



# Qualitätssicherung in der E-Book-Prozesskette

**XUGS@AKEP13**

BERLIN, 19. JUNI 2013

TOBIAS OTT, TOBIAS FISCHER

**pagina**   
Publikations  
technologien

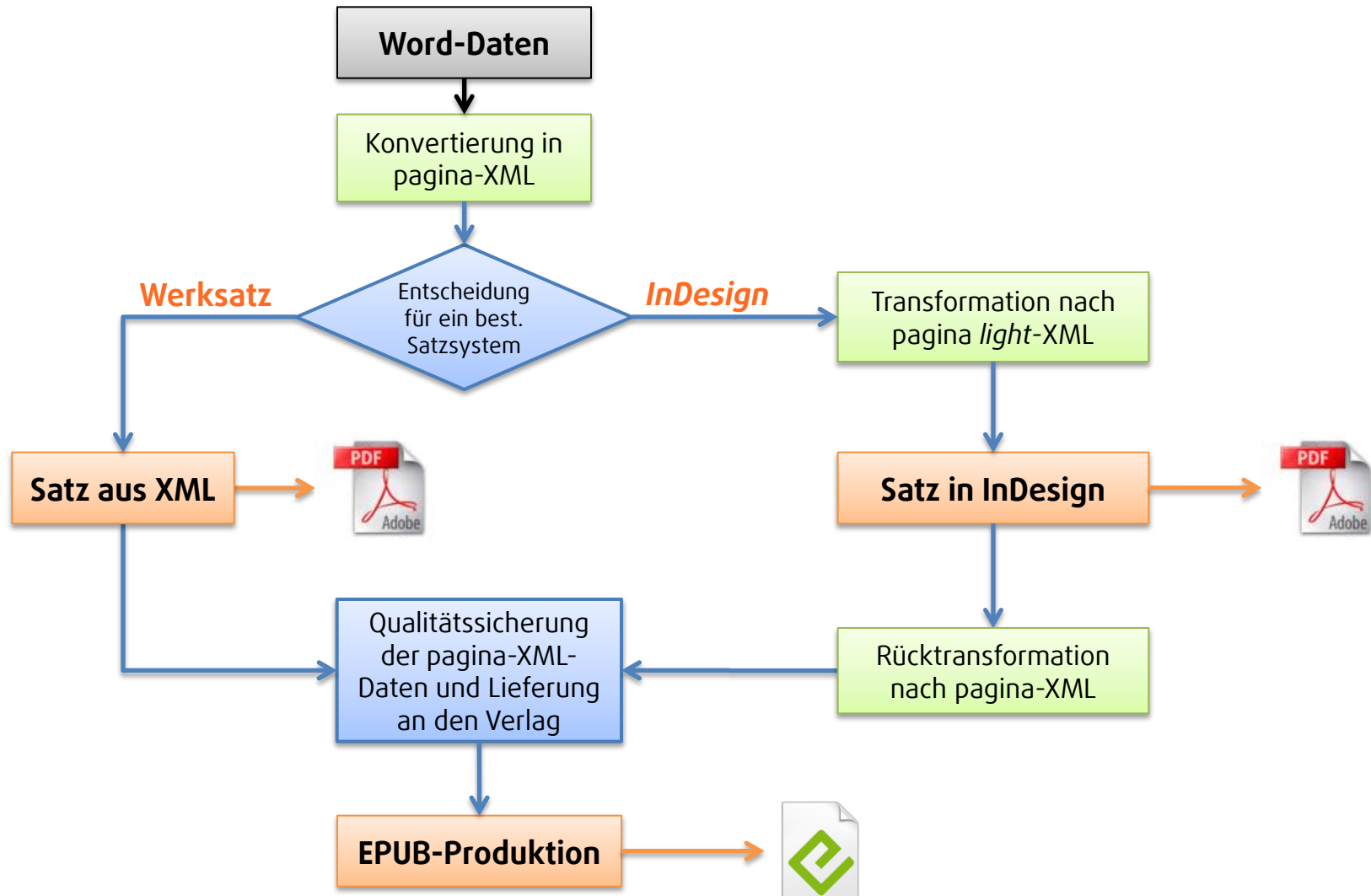
# Ausgangssituation

- XML-basierter Produktions-Workflow (*XML-first*)
- Automatisierte EPUB-Produktion



Wichtiger als die Qualitätssicherung des E-Books  
ist die Qualitätssicherung der XML-Quelldaten!

