

# Domain Specific Languages mit Xtext – Anwendung in der klinischen Forschung

Roland Mücke, itemis AG  
23.09.2011

## Vorstellung der itemis AG

- IT-Dienstleister
  - Beratung, Coaching, Werkzeugentwicklung
- itemis AG gegründet im Jahre 2003
- 130 Mitarbeiter an mehreren Standorten in Deutschland, Europa und Amerika
- Schwerpunkte
  - Modellgetriebene Softwareentwicklung (MDSD)
  - Enterprise und Embedded
  - Agile Softwareentwicklung mit Scrum
  - Mobile Applications
- Strategic Member der Eclipse Foundation



## Einstieg

- Während der Vorbereitungsphase einer klinischen Studie: Spezifizieren von Merkmalen (Items), die in der Studie erhoben werden sollen
- Hunderte von Items in einer Studie mit Dutzenden von Eigenschaften, PrüfregeIn und Bedingungen
  - Variablennamen, Wertebereiche, Merkmalsausprägungen, Datentypen, Eingabeformate, Namen auf dem Erhebungsbogen, Maßeinheiten etc.
- Einrichten einer Studie mit einem Studienmanagementsystem (SMS)
- Anlegen oder Auswählen der gewünschten Items über die Oberfläche des SMS
  - Verwenden von standardisierten Items oder Anlegen neuer Items
- Import von Item-Spezifikationen in das SMS
  - Verwenden von Standardformaten z. B. CDISC ODM oder proprietären Formaten (XML?)

## Sonderwünsche

- Besondere Anforderungen an die Zusammenstellung von Items einer Studie?
  - „Bei multizentrischen Studien müssen für Laborwerte die Normbereiche separat erfasst werden.“
  - „Die Variablennamen in Studien für Kunde XY müssen immer mit XY anfangen.“
- Zusammenarbeit mit Projektpartnern ohne Zugriff auf das SMS?
  - Austausch via E-Mail
  - Versionierung und Vergleich von Änderungen an der Spezifikation
- Item-Spezifikation mit Terminologie und Konzepten, die vom SMS abweichen?
  - „Items sind bei uns nicht in Modulen sondern in Merkmalsgruppen zusammengefasst.“
  - „Die Auswahl der Maßeinheit eines Laborwertes soll kein eigenständiges Item sein sondern als Unter-Item des Laborwert-Items spezifiziert werden.“

## Verbiegen

- Anpassen oder Erweitern des SMS, um Anforderungen umzusetzen
  - Teuer oder überhaupt machbar?
- Trainieren der Mitarbeiter, um sich der Terminologie und den Konzepten des SMS zu anpassen
  - Verwerfen oder Umwandeln von etablierten Regeln und Arbeitsprozeduren?
- Bedingungsloses befolgen von Standards für die Interoperabilität von Softwaresystemen (z. B. dem XML-basierten CDISC ODM)
  - Fachanwender mit Aspekten technischer Interoperabilität belasten?
- Oder: Man bietet den Fachanwendern ein Werkzeug, das auf sie maßgeschneidert ist, und nimmt Anpassungen und Umwandlungen für SMS und Interoperabilität unter der Haube vor.

## Unsere eigene Studienspezifikation

### Studienspezifikation

#### Studieninformation

**Kurztitel:** "Beispielstudie"  
**Name der Studie:** "Beispielstudie zum Testen der DSL"  
**EudracCT-Nummer:** "2010-123456-99"  
**Datumsformat:** "DD.MM.yyyy"  
**Patienten-ID einheitlich in:** *Klinik*  
**Art der Patienten-ID:** *numerisch*  
 falls "DDE", Ersteingabe sichtbar: *Ja*  
**Purge Discrepancies:** *Nie*  
**Audit-Trial:** *nein*

#### Verwendete Bibliotheken

**Bibliothek** "../Beispielbibliothek.itemlib"  
**Bibliothek** "Spezial-Bibo.itemlib"

#### Studienaufbau

```
Event Baseline als BASELINE
  Page BasisUntersuchung1 als B1
    Modul SozidemographischeAngaben als S vertikal
      Item SystolischerBlutdruck als SSYS pflicht
      Item DiastolischerBlutdruck als SDIA pflicht
      Item Untersuchungsdatum als SDATUM pflicht
      Item Alter als SAGE
      Item Geschlecht als SSEX
      Item Untersuchungsdauer als STIME
      Item Notiz als SNOTIZ
      Item Beispiele.bdSysStandard als SF00
      Item Beispiele.bdSysStandard als SBAR
    Page BasisUntersuchung2 als B2
      Modul AktuelleMedikation als M horizontal
        Item Notiz als MNOTIZ
        Item Psychologie.PsychotherapeutischeBehandlung mit
          Beginn als MPTH1VOM,
          Ende als MPTH1BIS,
          Dauer als MPTH1DAUER
  Event Rando als RANDO
```

