

# Spectateurs (Moralische Wochenschriften)

Spectateurs (Moralische Wochenschriften), Klaus-Dieter Ertler, Alexandra Fuchs, Michaela Fischer, Elisabeth Hobisch (ed.), 2011. <a href="http://gams.uni-graz.at/context:mws">http://gams.uni-graz.at/context:mws</a> (Last Accessed: 08.08.2017). Reviewed by Greta Franzini (Georg-August-Universität Goettingen), gfranzini (at) etrap.eu.

#### **Abstract**

The Spectateurs (Moralische Wochenschriften) project is collecting eighteenth century periodical essays from the Enlightenment period from different European countries in an effort to build a comprehensive database of this literary genre. The unique selling proposition of the corpus is its multilingualism, in that users can explore the genre across borders and compare the national socio-cultural movements of the time. The project offers exciting interdisciplinary research opportunities but more should be done to facilitate the reuse of the collection, starting from better documentation and an efficient download system.

#### Introduzione

Il progetto *Spectateurs* (*Moralische Wochenschriften*), in italiano *Spettatori*, mette in rete una collezione digitale di settimanali moralistici europei in continua crescita risalenti al diciottesimo secolo. Il progetto, che può essere considerato un corpus di riferimento in materia, prende il nome dal quotidiano inglese 'The Spectator', apparso per la prima volta nel 1711 e imitato in tutt'Europa durante l'Illuminismo sia come quotidiano che come settimanale. Il carattere brillante ed educativo di questa tipologia di

giornale sollecitò l'interesse di lettori e lettrici di diversa estrazione e classe sociale, nonostante fosse principalmente rivolto a un pubblico borghese (Ertler, 2012).

Spectateurs (Moralische Wochenschriften) – d'ora in poi MWS – si propone di raccogliere ed esporre tutti questi settimanali per uno studio approfondito della comunicazione di massa e del panorama socio-culturale illuminista, nonché dell'influenza esercitata da questo genere letterario-giornalistico sulla narrativa moderna. Il progetto, unico nel suo intento, è diretto da Klaus-Dieter Ertler dell'Istituto di Lingue Romanze dell'Università di Graz in collaborazione con il Zentrum für Informationsmodellierung austriaco. L'iniziativa più vicina è Europeana Newspapers. Il testo introduttivo del progetto è datato febbraio 2011 ma la durata ufficiale del progetto non è specificata. La pagina Imprint, obbligatoria per siti pubblicati in paesi germanofoni e solitamente utilizzata per dichiarazioni legali e recapiti, non fornisce informazioni circa le risorse economiche del progetto e neppure un recapito. L'unico indirizzo email visibile è nascosto nelle schede dei metadati Dublin Core allegate a ciascun giornale (Fig. 1).

#### Obiettivi e contenuto

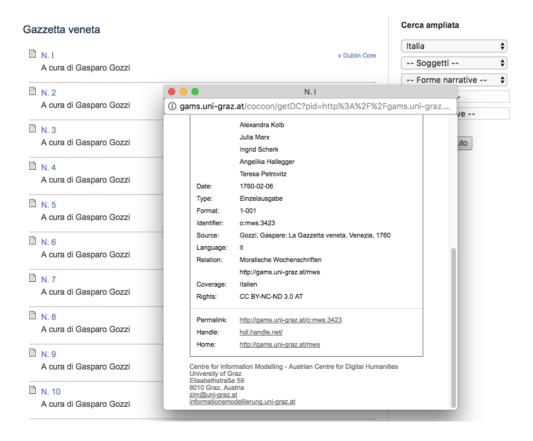


Fig. 1: L'unico contatto email del progetto è reperibile dalle schede metadati *Dublin Core* allegate ai giornali attualmente disponibili nel database.



Fig. 2: Schermata di MWS. In questo esempio, il quarto numero della *Gazzetta Veneta*, la citazione bibliografica si trova a capo pagina. La trascrizione del testo è disponibile in due formati, XML (TEI P5) e PDF, ed il menù di destra fornisce altre modalità di visualizzazione.