# Der pagina XML-Workflow



# Schritt 1: Datenkonvertierung

- **Ziel: valide XML-Daten**
- Arbeitsschritte abhängig von Datenquelle
  - **Konvertierung aus Word**
    - Normalisierung mit standardisierter Formatvorlage
    - Umwandlung Word-2003-XML in XML nach pagina-DTD (per Skript)

BL\_978-3-10-007626-7\_norm.docx - Microsoft Word

Start Einfügen Seitenlayout Verweise Sendungen Überprüfen Ansicht Entwicklertools pagina Word-Tools

GvH-Styles übertragen/aktualisieren  
Stilvorlagen schützen (StyleLock)

Stilformatierung Direktformatierung Tabelle(n) Mikrotypographie Rechtschreibung/Trennungen GvH-Automakro Unicode-Info Springe zu... Weiteres Dokumentation anzeigen Tastenkürzel Hilfe

Formatvorlagen Verwaltung von Stilvorlagen Formatierungstools Text-Korrekturwerkzeuge Sonstige Werkzeuge

01\_grundtext  
01\_grundtext  
03\_u3  
01\_grundtext  
01\_grundtext  
04\_einschub\_9  
04\_einschub\_9  
04\_einschub\_9  
04\_einschub\_9  
04\_einschub\_9  
04\_einschub\_9  
04\_einschub\_9  
01\_grundtext  
01\_grundtext

eigentliche Aussage ihrer Strophen, auf ein Minimum und widmeten ihr Hauptaugenmerk der Form, und die Feststellung, dass die Skaldendichtung in erster Linie Formkunst und weniger Inhalt Kunst ist, lässt sich nur schwer bestreiten. Naturphänomene, vor allem Wettererscheinungen, sind zwar fester Bestandteil der Bildsprache der Kenningar, aber Natur als solche und das Erlebnis von Natur wie auch Stimmungen und Gefühle sind selten Gegenstand skaldischer Dichtung. Dennoch gibt es bewegende Emotionen, z.B. in Egill Skalla-Grímssons langem Klagegedicht „Der Söhne Verlust“ - dies allerdings nicht im *dróttkvætt*, sondern im einfacheren *kviðuhátt* - oder in Kormáks Liebesstrophem.

## Eine Strophe des Skalden \$\$\$\$Pormóð Kolbrúnarskáld

Das folgende Beispiel veranschaulicht die formale Komplexität einer Dróttkvættstrophe von mittlerem Schwierigkeitsgrad. Es handelt sich um die letzte Strophe des von Þormóð Kolbrúnarskáld über seinen getöteten Freund und Schwurbruder Þorgeir Hávarson verfassten Nachrufgedichts (*erfidrápa*), das in Bruchstücken in der „Saga von den Schwurbrüdern“ überliefert ist. Das Gedicht muss um 1024 entstanden sein und zählt Þorgeirs Kämpfe und Siege auf und endet mit seinem Tod.

Olli fjör, áör fell,  
flugtrauðr hjarar dauða  
- sá vas *rækjandi* enn ríki  
*reggs* - Þrétían seggia;  
þar læt *hjaldrs* fyr *hólðum*  
*hins*, es þrek gat vinna  
- *mál téa mín* at deilask  
*mjúk* - vígatal *lúkask*