### Definitionen

#### Event-Definitionen

**Event** Baseline "Basis-Untersuchung"  
**Event** Rando "Randomisation"

#### Page-Definitionen

**Page** BasisUntersuchung1 "Basis-Untersuchung Seite 1"  
**Page** BasisUntersuchung2 "Basis-Untersuchung Seite 2"

#### Modul-Definitionen

**Modul** SozidemographischeAngaben "Sozidemographische Angaben"  
**Modul** AktuelleMedikation "Aktuelle Medikation"

### Item-Definitionen

```
Item SystolischerBlutdruck "Blutdruck systolisch" {
  Format: Number(4.1)
}
Item DiastolischerBlutdruck "Blutdruck diastolisch" {
  Format: Number(4.1)
  Range: Warning[40, 160]
  Prompt: "diastolisch"
  Hilfetext: "diastolischer Blutdruck im Stehen gemessen"
}
Item Untersuchungsdatum "Untersuchungsdatum" {
  Format: Date
  caDSR-ID: X3000
  Range: Check[1940, "aktuelles Datum"]
}
Item Untersuchungsdauer "Dauer der Untersuchung" {
  Format: Time
  caDSR-ID: X12392
}
Item Notiz "Bemerkung" {
  Format: Varchar2(240)
  Praeparat-ID: 12392-0
}
Item Alter "Alter" {
  Format: Number(3.0)
  Range:
    Warning[18, 60] "Warnung! Wert liegt ausserhalb des Prüfbereiches von [18-60]."  

    Error[16, 75] "Fehler! Wert liegt ausserhalb des Prüfbereiches von [16-75]."
```

### Codelisten-Definitionen

```
CodeListe CLYN "ja / nein" numerisch {
  1 : "ja"
  2 : "nein"
  -1 : "unbekannt"
}
CodeListe SSEX "Geschlecht" text {
  0 : "männlich" Praeparat-ID: 00456-0
  1 : "weiblich"
  -1 : "unbekannt"
}
```

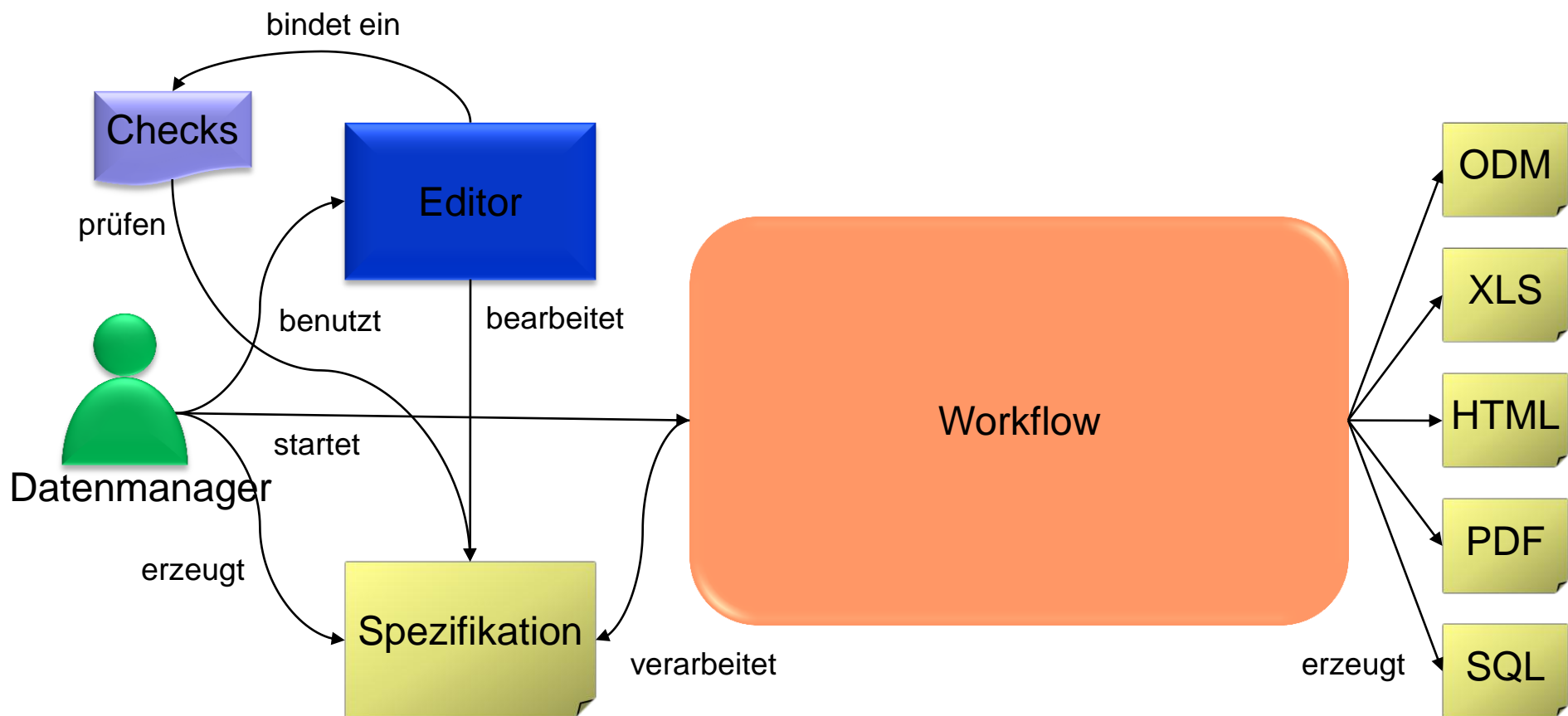
### Dictionary-Definitionen

```
Dictionary COUNTRY "Land" {
  Spalten:
    CODE 3 Text
    DESCRIPT 100 Text
  Werte:
    "DE", "Deutschland"
    "EN", "Großbritannien"
    "FR", "Frankreich"
    "CH", "Schweiz"
    "AT", "Österreich"
}
```

