

# DalfMod: een markupformaat voor de beschrijving van brieven

Raf Guns  
*raf.guns@telenet.be*

## Abstract

DALF is een op TEI gebaseerd formaat voor de beschrijving van literair-historische brieven. Dit artikel stelt DalfMod voor, een set van wijzigingen aan DALF. In tegenstelling tot DALF is DalfMod gebaseerd op P5, de laatste versie van TEI. Dit maakt het mogelijk XML-schema's, datatypes en namespaces te gebruiken en het formaat modulair op te bouwen. Daarnaast bevat DalfMod ook enkele andere voorstellen tot verbetering, waaronder een vereenvoudigde indeling en transcriptie van extern materiaal.

## 1. Inleiding

Dit artikel stelt DalfMod voor, een formaat dat is ontstaan in het kader van mijn verhandeling aan de UA. Dit betekent dat de aanpassingen grotendeels het werk van een individu zijn, veeleer dan een standaard. De uiteindelijke doelstelling van DalfMod is te komen tot een markupformaat voor de beschrijving van (literaire en historische) brieven. Dit formaat omvat zowel metadata als de transcriptie van de tekst. DalfMod is een aanpassing van DALF in twee opzichten: enerzijds is het een P5-versie van DALF, anderzijds bevat DalfMod ook algemene verbeteringen aan DALF. Waarom zijn deze aanpassingen van belang?

DalfMod is (zeker met betrekking tot het P5-gedeelte) te beschouwen als voorbereidend werk voor de toekomst van DALF. In 2006 zal immers het werk aan een nieuwe, aan P5 conformerende, versie van DALF worden aangevat. Momenteel wordt DalfMod nader bekeken door de auteurs van DALF om te onderzoeken welke wijzigingen in de nieuwe versie van DALF kunnen worden opgenomen.

Op langere termijn kan DalfMod een eerste stap betekenen voor de integratie van DALF in de officiële richtlijnen van TEI. TEI moedigt integratie van 'tagsets' voor nieuwe documenttypes aan, — ook een 'brievenformaat' zou opgenomen kunnen worden. DalfMod leunt dichter aan bij TEI, omdat (*a*) DalfMod algemener van aard is (DALF is gericht op specifieke praktische toepassingen), en (*b*) DalfMod gebaseerd is op P5, de meest recente versie van TEI. Bij zo'n integratie in TEI wordt de nieuwe tagset vanzelfsprekend grondig geëvalueerd en herbekeken; DalfMod bevat zowel een kritische evaluatie van DALF als concrete voorstellen tot verbetering.

Tot slot is DalfMod ook een test van TEI P5. Met name de manier waarop TEI kan worden uitgebreid en aangepast, is compleet anders in P5 dan in de vorige versies van TEI. DalfMod is een van de eerste projecten die deze nieuwe mogelijkheden gebruiken en testen.

## 2. Brieven

Wat is een brief? De traditionele editiewetenschap hanteert volgende criteria [4]:

- Tekst die contact legt of bestendigt tussen 2 of meer personen

- Bedoeld ter verzending/overhandiging
- Niet bedoeld voor openbaarmaking. Gezien het beperkte ‘publiek’ wordt de communicatieve context (participanten, tijd en plaats) vaak bekend verondersteld en niet meer expliciet vermeld. Deze context expliciteren is een van de taken van een markupformaat voor brieven.

### 3. TEI en TEI P5

De richtlijnen van het Text Encoding Initiative (TEI) vormen een set van markup-specificaties (XML) voor literaire, historische en taalkundige teksten. Door de modulaire, uitbreidbare opbouw en uitgebreide, in de praktijk bewezen basis is TEI een goed vertrekpunt voor een markupformaat voor allerlei tekstsoorten. Ieder TEI-document kent een rigide indeling in metadata enerzijds en tekst anderzijds:

```
<TEI>
  <teiHeader><!-- alle metadata hier --></teiHeader>
  <text><!-- de eigenlijke tekst hier --></text>
</TEI>
```

TEI P5 [7] is de laatste versie van TEI; deze is momenteel nog in ontwikkeling. Dit is de eerste versie die volledig **XML-gebaseerd** is, wat toelaat de mogelijkheden van XML ten volle uit te buiten (zoals namespaces voor integratie [5] met andere XML-gebaseerde formaten zoals SVG, MathML of DocBook). Waar TEI P4 volledig gebaseerd was op DTDs, kunnen in P5 verschillende **schemata** gebruikt worden: DTD, W3C XML Schema (WXS) en RELAX NG. Hiertoe maakt P5 gebruik van een eigen meta-formaat, ODD. ODD kan aan de hand van Roma<sup>1</sup> worden omgezet in een DTD of een XML-schema. Ter illustratie geven we een eenvoudige ODD-specificatie voor het element <closing>.

```
<elementSpec ident="closing" module="letdescription" mode="add">
  <content xmlns:rng="http://relaxng.org/ns/structure/1.0">
    <rng:ref name="macro.phraseSeq" />
  </content>
  <desc>Contains a salutation at the end of a letter.</desc>
</elementSpec>
```

### 4. Bestaande formaten

We maken een onderscheid tussen *metadata* en *tekst*. Tekst kan op zijn beurt worden onderverdeeld in microniveau (niveau van woord, zin of paragraaf) en macroniveau (niveau van ‘structurele elementen’, elementen die de structuur van de tekst aangeven).

