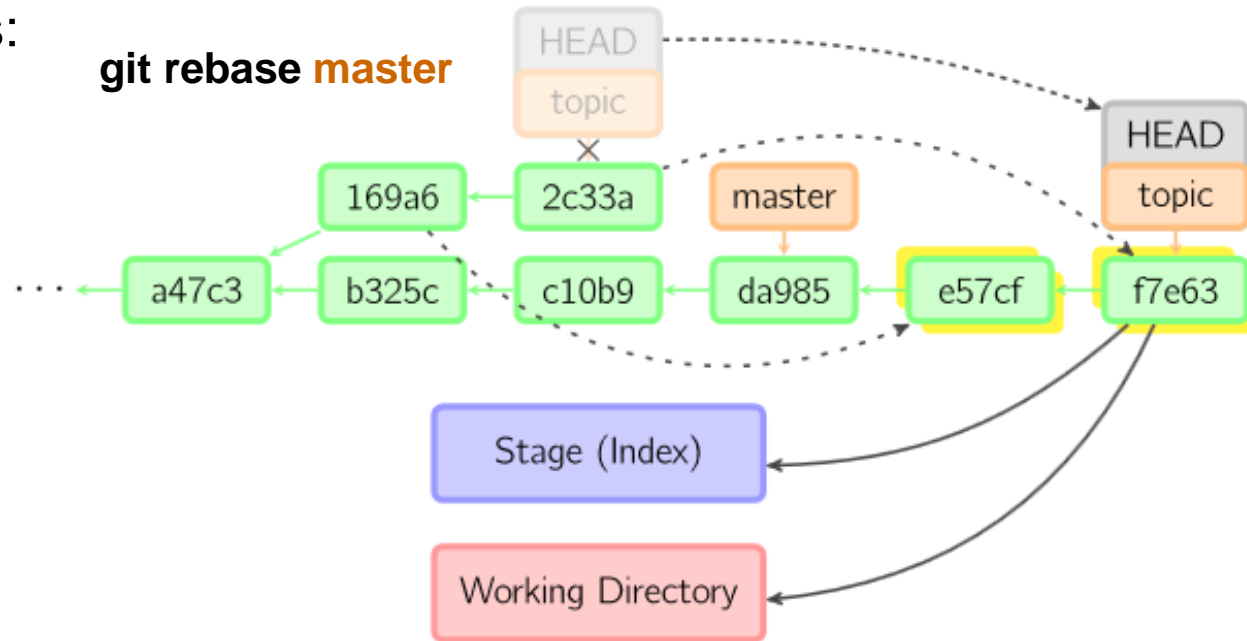
git cherry-pick commitID

- Der cherry-pick (Kirschen bzw. Rosinen rauspicken) Befehl „merged“ selektiv **genau einen Commit** in den aktuellen Branch:
 - ◆ die Änderungen dieses commits
 - ◆ seinen Kommentar
- Gegebenenfalls entstehen Konflikte (wie bei einem normalen merge)