- L'utente può consultare il sito di MWS in diverse lingue europee (tedesco, inglese, francese, italiano e spagnolo). La pagina iniziale di ciascuna versione riassume il contenuto del progetto e ne spiega le motivazioni. Manca tuttavia una documentazione dettagliata che fornisca i particolari della codifica XML (TEI P5) e delle scelte editoriali e che chiarisca il motivo per cui i giornali non siano trattati in ugual modo: alcuni, per esempio, sono accompagnati da indici di persone, luoghi e opere, altri no. Tali disparità sono probabilmente riconducibili a lavori in corso ma è difficile sapere con esattezza non dispondendo di informazioni correnti.
- Il progetto non risponde a domande di ricerca specifiche ma si presenta come "una banca dati centrale per tutti i giornali moralistici europei" aperta a coloro che nutrono un interesse per il periodo storico e il genere letterario. L'accesso alla risorsa è diretto: non richiede fastidiosi *login* o l'installazione di particolari programmi per l'utilizzo.
- I giornali sono catalogati per paese d'origine e divisi nei loro rispettivi volumi o numeri. Ciascun volume presenta una scheda di metadati esauriente (Fig. 1) e alcune di queste informazioni sono opportunamente riunite e ripetute a capo pagina in forma di citazione bibliografica (Fig. 2).



Fig. 3: Un esempio d'interattività nel testo della *Gazzetta Veneta*. La selezione della casella 'Opere' nel menù di destra evidenzia in rosso le opere citate nel testo.

La trascrizione del testo di ogni giornale è disponibile in due formati scaricabili, XML (TEI P5) e PDF, identificabili tramite le corrispondenti icone. Una barra menù laterale offre ulteriori opzioni di visualizzazione (Fig. 2): la 'Versione stampabile (Visualizzazione Originale)' è un duplicato del PDF già accessibile mediante l'icona rossa; la 'Versione stampabile (Visualizzazione dell'analisi)' consiste nella combinazione del formato PDF e di annotazioni testuali; 'XML (TEI P5)' è un duplicato del file già accessibile mediante l'icona gialla; e, infine, 'Voyant Tools' conduce l'utente a una visione interattiva e statistica del testo. Le caselle di controllo sottostanti conservano una certa interattività anche all'interno di MWS: la selezione della casella 'Opere', per esempio, evidenzia in rosso tutte le opere citate nel testo (Fig. 3), lo stesso vale per la casella 'Luoghi' come anche per le altre. La voce del menù 'Visualizzazione dell'analisi' permette di consultare le annotazioni di ciascun testo.

7 MWS non fornisce traduzioni dei testi né una bibliografia sull'argomento o elenco di pubblicazioni derivanti dal progetto.

#### Metodi

- Le scelte di grafica della risorsa non sono precisate ma la struttura organizzativa dei testi è intuitiva e semplice. Non è precisato quante e quali lingue andranno a costituire il database né quale sarà il numero totale dei giornali e quando essi diverrano elettronicamente fruibili. Attualmente (marzo 2017) il progetto mette a disposizione 11 giornali italiani, 22 spagnoli, 4 tedeschi, 15 francesi e 1 inglese per un totale di 53 giornali. Questo numero, tuttavia, non coincide con i 44 giornali riportati nella voce 'Lista Completa' del menù principale.
- 9 Il progetto non fornisce informazioni circa le fonti dei testi qui raccolti, per cui non è dato sapere se e come la digitalizzazione è stata effettuata o se i giornali provengano da

edizioni digitali esistenti. Non avendo accesso alle scannerizzazioni dei giornali e mancando di documentazione circa i criteri di codifica XML (TEI P5), non è chiaro se le trascrizioni di MWS siano diplomatiche, semi-diplomatiche o interpretative. Da una rapida analisi dell'ortografia, si potrebbe ipotizzare una diplomatica.

Difficile quantificare il corpus con esattezza: un *download* del progetto dal server austriaco conta circa 4.5 GB di dati fra giornali (inclusi alcuni testi non attualmente disponibili in rete ma probabilmente prossimi alla pubblicazione), metadati e file annessi. Le numerosissime cartelle e sottocartelle non permettono di calcolare il numero totale di parole ma, date le dimensioni del *download*, si tratta sicuramente di centinaia di migliaia di *tokens*.