## Vorteile einer textueller Spezifikation

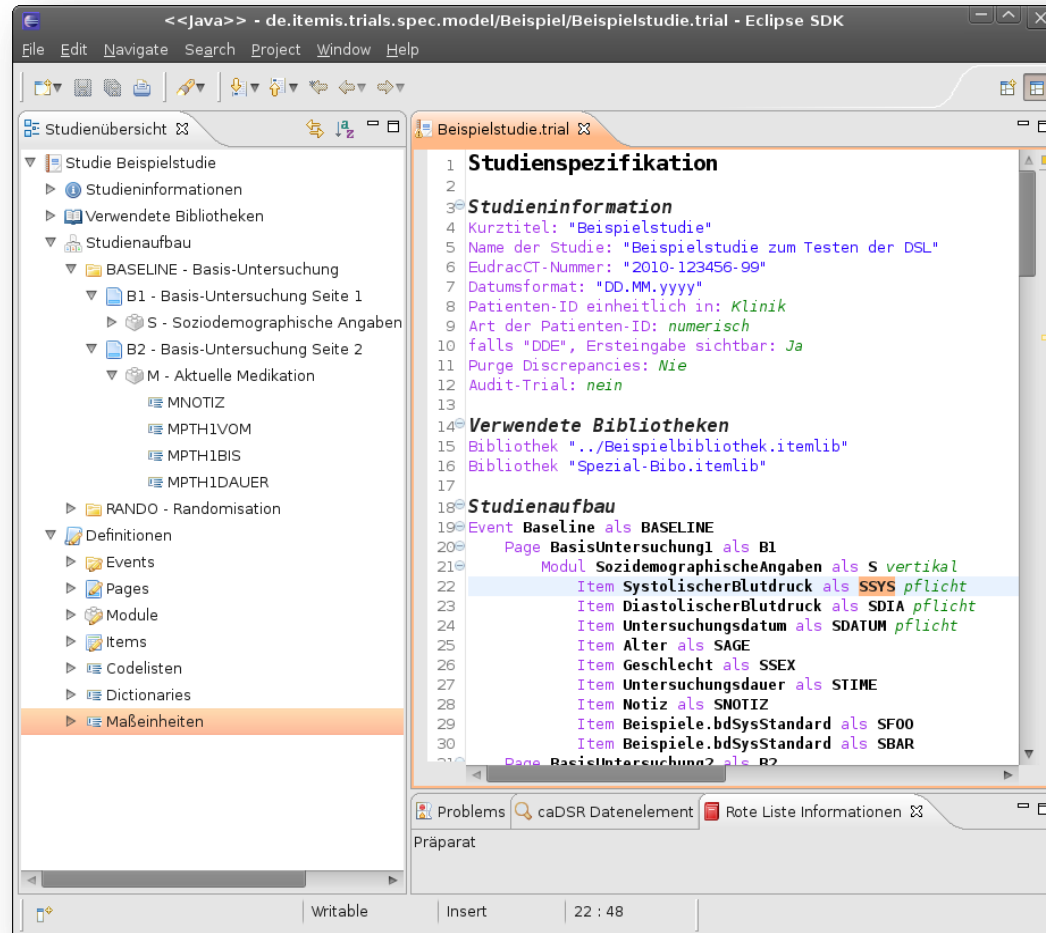
- Orientierung und Navigation in großen Spezifikationsdokumenten einfach
- Gleichzeitige Bearbeitung durch mehrere Personen möglich
  - Automatisches Verschmelzen von Änderungen
- Versionierung und Anzeige von Änderungen zwischen Versionen
- Aufteilen einer Spezifikation auf mehrere Dateien
  - Bessere Handhabbarkeit großer Spezifikationen
  - Referenzen zwischen den Dateien möglich
- Auch ohne besondere Software lesbar
  - Es handelt sich um simple Textdokumente

# Arbeitsablauf aus Sicht des Datenmanagers





## Ein Editor



## Unterstützung des Editors beim Spezifizieren

- **Farbliche Hervorheben von Schlüsselwörtern (Syntax Highlighting)**
- Navigation in einer Baumansicht (Outline View)
- Verlinken von Spezifikationselementen (Crosslink Navigation)
- Auto-Vervollständigung (Code Completion)
- Checks für Syntax und Semantik
- Ausblenden von nicht benötigten Abschnitten (Code Folding)
- Referenzierung externer Terminologien

### Codelisten-Definitionen

```

Codeliste CLYN "ja / nein" numerisch {
  1 : "ja"
  2 : "nein"
  -1 : "unbekannt"
}
Codeliste SSEX "Geschlecht" text {
  0 : "männlich"
  1 : "weiblich"
  -1 : "unbekannt"
}
  