Die Strophe ist durch eine deutliche Zäsur in zwei syntaktisch selbständige Halbstrophen geteilt, die aus vier Kurzversen zu je sechs Silben bestehen. Der Stabreim (*fett*) verbindet jeweils zwei Kurzverse zu einer Langzeile: zwei Stäbe im Anvers, der dritte auf der ersten Hebung des Abverses. Der Innenreim (*unterstrichen*) bleibt auf den Kurzvers beschränkt, der erste Reim liegt nicht fest, der zweite liegt auf der vorletzten Silbe. In den Zeilen 1,3,5,7 steht Halbreim (*skothending*, es reimen Konsonanten bei wechselnden Vokalen), in den geraden Zeilen steht Vollreim (*aðalhending*, es reimen Konsonanten plus Vokale). Die Syntax folgt nicht der gewöhnlichen Ordnung, den Erfordernissen von Stabreim und Innenreim entsprechend sind die Satzglieder relativ frei platzierbar. In die Strophe eingeschobene, syntaktisch selbständige Kurzsätze, die oft eine Bemerkung in der Ich-Form oder

Formatvorlagen

03\_u\_zwischen\_1  
03\_u\_zwischen\_2  
03\_u\_zwischen\_3  
03\_u\_zwischen\_4  
03\_u\_zwischen\_5  
03\_u1  
03\_u2  
03\_u3  
03\_u4  
03\_u5  
03\_u6  
03\_u7  
03\_untertitel  
03\_u-vor  
03\_u-zaehler  
04\_absatz\_joker  
04\_einschub\_1  
04\_einschub\_2  
04\_einschub\_3  
04\_einschub\_4  
04\_einschub\_5  
04\_einschub\_6

☒ Vorschau anzeigen  
☐ Verknüpfte Formatvorlagen deaktivieren  
Optionen...

Seite: 56 von 159 Wörter: 85.194 Deutsch [Deutschland]

# Schritt 1: Datenkonvertierung

- **Ziel: valide XML-Daten**
- Arbeitsschritte abhängig von Datenquelle
  - **Konvertierung aus Word**
    - Normalisierung mit standardisierter Formatvorlage
    - Umwandlung Word-2003-XML in XML nach pagina-DTD (per Skript)
  - **Daten aus Satzsystemen**
    - InDesign: Rücktransformation aus verflachter InDesign-DTD (eigenes InDesign-PlugIn für besseres Roundtripping)
    - Werksatz: von Pls und SteuerCodes bereinigtes Ausgabe-XML
  - **Backlist-Digitalisierung**
    - „normale“ Digitalisierung: Erfassung oder Konvertierung in die pagina-DTD
    - Zeilengleiche Redigitalisierung: Erfassung nach eigener Backlist-DTD, danach Transformation

# Schritt 2: Datenveredelung

- **Ziel: Einheitlich strukturiertes und mikrotypographisch angereichertes XML**
- Arbeitsschritte auf ...
  - **Strukturebene (XML-Struktur)**
    - Bei Daten aus InDesign:  
*Ergänzung nicht unterstützter Strukturen wie mehrstufige Register, Dramensatz, komplexe Listenstrukturen*
    - Bei Daten aus der Redigitalisierung:  
*Eliminierung print-spezifischer Informationen und redundanter Auszeichnungen*
    - Daten aus Word-Konvertierung:  
*Bereinigung von Tabellen, verschachtelte Strukturen, Vervollständigung von Container-Elementen (ähnlich wie aus InDesign)*
  - **Zeichenebene**
    - Regelbasiertes Einbringen von mikrotypographischen Auszeichnungen per Skript (sofern nicht bereits aus dem Satz vorhanden; nach Verlagskonventionen)