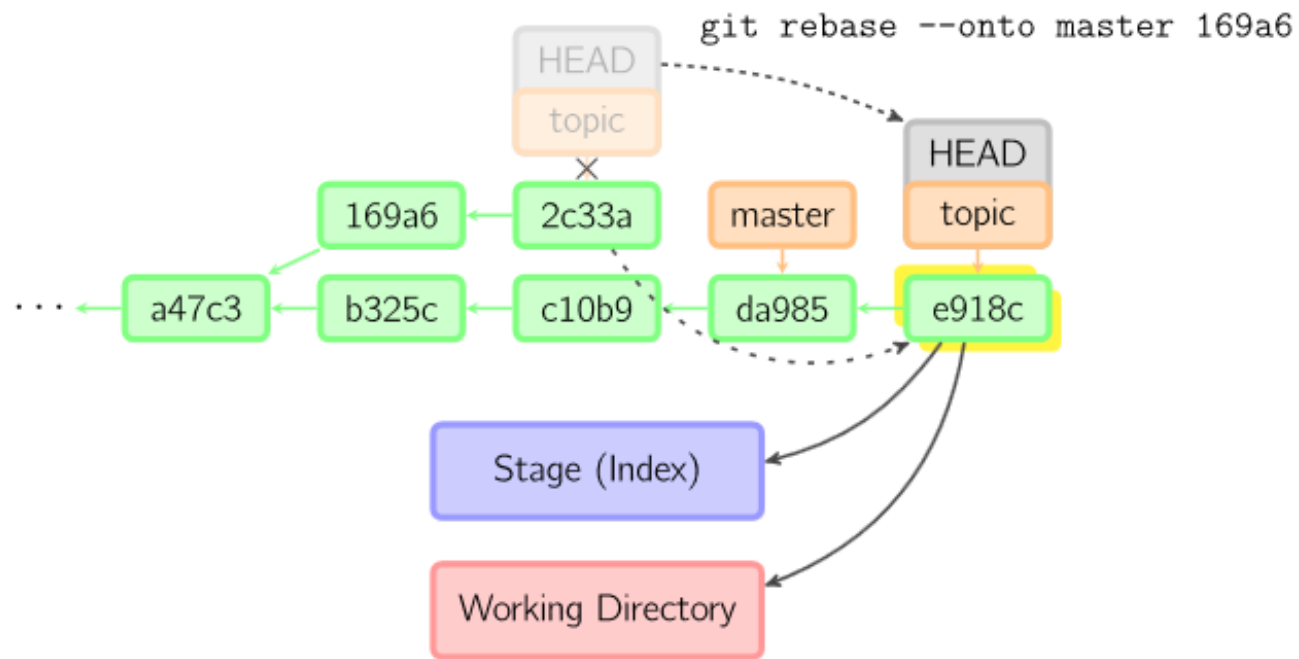
git rebase branch

- Rebase spielt die Commits des aktuellen Branches auf einen anderen Branch auf.
- Es hinterlässt eine Historie, an der nicht mehr erkennbar ist, dass die übertragenen commits aus einem anderen branch stammten.
- Es ist wie cherry picking **aller** commits seit der Verzweigung der beiden branches:



git rebase --onto **branch** **commitID**

Mit der Option `--onto` werden nur die Commits des aktuellen branches die **neuer als commitId** sind auf **branch** übertragen:

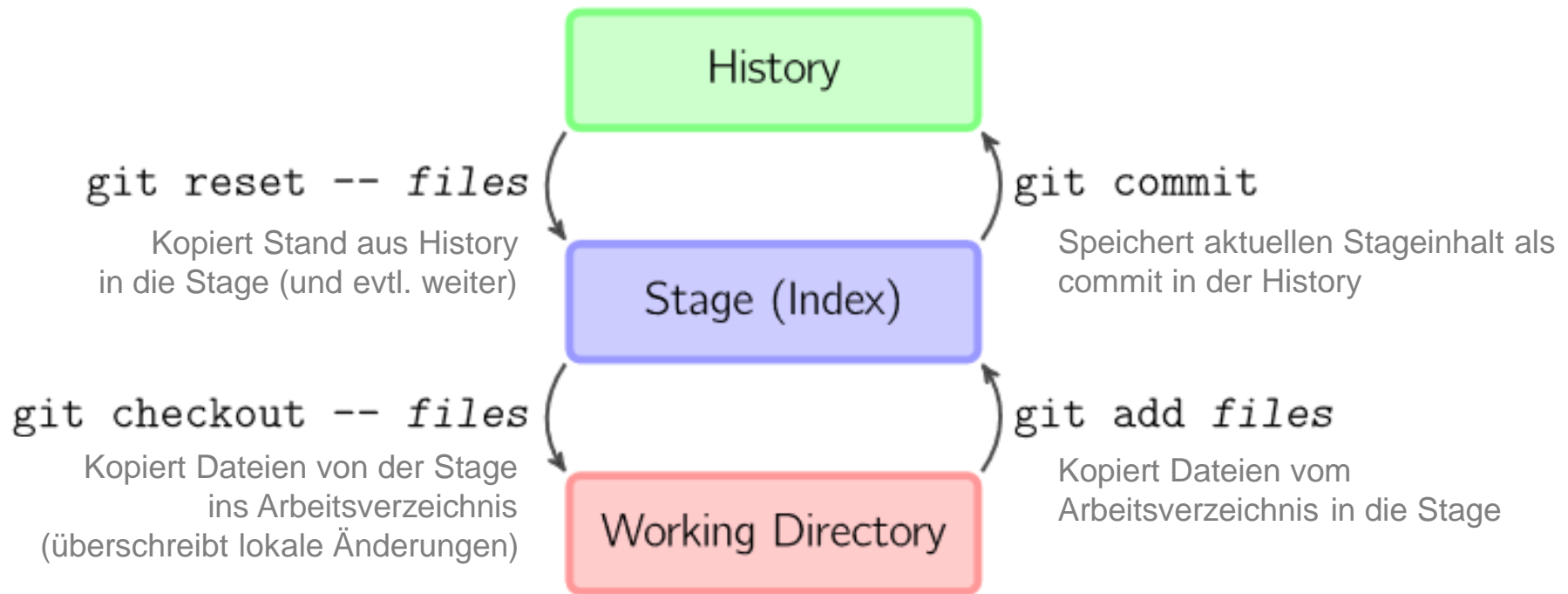


Änderungen rückgängig machen

- ▶ **Des Pudels Kern:** Fehler beseitigen!!!