```

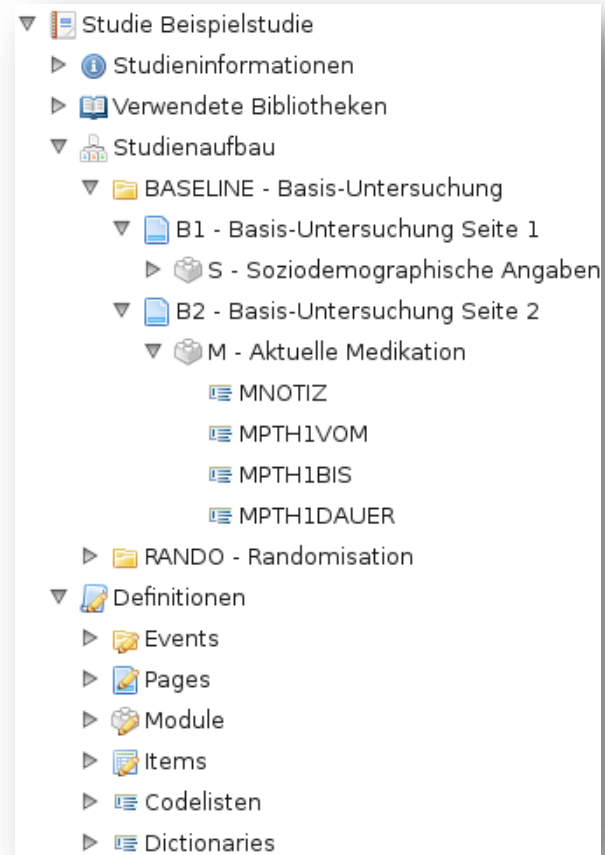
### Dictionary-Definitionen

```

Dictionary COUNTRY "Land" {
  Spalten:
    CODE 3 Text
    DESCRIPT 100 Text
  Werte:
    "DE", "Deutschland"
    "EN", "Großbritannien"
    "FR", "Frankreich"
    "CH", "Schweiz"
    "AT", "Österreich"
}
  
```

## Unterstützung des Editors beim Spezifizieren

- Farbliche Hervorheben von Schlüsselwörtern (Syntax Highlighting)
- **Navigation in einer Baumansicht (Outline View)**
- Verlinken von Spezifikationselementen (Crosslink Navigation)
- Auto-Vervollständigung (Code Completion)
- Checks für Syntax und Semantik
- Ausblenden von nicht benötigten Abschnitten (Code Folding)
- Referenzierung externer Terminologien



# Unterstützung des Editors beim Spezifizieren

- Farbliche Hervorheben von Schlüsselwörtern (Syntax Highlighting)
- Navigation in einer Baumansicht (Outline View)
- **Verlinken von Spezifikationselementen (Crosslink Navigation)**
- Auto-Vervollständigung (Code Completion)
- Checks für Syntax und Semantik
- Ausblenden von nicht benötigten Abschnitten (Code Folding)
- Referenzierung externer Terminologien

```
Modul AktuelleMedikation als M horizontal
Item Notiz als MNOTIZ
Item Psychologie.PsychotherapeutischeBehandlung mit
  Beginn als MPTH1VOM,
  Ende als MPTH1BIS,
```

## Unterstützung des Editors beim Spezifizieren

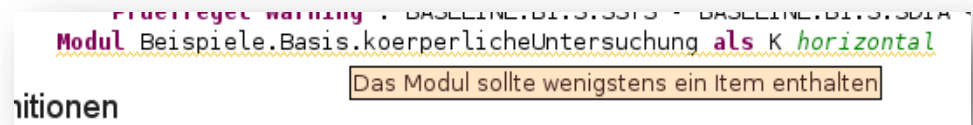
- Farbliche Hervorheben von Schlüsselwörtern (Syntax Highlighting)
- Navigation in einer Baumansicht (Outline View)
- Verlinken von Spezifikationselementen (Crosslink Navigation)
- **Auto-Vervollständigung (Code Completion)**
- Checks für Syntax und Semantik
- Ausblenden von nicht benötigten Abschnitten (Code Folding)
- Referenzierung externer Terminologien

The screenshot displays a code editor window titled "Studienaufbau". The code is structured hierarchically with the following elements:

- Event** Baseline als BASELINE
- Page** BasisUntersuchung1 als B1
  - Modül** SozidemographischeAngaben als S *vertikal*
    - Item** SystolischerBlutdruck als SSYS *pflicht*
    - Item** DiastolischerBlutdruck als SDIA *pflicht*
    - Item** Untersuchungsdatum als SDATUM *pflicht*
    - Item** Alter als SAGE
    - Item** Geschlecht als SSEX
    - Item** Untersuchungsdauer als STIME
    - Item** Notiz als SNOTIZ
- Page** BasisUnt
  - Modül** Akt
    - Item** Coolness
    - Item** DiastolischerBlutdruck
- Event** Rando als R
  - Page** BasisUnt
    - Modül** Soz
      - Item** PsychotherapeutischeBehandlung - Psychologie.Psychothe
      - Item** SystolischerBlutdruck
      - Item** Untersuchungsdatum