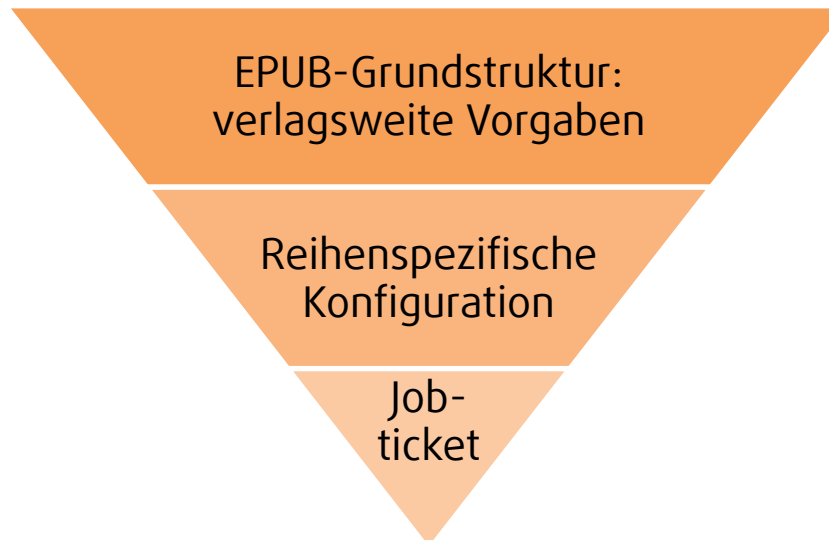
## Schritt 3: Qualitätssicherung

- **Ziel: XML-Daten, die den Verlagskonventionen entsprechen**
- Werkzeuge:
  - **Schematron:** *ergänzende Struktur- und Konventionsprüfung*
    - Einhaltung der Namenskonvention für Abbildungsaufrufe
    - Konsistenzprüfung für Tabellen
    - Prüfung von Querverweisen, etc.
  - **Regular Expressions:** *Prüfung der Mikrotypographie*
    - Anzeige falscher oder fehlender Festabstände
    - Manuelle Prüfung von semantisch zu unterscheidenden Fällen
  - **pagina DokuChecker:** *Protokoll für schnelle Datenübersicht*
    - Prüfprotokoll, u. A. zu Dokumentations- und Archivierungszwecken
    - Webanwendung, steht auch Kunden und deren Dienstleistern zur Verfügung



# Schritt 4: EPUB-Produktion

- **Ziel: EPUB nach einheitlichen Verlagskonventionen**
- Qualitätssicherungsmaßnahmen:
  - **Automatisierte EPUB-Produktion**
    - Stellt einheitliche Grundstruktur der EPUBs sicher



# Schritt 4: EPUB-Produktion

- **Ziel: EPUB nach einheitlichen Verlagskonventionen**
- Qualitätssicherungsmaßnahmen:
  - **Automatisierte EPUB-Produktion**
    - Stellt einheitliche Grundstruktur der EPUBs sicher
  - **Reihenspezifische Konfigurationsdateien**
    - 140 Einstellungsmöglichkeiten zur reihenspezifischen Anpassung der Grundstruktur

Kapitälchen in Versalien umwandeln:

Ja

Sollen Kapitälchen komplett in Versalien umgewandelt werden?  
*Hintergrund: Umwandlung per CSS funktioniert nicht auf allen Lesegeräten!*

Versalien "hart" großschalten:

Ja

Sollen Versalien im Quellcode "hart" großgeschaltet werden?  
*Hintergrund: Umwandlung per CSS funktioniert nicht auf allen Lesegeräten!*

<sprecher> "hart" großschalten:

Ja

Sollen Inhalte des Elements <sprecher> im Quellcode "hart" großgeschaltet werden? (Inhalte vom Kindelement <regie\_inline> bleiben unberücksichtigt!)  
*Hintergrund: Umwandlung per CSS funktioniert nicht auf allen Lesegeräten!*

Tabulator-Fallback:

Tabulatoren (zur Zeit nur *tabulator\_rechts*) werden durch Weißraum ersetzt. Definieren Sie hier diesen Weißraum (z.B. drei Gevierte, 6 nicht umbrechende Leerzeichen, o.Ä.) **als Unicode-Glyphe** (KEINE Entities!).  
**Achtung:** Die Regeln der Festabstands-Ersetzung (*Char-Mapping*) greifen auch hier! Definieren Sie also keine Festabstände als Tabulator-Ersetzung, welche später selbst ersetzt werden!

### Einstellungen für die Darstellung der Original-Seitenzahlen aus dem gedruckten Buch

Anzeige der Print-Seitenzahlen?