Fig. 4: Esempio di un testo annotato e del rapporto gerarchico delle annotazioni.

```
▼<encodingDesc>
 ▼<editorialDecl>
   ▼<interpretation>
    ▼<ab type="interpGrp">
      ▼<interpGrp type="Narrative_Darstellungsebenen">
         <interp xml:id="E1">Ebene 1</interp>
         <interp xml:id="E2">Ebene 2</interp>
         <interp xml:id="E3">Ebene 3</interp>
         <interp xml:id="E4">Ebene 4</interp>
         <interp xml:id="E5">Ebene 5</interp>
         <interp xml:id="E6">Ebene 6</interp>
       </interpGrp>
      ▼<interpGrp type="Narrative Darstellungsformen">
         <interp xml:id="AE">Allgemeine Erzählung</interp>
         <interp xml:id="SP">Selbstportrait</interp>
         <interp xml:id="FP">Fremdportrait</interp>
         <interp xml:id="D">Dialog</interp>
         <interp xml:id="AL">Allegorisches Erzählen</interp>
         <interp xml:id="TR">Traumerzählung</interp>
         <interp xml:id="F">Fabelerzählung</interp>
         <interp xml:id="S">Satirisches Erzählen</interp>
         <interp xml:id="EX">Exemplarisches Erzählen</interp>
         <interp xml:id="UT">Utopische Erzählung</interp>
         <interp xml:id="MT">Metatextualität</interp>
         <interp xml:id="ZM">Zitat/Motto</interp>
         <interp xml:id="LB">Leserbrief</interp>
       </interpGrp>
      </ab>
    </interpretation>
  </editorialDecl>
 </encodingDesc>
```

Fig. 5: L'elemento <encodingDesc> nel <teiHeader> del giornale *II Filosofo alla Moda:*Lezione IV contiene la lista, in tedesco, di possibili annotazioni, ognuna contrassegnata da

un xml:id.

La qualità dei dati è alta e non necessita quindi di correzioni. La mole del corpus, sebbene ancora in fase di elaborazione, e la cura dei testi permettono all'utente di usare questo progetto per analisi su larga scala.

```
▼<seg synch="#FR.6" type="E3">
▼<seg synch="#FR.7" type="EX">
Ho conosciuto tal'uno, che per avere veduta una stella volante, non ha potuto dormire in tutta la notte. Ho veduto un languido amante impallidire, e lasciar di mangiare, per avere rotta la lunetta d'un volatile. E'accaduto più volte, che il canto d'un Guffo, udito a mezza notte, ha cagionato più sconcerto in una Famiglia di quello avrebbe fatto una truppa di ladri. Che più! la debole voce d'un Grillo ha impresso maggiore spavento di quello avesse fatto il ruggito d'un Leone. La minima bagatella puole deventare spaventacchio ad una immaginazione offesa. Un Chiodo vecchio arrugginito; una spilla uncinata; l'incontro di certa specie di Persone si convertono in prodigj. </seg>
</seg>
</seg>
```

Fig. 6: Nell'elemento <body> dello stesso giornale le relative annotazioni vengono invocate tramite l'xml:id dichiarato nel <teiHeader> (vedi <u>Fig. 5</u>). Qui viene riportato un *Livello 3* (@type=E3) e un *esempio* (@type=EX).

Le annotazioni esistenti sono divise in due categorie: i *Livelli di rappresentazione narrativi* dividono il testo in unità narrative (per un massimo di sei) e ne illustrano i rapporti gerarchici; le *Forme narrative*, invece, descrivono la tipologia o forma, appunto, di queste unità. L'esempio in Figura 4 riporta solo tre *livelli di rappresentazione narrativa*: il Livello 1 contiene la forma *citazione* e il Livello 2; il Livello 2 contiene il Livello 3, il quale, a sua volta, contiene la forma narrativa *Racconto generale*. L'utente può selezionare e spuntare livelli e forme narrative a piacere. Le annotazioni di ciascun giornale non vengono effettuate separatamente ma all'interno della corrispettiva trascrizione XML (TEI P5), come illustrato in Fig. 5 e 6. Bisogna sottolineare nuovamente come l'assenza di documentazione non dia informazioni sulla visualizzazione di tali annotazioni o di come queste possano essere gestite nel *workflow*. Si suppone, pertanto, che, dato il formato di codifica, ci siano uno o più fogli di stile XSLT (Extensible Stylesheet Language Transformations).

```
v<TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance
xsi:schemaLocation="http://www.tei-c.org/ns/1.0 http://gams.uni-graz.at/mws/schema/mws.xsd">
  ▼<teiHeader type="mws"
    ▼<fileDesc>
         <titleStmt>
  <title>Lezione IV</title</pre>
           <author>Cesare Frasponi</author>
         </titleStmt>
      ▼<editionStmt>
         <edition>Moralische Wochenschriften</edition>

             <name>Alexandra Fuchs</name>
             <resp>Editor</resp>
      </respStmt>
</editionStmt>
v<publicationStmt>
<publicationStmt>
<publicationStmt>
<publicationStmt>
<publicationStmt>
<publicationStmt>
<publicationStmt>
<idate when="2016-05-03">03.05.2016</date>
<idate when="2016-05-03">03.05.2016</date>
<idate vpublicationStmt>
v<publicationStmt>
v<publicationStmt>
v<publicationStmt>
v<publicationStmt>
v<publicationStmt>
v<publicationStmt>
v<publicationStmt>
v<publicationStmt</pr>
vprespond</pr>
courseDesc
             Frasponi, Cesare: Il Filosofo alla Moda, ovvero, Il Maestro Universale, Venezia: Giovanni Malachino 1727, 21-27
         <date>1727</date>
          <placeName key="#GID.1">Italien</placeName>
</bibl>
      </sourceDesc>
     <encodingDesc>...</encodingDesc>
      chrofileDesc>
        ▼<langUsage>
        <language ident="it">Italian</language>
</language>
</textClass>...</textClass>
         <textClass>
         ▼<keywords scheme="cirilo:normalizedPlaceNames">
           ▼<list>
                ▼<location>
                    <geo>12.83333,42.83333</geo>
</location>
```