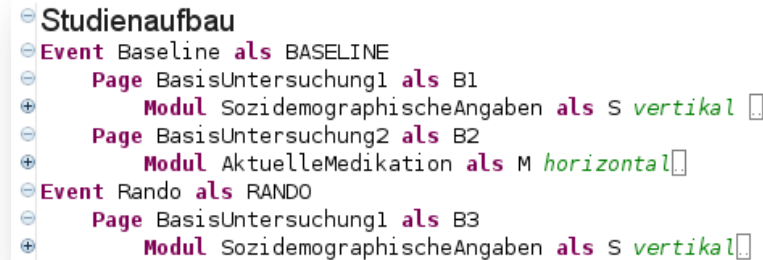
## Unterstützung des Editors beim Spezifizieren

- Farbliche Hervorheben von Schlüsselwörtern (Syntax Highlighting)
- Navigation in einer Baumansicht (Outline View)
- Verlinken von Spezifikationselementen (Crosslink Navigation)
- Auto-Vervollständigung (Code Completion)
- **Checks für Syntax und Semantik**
- Ausblenden von nicht benötigten Abschnitten (Code Folding)
- Referenzierung externer Terminologien



# Unterstützung des Editors beim Spezifizieren

- Farbliche Hervorheben von Schlüsselwörtern (Syntax Highlighting)
- Navigation in einer Baumansicht (Outline View)
- Verlinken von Spezifikationselementen (Crosslink Navigation)
- Auto-Vervollständigung (Code Completion)
- Checks für Syntax und Semantik
- **Ausblenden von nicht benötigten Abschnitten (Code Folding)**
- Referenzierung externer Terminologien



```
Studienaufbau
- Event Baseline als BASELINE
  - Page BasisUntersuchung1 als B1
    + Modul SozidemographischeAngaben als S vertikal
  + Page BasisUntersuchung2 als B2
    + Modul AktuelleMedikation als M horizontal
- Event Rando als RANDO
  - Page BasisUntersuchung1 als B3
    + Modul SozidemographischeAngaben als S vertikal
```

## Unterstützung des Editors beim Spezifizieren

- Farbliche Hervorheben von Schlüsselwörtern (Syntax Highlighting)
- Navigation in einer Baumansicht (Outline View)
- Verlinken von Spezifikationselementen (Crosslink Navigation)
- Auto-Vervollständigung (Code Completion)
- Checks für Syntax und Semantik
- Ausblenden von nicht benötigten Abschnitten (Code Folding)
- **Referenzierung externer Terminologien**

The screenshot shows a code editor window with the following content:

```

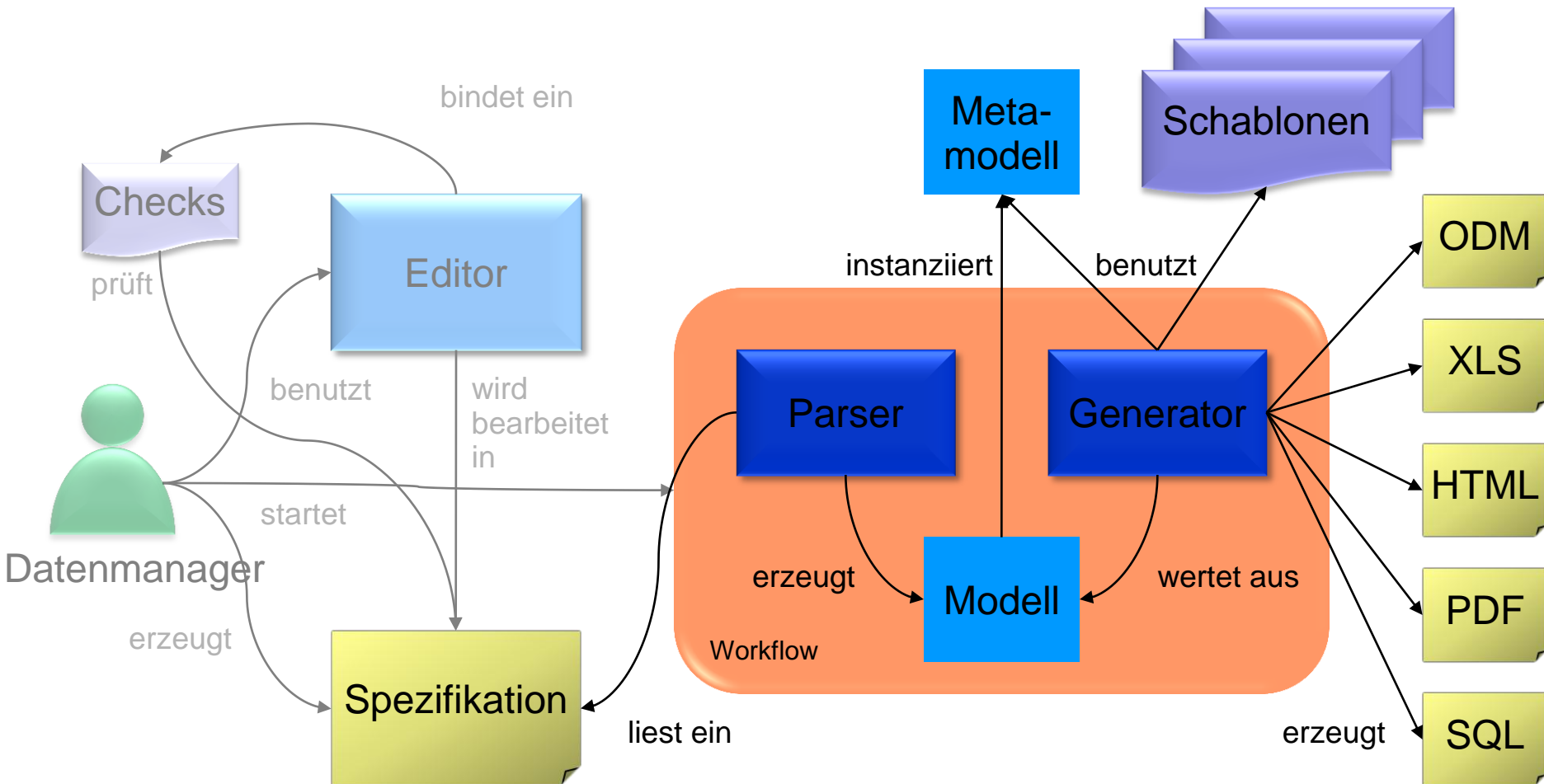
80 Item Untersuchungsdauer "Dauer der Untersuchung" {
81     Format: Time
82     caDSR-ID: X12392
83 }
84 Item Primaermedikament "Zuerst verabreicht" {
85     Format: Varchar2
86     Praeparat-ID: 12392-0
87 }