Nein

Sollen die Original-Seitenzahlen aus dem gedruckten Buch im EPUB angezeigt werden?

### Options-Sets

[Start]

Metadaten

Styling / CSS

Seitenzahlen

Abbildungen

Thumbnails

Seite: Cover

Seite: Haupttitel

Seite: Navigationsseite

Seite: Inhaltsübersicht

Seite: Über den Autor

Seite: Über das Buch

Seite: Impressum

Seite: Hinweise des Verlags

Kapitel allgemein

Fuß- und Endnoten allg.

Seite: Fußnoten

Seite: Endnoten

Seite: Buchfrage

Seite: Abbildungen

Seite: PopUp-Fallback

TOC: Allgemeines

TOC: Kapitel-Fallbacks

TOC: Überschrifts-Fallbacks

TOC: Autom. Sequenzziffer

TOC: Kapitelnummern

TOC: Vor- und Nachspann

Jobticket

EPUB-Leseprobe

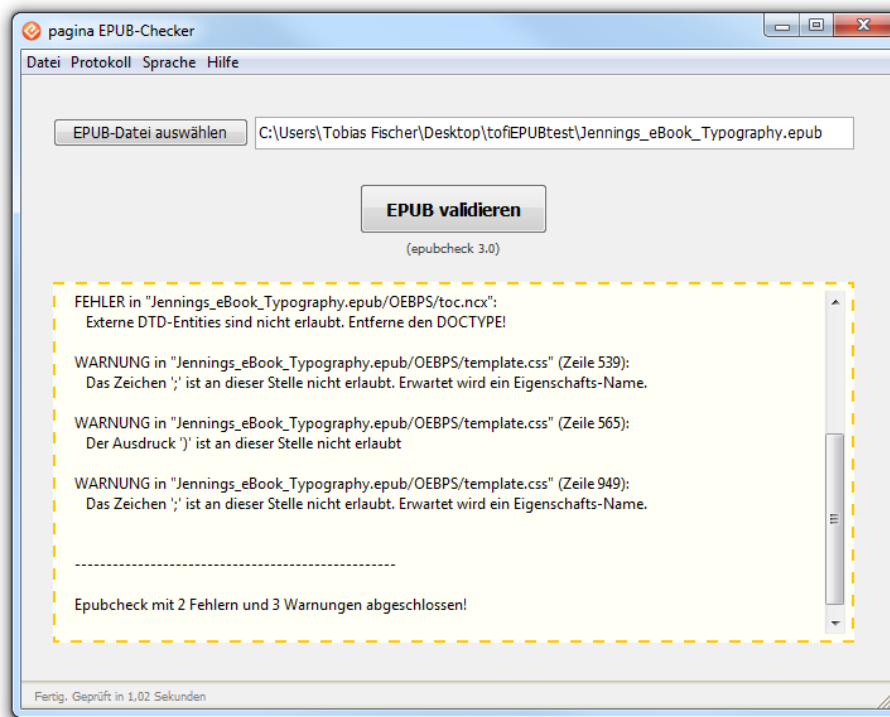
EPUB-Datei

# Schritt 4: EPUB-Produktion

- **Ziel: EPUB nach einheitlichen Verlagskonventionen**
- Qualitätssicherungsmaßnahmen:
  - **Automatisierte EPUB-Produktion**
    - Stellt einheitliche Grundstruktur der EPUBs sicher
  - **Reihenspezifische Konfigurationsdateien**
    - 140 Einstellungsmöglichkeiten zur reihenspezifischen Anpassung der Grundstruktur
  - **pagina GlyphenChecker & FontSubsetter**
    - Prüft eingebettete EPUB-Schriften auf nicht vorhandene Glyphen
    - Ermöglicht den Einsatz umfassender Unicode-Fonts bei zugleich geringer Dateigröße
  - **Automatische Konventionsprüfungen**
    - Prüfung verlagsspezifischer oder allgemeiner Konventionen
  - **Automatische EPUB-Validierung mit "epubcheck"**