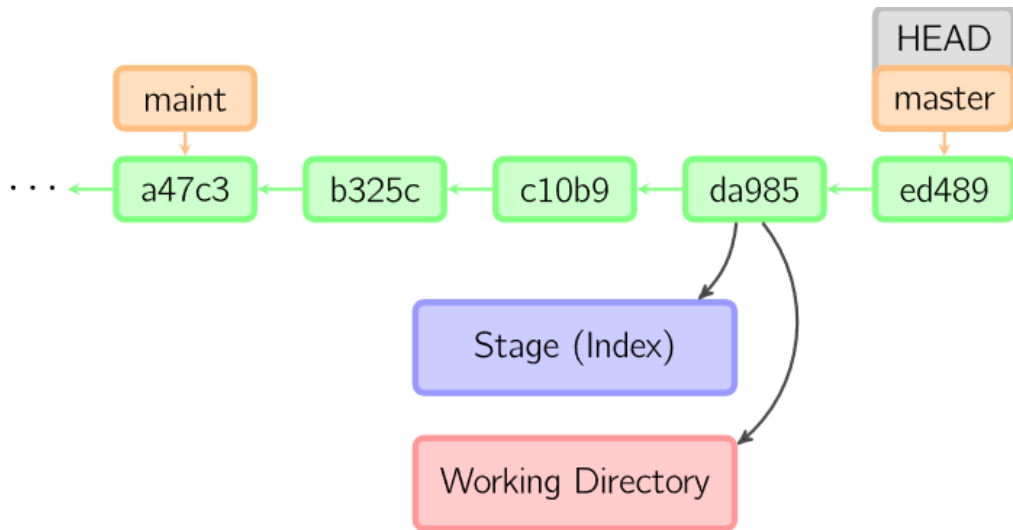
Was wollen wir evtl. rückgängig machen?

- Änderungen im working tree → Überschreiben mit altem Inhalt
 - ◆ Hinzufügen, Löschen oder Editieren von Dateien wird rückgängig gemacht
 - ◆ Zurücksetzen auf Stand eines bestimmten commit
 - ◆ **Achtung, hier kann es Datenverlust geben ☹, denn die neuesten Änderungen im Working Tree sind nirgends sonst gesichert (wenn sie es vorher nicht selbst getan haben)!!!**
- Stage → Unstage
 - ◆ Für nächsten commit vorgesehene Änderungen aus der Stage entfernen
 - ◆ Arbeitsbereich wird nicht verändert
- Commit → Uncommit
 - ◆ Aktueller Branch wird zurückgesetzt auf einen bestimmten commit
 - ◆ Neuere commits werden vergessen
 - ◆ Stage und Arbeitsbereich werden nicht verändert



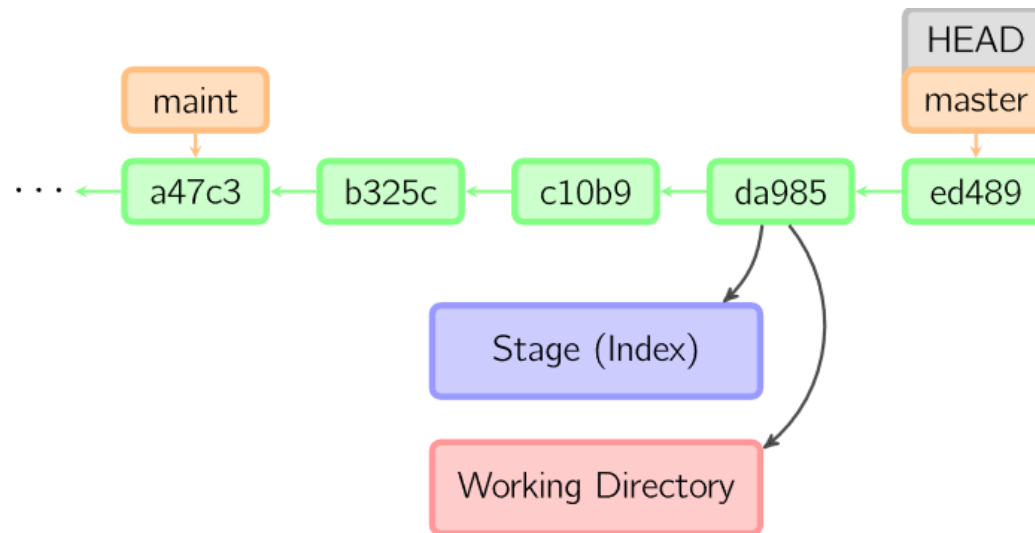
git checkout **commit** [paths]