Fig. 7: I metadati relativi a *Il Filosofo alla Moda: Lezione IV* sono forniti nel <teiHeader> del file di trascrizione XML (TEI P5).

Nei file di trascrizione XML (TEI P5) troviamo anche i metadati relativi ai fogli e agli editori responsabili della codifica (Fig. 7). Alcuni di questi dati si sovrappongono ai

metadati archivistici forniti dalle schede *Dublin Core* di ciascun giornale. <sup>5</sup> Le trascrizioni XML (TEI P5) del progetto sono governate da uno schema XML <sup>6</sup> e l'infrastruttura del progetto è predisposta per la connessione ad altre risorse digitali tramite metodologie di *Linked Open Data* (LOD), come vedremo più avanti.

#### Dati forniti e riutilizzo

Tutti i testi di MWS, come anticipato in precedenza, sono individualmente scaricabili sia in formato PDF che XML (TEI P5). Una versione TXT, spesso preferita da coloro che svolgono analisi testuali automatiche, non è fornita ma reperibile in tre modi: rimuovendo tutti i marcatori dal formato XML; copiando, incollando e ripulendo il testo HTML così come presentato nel sito; oppure, se in possesso di abilità tecniche, scaricando l'intera risorsa e rimuovendo i dati non necessari in maniera automatica. Attualmente, infatti, il progetto manca di un metodo efficiente per il download completo del corpus, sia esso per mezzo di un API, di un'architettura REST o di un archivio .zip. 8

#### Interfaccia

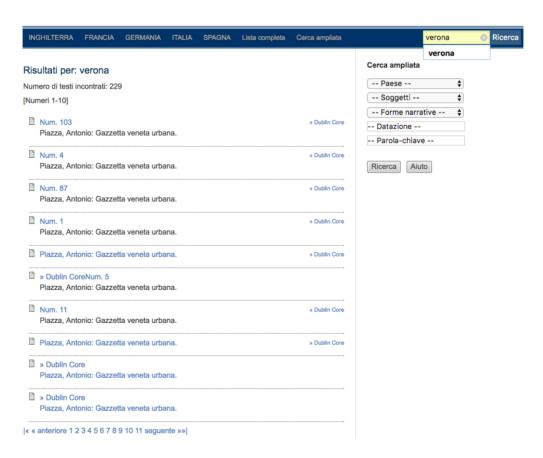


Fig. 8: La schermata dei risultati relativi alla ricerca semplice del termine 'verona'. La schermata è ambigua in quanto le opzioni di ricerca avanzata nel menù di destra non servono a filtrare questi risultati ulteriormente.