# pagina EPUB-Checker

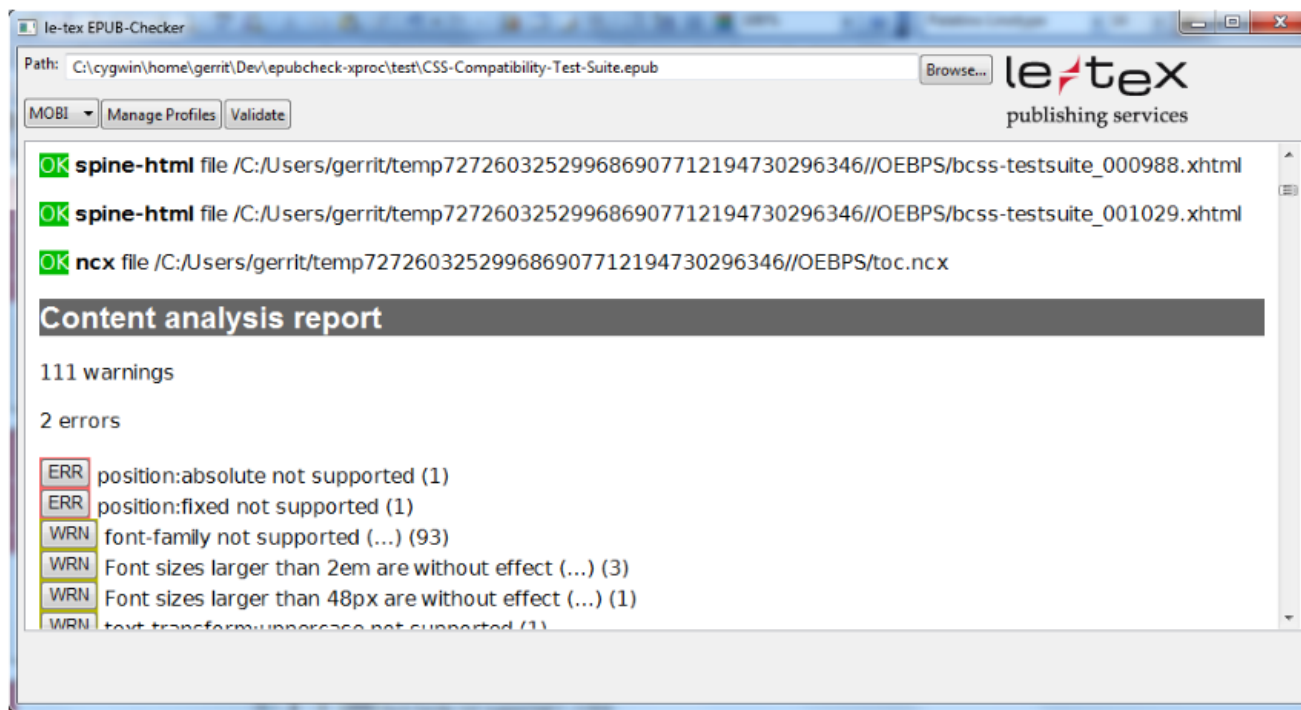
- Grafische Oberfläche und deutsche Fehlermeldungen für den offiziellen „epubcheck“



Kostenloser Download: <http://www.pagina-online.de/software/epub-checker/>

# le-tex epubcheck-xproc

- Möglichkeit gerätespezifische Prüfprofile zu hinterlegen, z.B. für Mobipocket



PDF-Präsentation: [http://www.le-tex.de/img/presentation\\_2014-10\\_lmsieke\\_epubcheck-xproc.pdf](http://www.le-tex.de/img/presentation_2014-10_lmsieke_epubcheck-xproc.pdf)

Download/Quellcode: <https://github.com/gimsieke/epubcheck-xproc>

# </vortrag>

## <berlin-digital-night>

**TOBIAS OTT**

tobias.ott@pagina-tuebingen.de



**TOBIAS FISCHER**

tobias.fischer@pagina-tuebingen.de