- Beispiel: Checkout aller Dateien im letzten commit vor HEAD (= HEAD~)
- ◆ git checkout HEAD~



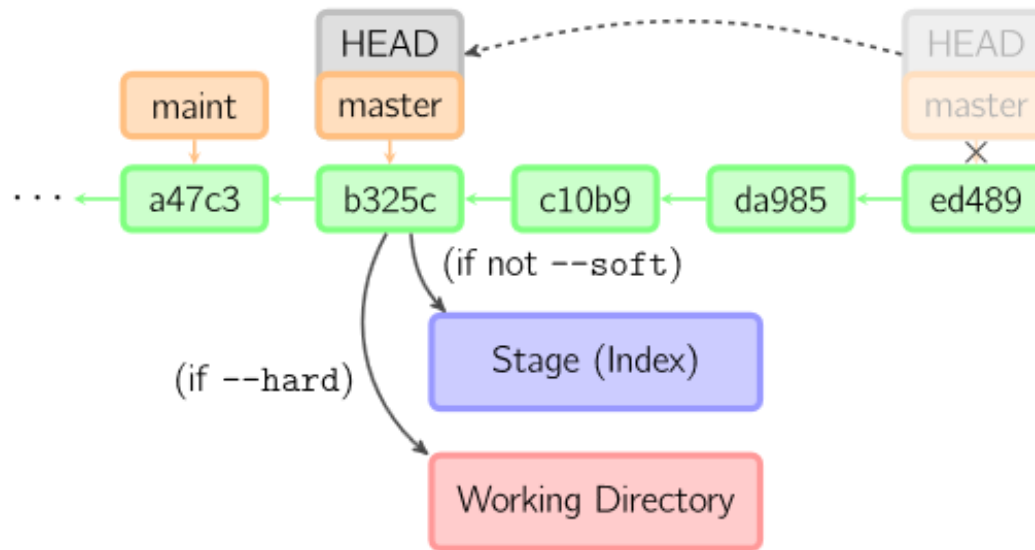
- **commit** angegeben
 - ◆ → Inhalt des **commit** überschreibt Stage und Arbeitsbereich
- Kein **commit** angegeben
 - ◆ → Inhalt der Stage überschreibt Arbeitsbereich
- Pfad angegeben → git überschreibt nur diese Datei(en)

git checkout **commit** [paths]



git reset [--option] commit

- Verschiebt aktuellen Branch zu **commit**
 - ◆ Im aktuellen Branch alle commits vergessen die neuer sind als **commit**
 - ◆ Implementierung: Aktueller Branch-Zeiger wird auf **commit** zurückgesetzt
 - “uncommit”
- Kann zusätzlich den Inhalt von commit in den Index (**--mixed**) oder in den Index und das Arbeitsverzeichnis (**--hard**) kopieren:



git reset HEAD~ 3

git reset [--option] commit

Je nach Option:

- **--soft**

- ◆ Im aktuellen Branch alle commits vergessen die neuer sind als **commit**
- ◆ Implementierung: Aktueller Branch-Zeiger wird auf **commit** zurückgesetzt
 - “uncommit”

- **--mixed** (*default wenn keine Option angegeben*)

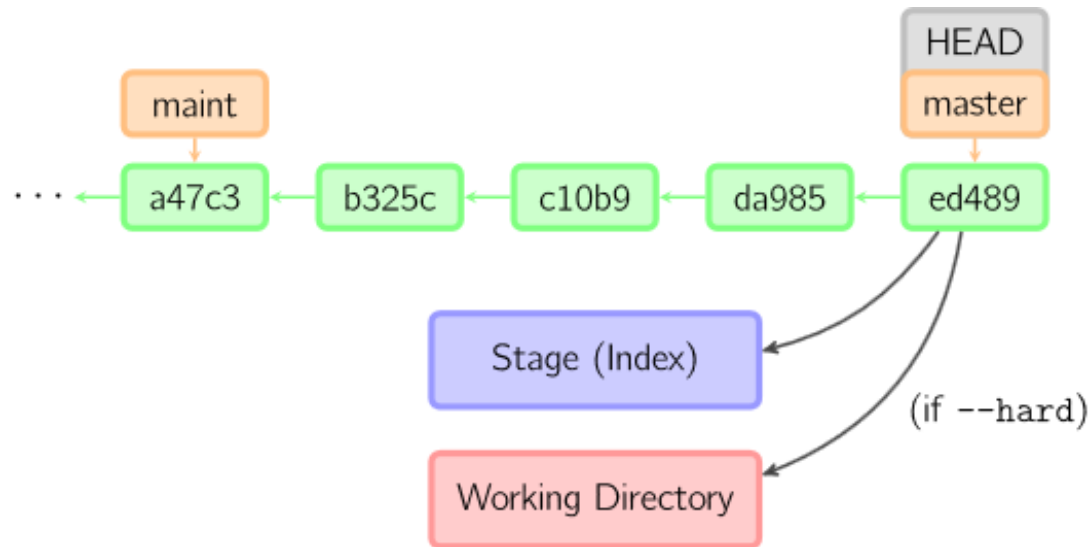
- ◆ Zusätzlich zu 1. Inhalt des **commit** in Stage übertragen
- ◆ Somit in Stage vorgemerkte Änderungen rückgängig machen
 - “uncommit and unstage”