TEI is bijzonder goed geschikt voor de transcriptie op microniveau; zo bevat TEI uitgebreide mogelijkheden voor de aanduiding van correcties, toevoegingen, onleesbare delen en weglatingen. TEI is echter zeer beperkt op het vlak van briefspecifieke metadata en het macroniveau van brieven.

---

<sup>1</sup> <http://tei.oucs.ox.ac.uk/Roma/>

Model Editions Partnership (**MEP**) [3] is een formaat dat vrij specifiek voor brieven bedoeld is. MEP kent echter een zeer specifiek toepassingsveld: het is uitsluitend bedoeld voor ambtelijke brieven en de transcriptie van reeds bestaande ‘papieren’ edities.

Manuscript Access through Standards For Electronic Records (**MASTER**) [2] is een uitbreiding van TEI, bedoeld voor de beschrijving van (middeleeuwse) manuscripten. MASTER voegt een extra element `<msDescription>` toe, waarbinnen alle manuscriptspecifieke metadata worden ondergebracht. Dit element is een inspiratie geweest voor de `msdescription`-module van TEI P5 en voor de metadata van DALF.

Digital Archive of Letters in Flanders (**DALF**) [8] is een uitgebreid en aan de praktijk getoetst formaat, dat expliciet bedoeld is voor ongepubliceerde, literair-historische brieven. DALF bestaat uit XML met een DTD en is gebaseerd op TEI P4 [6]. De briefspecifieke metadata van DALF zijn gebaseerd op MASTER. Deze worden opgenomen in het extra element `<letDesc>` in de header. Op het niveau van de tekst introduceert DALF verschillende nieuwe structurele elementen.

## 5. DalfMod

DalfMod is een set van voorgestelde wijzigingen aan DALF. Deze wijzigingen zijn in twee stappen aangebracht. De eerste stap was een ruwe P5-versie van DALF, die doelbewust zo dicht mogelijk bij het oorspronkelijke DALF-formaat bleef. De tweede stap bestond erin (a) DalfMod beter te integreren met P5 en (b) enkele andere problemen uit het oorspronkelijke DALF op te lossen.

### 5.1. Stap 1: DALF en TEI P5

Stap 1 bestond uit een ‘letterlijke’ omzetting van de DALF-DTD naar TEI P5 en had in de eerste plaats tot doel de mogelijkheden en potentiële problemen te onderzoeken van een omzetting naar P5. De belangrijkste vaststellingen waren:

- Voor deze eerste stap werd gebruik gemaakt van de webinterface van Roma. De mogelijkheden van deze interface bleken relatief beperkt; door manuele aanpassing van het ODD-bestand konden wel de gewenste resultaten bereikt worden.
- DALF gebruikt de module `TEI.drama`, die elementen bevat i.v.m. toneel, regieaanwijzingen, filmopnames enz.; het is onduidelijk waarom deze module is opgenomen.
- DALF geeft de belangrijkste elementen van de ontstaanscontext aan in `<letHeading>`. P5 heeft geen echt alternatief hiervoor. Dit element kan grotendeels uit DALF worden overgenomen, al is de elementnaam ‘`letHeading`’ enigszins onduidelijk over de inhoud.
- Zowel DALF als de `msdescription`-module van P5 bevatten elementen om extern materiaal — zoals illustraties of toegevoegde objecten — te beschrijven. Aan beide benaderingen zijn echter nadelen verbonden. De benadering van `msdescription` is duidelijk bedoeld voor oude manuscripten en valt niet goed te extrapoleren naar moderner brievenmateriaal. De indeling van DALF is tamelijk ingewikkeld en op sommige vlakken dubbelzinnig.
- Tot slot zijn er enkele kleine praktische problemen. Deze hebben tot gevolg dat de ruwe P5-versie niet direct compatibel is met. In de eerste plaats worden de entity’s uit P4 niet ondersteund, aangezien XML-schema’s geen entity’s kennen. Bij gebruik van een DTD kunnen deze wel terug toegevoegd

worden. Daarnaast zijn er ook verschillen met betrekking tot de namen van een standaardelement en twee attributen.