```

Below the code editor, a breadcrumb trail is visible: `caDSR Datenelement` with a search icon.



## Dem Workflow unter die Haube geschaut



## Das Metamodell als EMF Ecore

The screenshot displays the Eclipse IDE interface for editing an Ecore model. The top pane shows the project structure for 'TrialDSL.ecore', with the 'definitionen : Definitionen' package selected. The bottom pane shows the Properties view for this package, with the following data:

Property	Value
Changeable	true
Container	false
Containment	true
Default Value Literal	
Derived	false
EKeys	
EOpposite	
EType	Definitionen

- „Standard-Format“ des Eclipse Modeling Frameworks zur Beschreibung von Metamodellen.
- Technische Beschreibung dessen, was eine Studienspezifikation enthalten kann oder muss.
- Mühsam zu erstellen und zu pflegen.
- Warum nicht auch textuell definieren?

# Grammatik einer Domain Specific Language (DSL)

```
grammar de.itemis.trials.spec.TrialDSL with de.itemis.trials.item.ItemDSL

generate trialDSL "http://www.itemis.de/trials/TrialDSL"

import "http://www.eclipse.org/emf/2002/Ecore" as.ecore
import "http://www.itemis.de/trials/LibraryDSL" as.itemlib
import "http://www.itemis.de/trials/ItemDSL" as.itemdef

/*****
 * Hauptstruktur
 *****/
Spezifikation:
    "Studienspezifikation"
        info=Studieninformation
        (bibliotheken=Bibliotheken)?
        aufbau=Studienaufbau
        (definitionen=Definitionen)?
;

Studieninformation: {Studieninformation}
    "Studieninformation" (informationen+=StudieninformationsElement)*;

Bibliotheken: {Bibliotheken}
    "Verwendete Bibliotheken" (imports+=Import)*;

Studienaufbau: {Studienaufbau}
    "Studienaufbau" (events+=EventReferenz)*;
```