- **--hard**

- ◆ Zusätzlich zu 2. Inhalt des Working Tree mit **commit**-Zustand überschreiben
- Alle Änderungen seit **commit** rückgängig machen ← **Achtung, Datenverlust möglich!!! ☹**

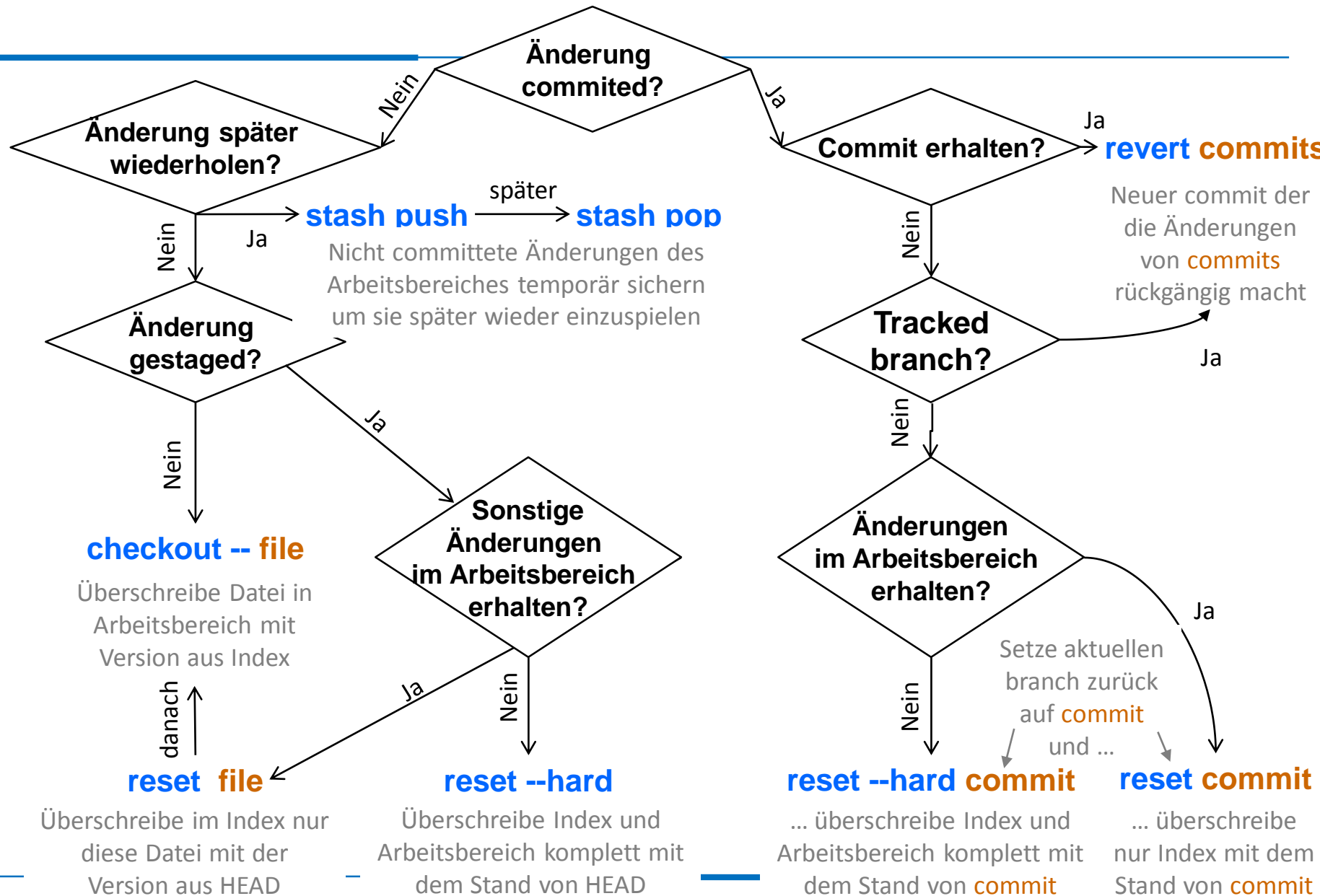
git reset filePath

- Gegenstück zu **git add filePath (= stage)**
 - ◆ Überschreib den Zustand einer Datei in der Stage/Index mit ihrem Zustand aus dem letzten Commit des aktuellen Branches (HEAD)
 - ◆ Damit wird die Stage-Operation Rückgängig gemacht (**= unstage**)
- Mit der Option **-- hard** wird auch der Zustand der Datei im Arbeitsbereich überschrieben ← **Achtung, Datenverlust!**