	TEI P4 en DALF	TEI P5
root-element	<TEI.2>	<TEI>
attributen voor unieke identificatie en taalaanduiding	id en lang	xml:id en xml:lang

## 5.2. Stap 2: DalfMod

DalfMod bouwt voort op de resultaten van stap 1; er is getracht DalfMod verder op te bouwen volgens de praktijk van P5 en de vastgestelde problemen te verbeteren. In tegenstelling tot de eerste stap was compatibiliteit met DALF hier geen doelstelling meer.

Naar analogie van P5 streeft DalfMod naar een modulaire opbouw. Daarom is DalfMod opgedeeld in twee modules, **primarymeta** en **letdescription**. De module **primarymeta** bevat metadata die typisch zijn voor alle primaire bronnen (brieven, manuscripten, dagboeken enz.). Deze module kan in principe gedeeld worden door alle formaten die primaire bronnen beschrijven. Momenteel is **primarymeta** opgesteld op basis van gelijkaardige elementen in DalfMod en de **msdescription**-module van TEI P5. Als we ook rekening willen houden met andere documenttypes, is deze module zeker nog vatbaar voor verdere uitbreiding en verfijning. Externe documenttypes kunnen de **primarymeta**-module toevoegen aan het ODD-bestand: `<moduleRef url="primarymeta.rng"/>`. In **primarymeta** zijn onder andere de volgende elementen gedefinieerd: `<accMat>`, `<condition>`, `<history>` en `<repository>`.

De module **letdescription**, anderzijds, bevat zowel metadata als tekstuele elementen die typisch zijn voor brieven. Binnen de metadata vinden we bijvoorbeeld `<letIdentifier>`, een formele identificatie en localisatie van de brief. Op het vlak van de eigenlijke tekst geven we het voorbeeld van `<ps>` voor het postscriptum of `<closing>` voor de afsluitende groet aan het einde van een brief.

Een interessante nieuwigheid in P5 is de introductie van **datatypes** [1]. Zowel RELAX NG als WXS kunnen datatypes gebruiken en valideren. Datatypes maken het mogelijk de inhoud van elementen en attributen zeer precies te bepalen (b.v. “element x bevat een datum in ISO-formaat”); afwijkingen worden bij validatie als fout gemarkeerd.

Zo geeft het element `<envOcc>` aan of de omslag al dan niet bewaard is gebleven. In DalfMod kunnen hiervoor eenvoudig booleaanse waarden (‘true’ en ‘false’) gebruikt worden: `<envOcc>true</envOcc>`. In andere gevallen heeft het gebruik van datatypes voornamelijk een documenterende functie. Bijvoorbeeld: als een attribuut het type `datatype.Key` heeft, weten we dat het de bedoeling is een “value from a set of possible keys” te verstrekken en dat “[i]t is expected that the set of possible values will be defined in the document header”.

Zoals al eerder (in 2) is aangegeven, is het van belang dat de communicatieve context van een brief expliciet wordt opgenomen. DALF gebruikt hiervoor `<letHeading>`. In DalfMod is `<letHeading>` vervangen door `<origContext>`. `<origContext>` is overgenomen uit de **msdescription**-module van TEI; op dezelfde manier zijn de DALF-elementen `<letDate>` en `<letPlace>` vervangen door de **msdescription**-elementen `<origDate>` en `<origPlace>`. Deze aanpassingen hebben tot gevolg dat de

ontstaanscontext van primaire bronnen op een uniforme manier behandeld kan worden. De elementen <origContext>, <origPlace> en <origdate> zijn dan ook lid van primarymeta.

Het element <additional> stamt oorspronkelijk uit MASTER. De naam zegt veel over de inhoud: <additional> bevat alles wat niet elders kon worden ingedeeld, nl. “additional information concerning the present custody and availability, other representations of the letter, accompanying materials, and further bibliographical information” [8].

Om de inhoud van <additional> consistentier te maken, stel ik voor <accMat> (beschrijving van toegevoegd extern materiaal) te verplaatsen. Na verplaatsing van <accMat> bevat <additional> informatie over de huidige context van een document: de huidige bewaarplaats en beschikbaarheid, andere representaties (b.v. digitaal, microfilm) en bibliografische referenties over het document. Aldus maken we de weg vrij voor een hernoeming van <additional> naar (bijvoorbeeld) <currentContext>.

De bestaande indelingen van **extern materiaal** bleken ofwel ongeschikt ofwel te ingewikkeld. Daarom heeft DalMod een eenvoudige eigen indeling voor extern materiaal in ruime zin. Waar mogelijk zijn standaard TEI-elementen gebruikt.

We maken een onderscheid tussen toevoegingen die apart of los van de tekst staan en toevoegingen die een geheel vormen met de tekst. Concreet komt dit neer op de vraag: bevindt de toevoeging zich op dezelfde drager als de tekst staat of niet? In het algemeen gaan we ervan uit dat tekeningen, illustraties en tekst op dezelfde drager een geheel vormen, terwijl voorwerpen of aparte documenten niet tot het geheel van de tekst behoren.