## Grammatik einer Domain Specific Language (DSL)

```

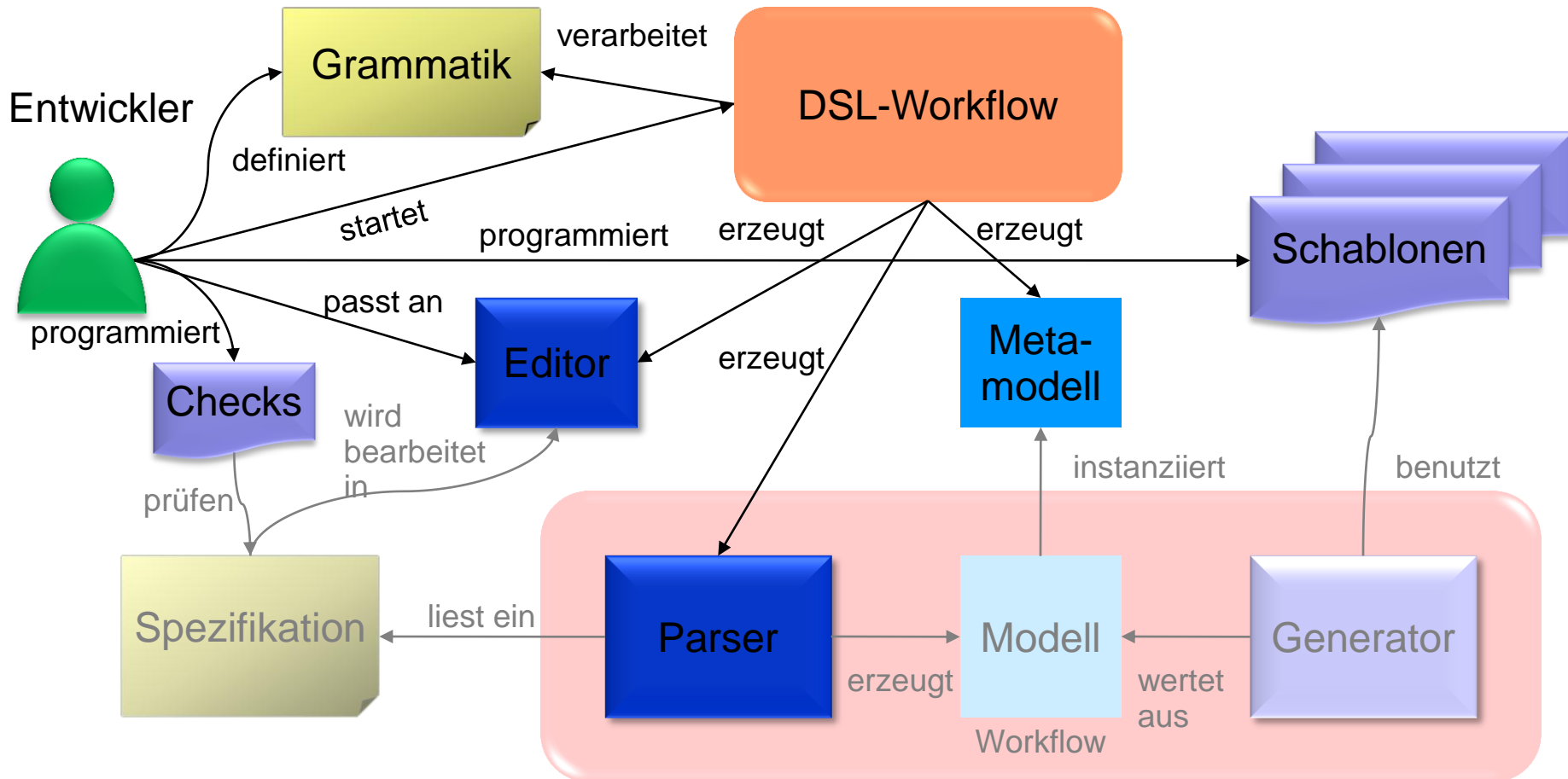
ItemDefinition:
  "Item" name=ID bezeichnung=STRING "{"
    (spec=ItemSpezifikation | (teilItems+=TeilItemDefinition)+)
  "}"
;

TeilItemDefinition:
  "Teil-Item" name=ID bezeichnung=STRING "{"
    spec=ItemSpezifikation
  "}"
;

ItemSpezifikation:
  "Format:" formatTyp=ItemFormatTyp ("(" formatParam1=INT ( "." formatParam2=INT)?)")?
  ("caDSR-ID:" caDSRRef=[CaDSRIdentifizier|ID])?
  ("Praeparat-ID:" roteListePraepId=[RoteListePraepIdentifizier|ROTE_LISTE_ID])?
  (
    (
      ("Codelist:" codelist=[CodelistDefinition|FQN] darstellungsTyp=CodelistDictionaryDarstellungsTyp) |
      ("Dictionary:" dictionary=[DictionaryDefinition|FQN] darstellungsTyp=CodelistDictionaryDarstellungsTyp)
    ) | (
      "Range:" (ranges+=ItemRange)+
    )
  )?
  (itemInfos+=ItemInfo)*
  ("Pruefregeln:" (pruefRegeln+=Pruefregel)+)?
;

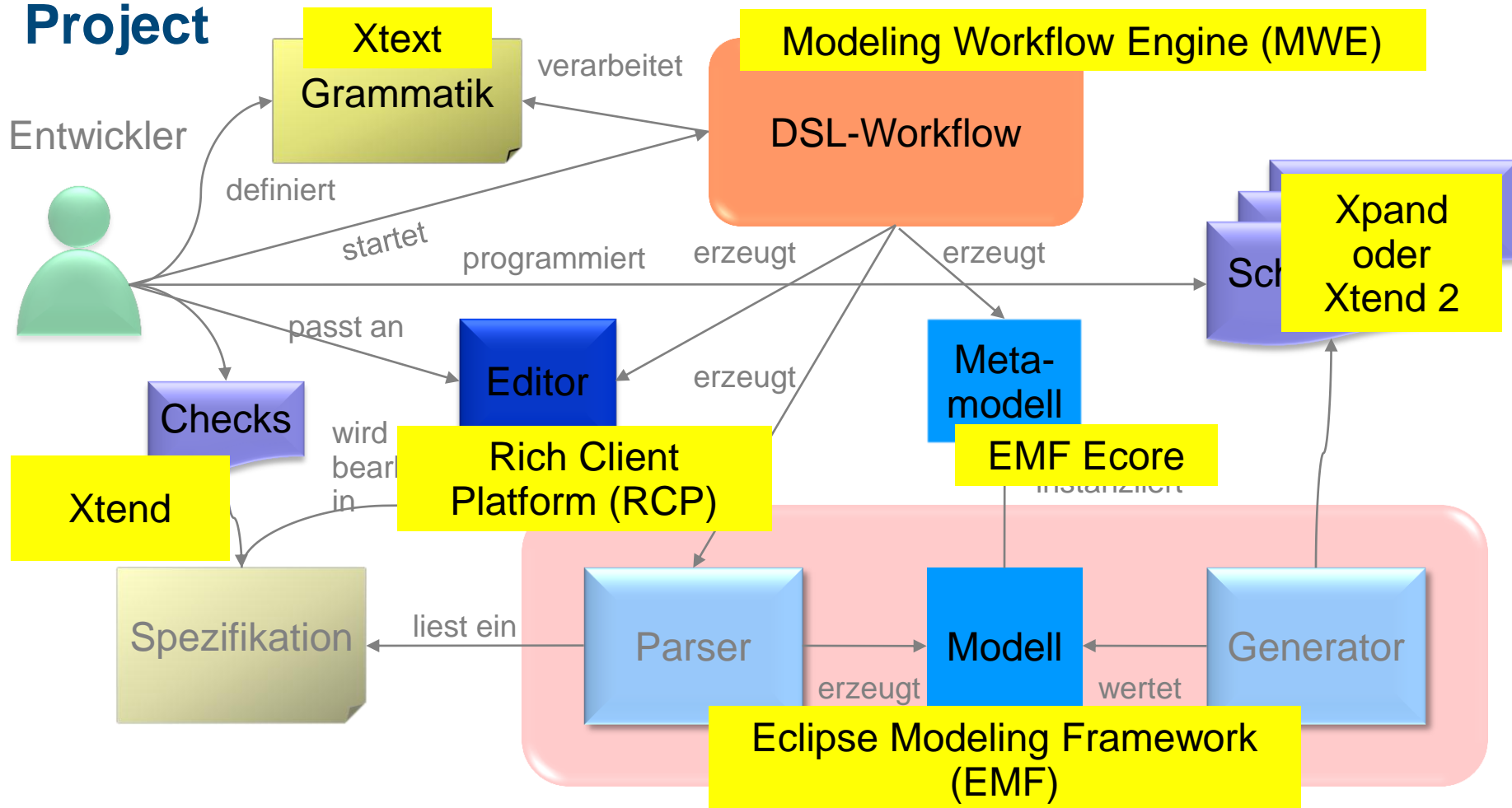
```