git reset myBrokenFile.c

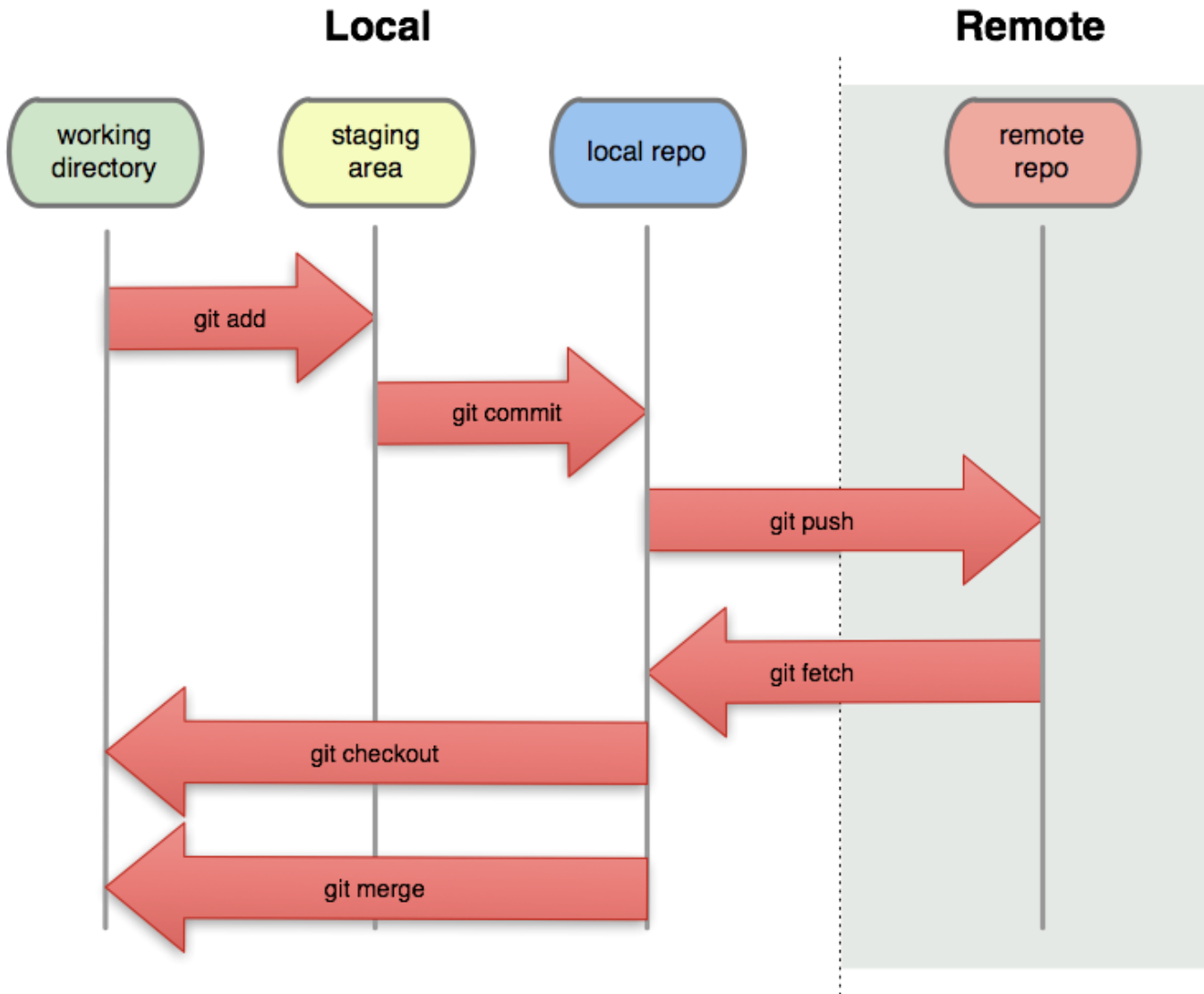
Änderung rückgängig machen



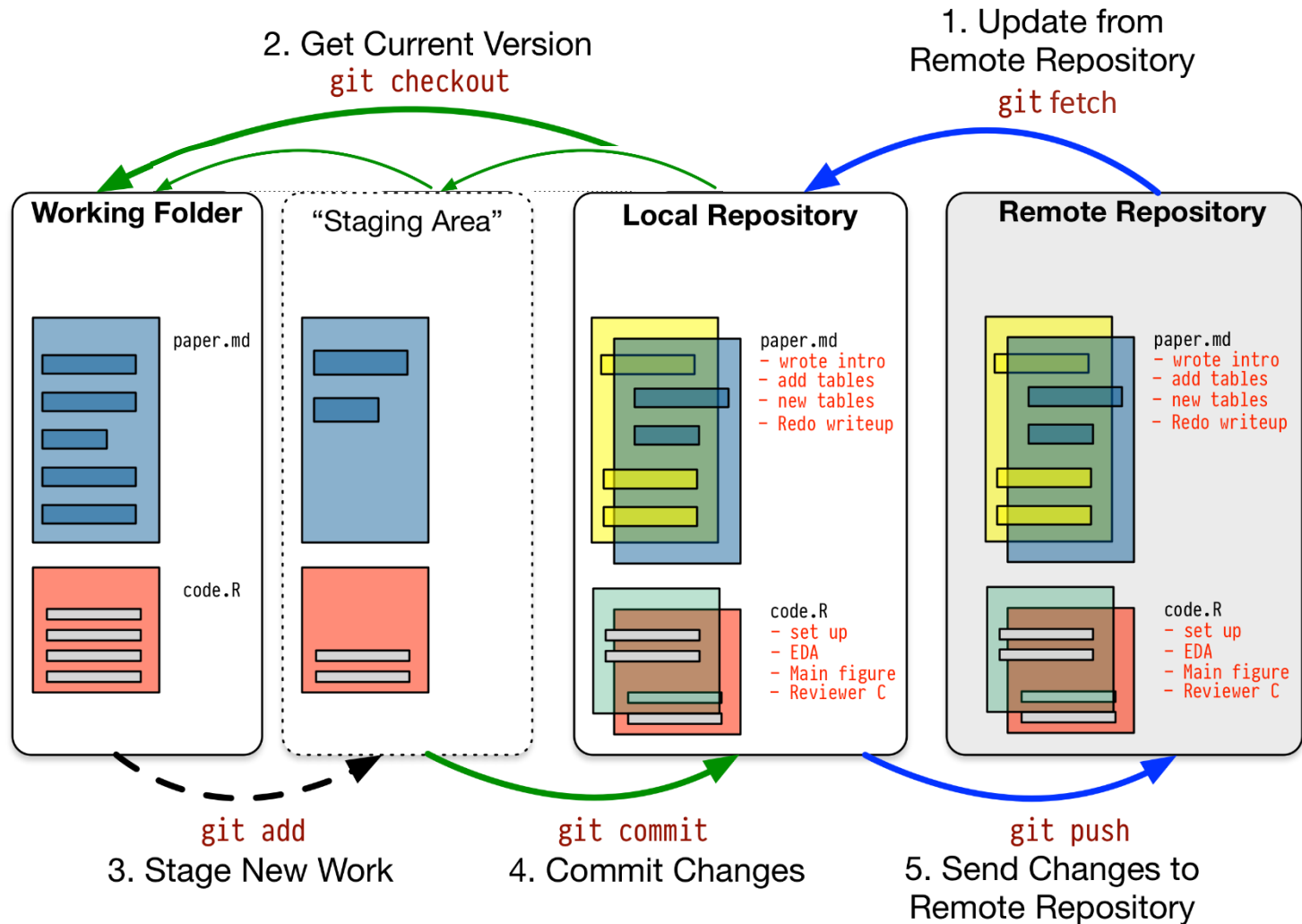
Verteiltes Arbeiten

- ▶ **Terminologie:** Remotes, origin, tracking
- ▶ **Operationen:** clone, fetch, pull, push

Gesamtbild: Lokal und verteilt



Gesamtbild



Plattformunabhängige GIT-Clients

- SmartGIT

- ◆ Kommerzielles Produkt aber für nichtkommerzielle Anwendung frei
- ◆ Stabil, ausgereift
- ◆ Leider nicht in Eclipse integriert
 - ⇒ Bei Wechsel nach Eclipse „F5“ / „Refresh“ nicht vergessen!
- ◆ <http://www.syntevo.com/smartgithg/>

- Egit

- ◆ Freies Plugin für Eclipse
- ◆ Sehr umfangreich, in Eclipse integriert
- ◆ Sehr ausführliches Tutorial (http://wiki.eclipse.org/EGit/User_Guide)
- ◆ <http://www.eclipse.org/egit/>

Zusammenfassung und Ausblick

Vergleich der Ansätze
Weiterführende Informationen

SCM-Ansätze und Werkzeuge ► Vergleich

Zentral (z.B. SVN)