	Apart	Bij de tekst	
		Illustratie	Voor-/Nabedrukt
Metadata (header)	<accMat> en <accMatItem>	—	<support>
Tekst	<ab> en <anchor>, gelinkt aan <accMatItem>	<figure> en <graphic>	<print>

Het voorbeeld uit DALF (p. 30–31) kan dan als volgt worden geëncodeerd. Dit voorbeeld illustreert het gebruik van <accMat>, <accMatItem>, <anchor>, <figure> en <graphic>.

```
<TEI>
  <teiHeader>
    [...]
    <accMat>
      <accMatItem xml:id="obj1-desc" corresp="obj1">dried
        petal of a poppy</accMatItem>
    </accMat>
    [...]
  </teiHeader>
  <text>
    [...]
```

```
<p>Een grote boom stond daar
  <figure rend="inline">
    <graphic url="fig1.jpg"/>
    <figDesc>small black-white drawing of a tree</figDesc>
  </figure>,
  met aan zijn voet een mooie papaver
  <anchor xml:id="obj1" corresp="obj1-desc"/>.</p>
[...]
```

### 6. Concluesies

Zoals de naam suggereert, is DalfMod geen volledig nieuw formaat, maar een set van (voorstellen tot) wijzigingen aan DALF. We bespreken hier een drietal verschillen.

DalfMod is in de eerste plaats lichter en eenvoudiger dan DALF: DalfMod maakt geen gebruik van TEI.drama, en laat elementen als <calc> achterwege. Vanzelfsprekend houdt dit ook een ‘kost’ in: in sommige gevallen zal DALF meer kunnen uitdrukken dan DalfMod.

Vervolgens is DalfMod — dankzij de P5-basis — sterker modulair opgebouwd dan DALF. Enerzijds bevat DalfMod zelf twee verschillende modules, letdescription en primarymeta. Anderzijds is het dankzij namespaces ook gemakkelijker DalfMod te combineren met andere XML-formaten.

Ten slotte heeft DalfMod nog meer dan DALF de bedoeling zoveel mogelijk voort te bouwen op de bestaande basis van TEI. De encodingswijze voor extern materiaal illustreert dit. Met uitzondering van <accMatItem> (en <accMat> als we geen rekening houden met de msdescription-module) komen alle elementen gewoon uit TEI zelf.

In 2006 wordt DALF herzien. Daarbij zullen wellicht sommige ideeën uit DalfMod worden overgenomen en andere ideeën worden verworpen. Twee zeer praktische uitdagingen voor die besprekingen zijn: (a) een verfijning en verbetering van de primarymeta-module, en (b) een scherper stellen van de inhoud en naam van <additional>.

### 7. Referenties

Bronbestanden en uitgebreidere documentatie zijn online beschikbaar op <http://randomseed.be/markup-brieven/>.

1. BIRON, P.V. & A. MALHOTRA. 2004. *XML Schema Part 2: Datatypes Second Edition*. URL: <http://www.w3.org/TR/xmlschema-2/>.
2. BURNARD, L. (red.). 2001. *Reference Manual for the MASTER Document Type Definition. Discussion Draft*. URL: <http://www.tei-c.org.uk/Master/Reference/oldindex.html>.

3. CHESNUTT, D. R., S. M. HOCKEY & C. M. SPERBERG-MCQUEEN. 1999. *Markup Guidelines for Documentary Editions*.  
URL: <http://adh.sc.edu/MepGuide.html>.
4. MATHIJSEN, M. 1997. *Naar de letter. Handboek editiewetenschap*.  
Den Haag: C. Huygens Instituut.
5. RAHTZ, S., N. WALSH & L. BURNARD. 2004. "A Unified Model for Text Markup: TEI, Docbook, and Beyond".  
URL: <http://www.tei-c.org.uk/Talks/xml europe2004/index.pdf>.
6. SPERBERG-MCQUEEN, C. M. & L. BURNARD (red.). 2002. *TEI P4: Guidelines for Text Encoding and Interchange: XML-Compatible Edition*.  
URL: <http://www.tei-c.org/P4X/>.
7. SPERBERG-MCQUEEN, C. M., L. BURNARD & S. BAUMAN (red.). 2005. *TEI P5: Guidelines for Text Encoding and Interchange*.  
URL: <http://www.tei-c.org.uk/P5/Guidelines/>.
8. VANHOUTTE, E. & R. VAN DEN BRANDEN (red.). 2003. *DALF Guidelines for the Description and Encoding of Modern Correspondence Material Version 1.0*.  
URL: <http://www.kantl.be/ctb/project/dalf/dalfdoc/>. Gent: CTB-KANTL.