## Die DSL zum Leben erwecken





# Umsetzung mit Bausteinen aus dem Eclipse Modeling Project



## Schablonen für den Generator

```
«DEFINE createTables FOR Studienaufbau-»
/* Erzeuge Tabellen. */
«FOREACH this.eAllContents.typeSelect(ModulReferenz).sortBy(e|e.getTable()) AS modulRef-»
CREATE TABLE «modulRef.getTable()» (
  ID INT NOT NULL PRIMARY KEY,
  «FOREACH modulRef.items AS itemRef-»
  «itemRef.name» «itemRef.itemDef.spec.formatTyp»,
  «ENDFOREACH-»
);
«ENDFOREACH-»
«ENDDDEFINE»
```

```
«DEFINE items FOR ModulReferenz»
<table>
<tbody>
  «FOREACH this.items AS itemRef-»
  <tr>
  <th><label for="item«this.name.toFirstUpper()_«itemRef.itemDef.name»">
    «itemRef.itemDef.getLabel():«EXPAND mandatoryMarker FOR itemRef-»
  </label></th>
  <td>
  «IF itemRef.itemDef.teilItems.isEmpty-»
  «EXPAND simpleItem(this.modulDef, itemRef.name, null) FOR itemRef.itemDef.spec-»
  «ELSE-»
  «FOREACH itemRef.itemDef.teilItems AS teilItem»
  «EXPAND simpleItem(this.modulDef, itemRef.name, teilItem.bezeichnung) FOR teilItem.spec»
  «ENDFOREACH»
  «ENDIF-»
  </td>
  </tr>
  «ENDFOREACH-»
</tbody>
</table>
«ENDDDEFINE»
```

## Zeitlicher Arbeitsaufwand

- Definieren der Grammatik für die DSL: Stunden bis Tage
- DSL-Workflow zum Erzeugen des Metamodells, Editors und Parsers: Sekunden
- Programmieren von Checks und Schablonen: Tage
- Anpassen des Editors: Tage
- Spezifizieren einer Studie: Tage oder Wochen
- Erzeugen von Ausgabedokumenten zur Weiterverarbeitung z. B. Import in SMS: Sekunden
  
- Einmaliger Entwicklungsaufwand für DSL, Editor und Schablonen (ggf. Anpassungen bei neuen Anforderungen)
- Fortwährende Nutzung von Editor und Generator für beliebig viele Studien



## DSL on demand

- DSL sind zwar domänenspezifisch aber nicht auf die Domäne der Spezifikation klinischer Studien festgelegt
- Für welche Domäne und welchen Anwendungsfall ein DSL gebaut wird, bestimmt man selbst
- Mehrere Anwendungsfälle in mehreren Domänen? Mehrere DSL!
  - Verknüpfung von verschiedenen DSL möglich
  - Referenzen zwischen den Inhalten der Dokumente aus verschiedenen DSL
- DSL für ...
  - Auswertung von Studiendaten
  - Layout von CRF
  - Konfiguration des SMS
  - ...

**Vielen Dank  
für Ihre Aufmerksamkeit!**

**Roland Mücke, itemis AG  
23.09.2011**