- Operationen

- ◆ Verteilt

- ⇒ —

- ⇒ —

- ⇒ —

- ◆ Lokal, linear

- ⇒ Checkin

- ⇒ Checkout

- ⇒ Update (incl. 3-Wege-Merge)

- ⇒ Commit

- ◆ Lokal, branch management

- ⇒ Branch

- ⇒ Switch

- ⇒ Merge

- ⇒ —

Verteilt (z.B. GIT)

- Operationen

- ◆ Verteilt

- ⇒ Clone

- ⇒ Pull (incl. 3-Wege-Merge)

- ⇒ Push

- ◆ Lokal, linear

- ⇒ Create repository

- ⇒ —

- ⇒ — (implizit bei „pull“)

- ⇒ Commit

- ◆ Lokal, branch management

- ⇒ Branch

- ⇒ Switch

- ⇒ Merge

- ⇒ Cherry pick

SCM-Ansätze und Werkzeuge ► Vergleich

Zentral (z.B. SVN)

- Konzepte
 - ◆ Es gibt nur ein Repository
 - ◆ Interaktion nur über zentralen Server
 - ◆ Dateiweise Historie
 - ◆ Tracking von Umbenennungen

Verteilt (z.B. GIT)

- Konzepte
 - ◆ Gleichberechtigte Repositories
 - ◆ Zentrale Instanz organisatorisch möglich
 - ◆ Repository-weite Historie
 - ◆ Änderung hat eindeutige Id

Vorteile verteilter Versionskontrolle

- Eigene, lokale Versionskontrolle
 - ◆ Nicht erst dann einchecken, wenn wirklich alles läuft
 - ◆ Eigene inkrementelle Historie, Rollbacks auch lokal möglich
- Lokal Arbeiten ist sehr schnell
 - ◆ Diff, Commit und Revert sind rein lokal → kein Netzwerkverkehr nötig
- Branching und Merging ist leicht
 - ◆ „Branch = Clone“ ► Branch im repository oder clone des repository.
 - ◆ „Merge = Pull“ ► Beim pull wird automatisch ein merge durchgeführt
- Partielle Integration ist leicht
 - ◆ Entwickler können ihre Änderungen untereinander abgleichen
 - ◆ Änderung eines „zentralen“ Masters erst nach partieller Integration
- Wenig Management erforderlich
 - ◆ Schneller Start (einfach „clone“ oder „create repository“)
 - ◆ Nur der eigene Rechner muss andauernd laufen
 - ◆ Kein Usermanagement erforderlich

Nachteile verteilter Versionskontrolle

- Es gibt keine globalen Versionsnummern
 - ◆ Jedes Repository verwaltet seine eigene Änderungshistorie
 - ◆ Aber man kann Releases mit sinnvollen Namen taggen
- Es gibt keine „neueste Version“
 - ◆ Kann man durch ein dediziertes Repository emulieren
- Ein zentrales Backup ist dennoch sinnvoll
 - ◆ auch wenn gerne dagegen argumentiert wird
 - ◆ eventuell hat nicht jeder alle Änderungen mitbekommen, also kann man sich nicht auf andere Repositories als Backup verlassen

Listen weiterer SCM Werkzeuge

- Wikipedia

- ◆ http://en.wikipedia.org/wiki/List_of_revision_control_software
- ◆ Sehr umfangreiche Liste, mit oft guten Fortsetzungen zu den einzelnen Werkzeugen

- Wikipedia

- ◆ http://en.wikipedia.org/wiki/Comparison_of_revision_control_software
- ◆ Versuch eines tabellarischen Vergleichs der Systeme
- ◆ Teilweise gut, aber hoffnungslose Sisyphos-Arbeit...

- Andere

- ◆ <http://www.software-pointers.com/en-configuration-tools.html>
- ◆ <http://www.thefreecountry.com/programming/versioncontrol.shtml>
- ◆ http://www.dmoz.org/Computers/Software/Configuration_Management/Tools/

Vergleiche von SCM Werkzeuge

- Wikipedia (allgemein)
 - ◆ http://en.wikipedia.org/wiki/Comparison_of_revision_control_software
 - ◆ Versuch eines tabellarischen Vergleichs der Systeme
 - ◆ Teilweise gut, aber hoffnungslose Sisyphos-Arbeit...
- Wikipedia (nur open source)
 - ◆ http://en.wikipedia.org/wiki/Comparison_of_open_source_configuration_management_software
- Andere neutrale Vergleiche
 - ◆ <http://better-scm.berlios.de/comparison/> (Stand 29. Feb. 2008)
- Nicht ganz unparteiische Vergleiche...
 - ◆ http://www.relisoft.com/co_op/vcs_compare.html
 - ◆ http://www.accurev.com/scm_comparisons.html