- MWS è fruibile in francese, inglese, italiano, spagnolo e tedesco, ma alcune pagine sono solamente disponibili in inglese e tedesco. Ciononostante, grazie alla semplicità dell'interfaccia, la ricerca dei testi risulta estremamente intuitiva. Le voci del menù non sono ambigue ma ci si aspetterebbe più contenuto nelle pagine d'introduzione al progetto.
- 16 L'interfaccia permette di consultare i giornali *singolarmente*, non l'intero corpus. Si possono effettuare sia ricerche semplici mediante l'apposita barra di ricerca sia ricerche più avanzate cliccando su 'Cerca ampliata' nel menù principale. Come si può osservare dalla Fig. 8, una ricerca semplice del termine 'verona' riporta i testi che lo contengono. La schermata dei risultati diventa però ambigua quando l'utente vuole fare delle ricerche più raffinate. La ricerca ampliata a destra dei risultati, suggerisce la possibilità di aggiungere ulteriori filtri. Selezionando, però, ad esempio, 'Passione' dalla voce 'Soggetti', i risultati precedenti vengono annullati e sostituiti con quelli pertinenti a 'Passione'. Per ottenere risultati che contengano il termine 'verona' assieme a un ulteriore soggetto come 'Passione' bisogna, quindi, usare un operatore Boolean nella barra di ricerca semplice, ovvero 'verona AND passione'. Le modalità di ricerca sono descritte in una pagina separata, la pagina 'Aiuto' attualmente disponibile solo in inglese e tedesco e accessibile tramite il bottone posto sotto 'Cerca ampliata'. Per risolvere questa ambiguità e assistere gli utenti nella ricerca, tali istruzioni potrebbero essere rese più visibili nelle schermate di ricerca.
- Per l'analisi dei testi, invece, l'utente può scegliere di manipolarli tramite l'integrazione con *Voyant Tools* o in via del tutto indipendente. *Voyant Tools* è uno strumento web di analisi testuale che offre visualizzazioni statistiche d'insieme per ogni testo (word-cloud, occorrenze di parole, dimensioni del testo, ecc.). L'utilità di tali visualizzazioni è soggettiva ma l'intento di questo strumento è di riproporre i testi in chiave statistica e relazionale. Per analisi più avanzate, è necessario scaricare i file di cui si ha bisogno in locale. Il progetto purtroppo non fornisce l'opzione di scaricare la totalità dei dati come file unico. Una soluzione è di recuperare i dati con il comando *wget* per mezzo dell'interfaccia a linea di comando. In questo modo però non vengono solo scaricati i file XML (TEI P5) ma una quantità ingente di file attinenti, quali metadati e le pagine di presentazione per ogni giornale nelle lingue proposte. 10
- L'unico elemento interattivo dell'interfaccia di questo progetto si basa sulla possibilità di (de)selezionare diversi livelli d'annotazione.

#### **Preservazione**

La documentazione attuale del progetto disponibile sul sito MWS è disomogenea e non sufficientemente dettagliata soprattutto per quanto riguarda le tempistiche e le scelte editoriali. Una panoramica tecnica è invece accessibile tramite il sito ufficiale del Zentrum für Informationsmodellierung, responsabile per lo sviluppo dell'infrastruttura GAMS (Geisteswissenschaftliches Asset Management System) che sostiene il progetto. GAMS si basa su FEDORA (Flexible Extensible Digital Object Repository Architecture) e segue perciò gli standard di preservazione da esso previsti, nonché quelli di OAIS (Open Archival Information System) e di Linked Open Data (LOD). La documentazione di GAMS dedica un'intera sezione all'esposizione del concetto di preservazione digitale a lungo termine e a come GAMS ne adempia i requisiti. Nel 2014, infatti, la certificazione *Data Seal of Approval* ha riconosciuto GAMS come "trusted digital repository" e, pertanto, come iniziativa impegnata a garantire il continuo accesso alle proprie risorse digitali.

Il progetto è protetto da una licenza *Creative Commons Attribution-NonCommercial-ShareAlike*, la quale regola l'uso di tutti i dati costituenti MWS. Per ogni giornale, GAMS fornisce un URL permanente generato dal server del Zentrum für Informationsmodellierung 13 nonché un identificatore permanente (PID) nel <teiHeader> di ciascun documento.

### **Conclusione**

In conclusione, MWS, sebbene ancora in fase di elaborazione, raggiunge lo scopo di rendere elettronicamente fruibili una collezione di giornali moralistici europei del diciottesimo secolo. Lo *USP* o "argomentazione esclusiva di vendita" del corpus è un progetto di respiro europeo, in quanto gli utenti possono studiare e paragonare lo sviluppo di questo genere letterario nelle diverse nazioni in un determinato periodo e contesto storico. Le opportunità di ricerca offerte dal progetto sono molte e di potenziale interesse per un pubblico vasto (linguisti, storici, critici letterari, ecc.).

C'è tuttavia spazio per migliorare: quanto più la qualità dei dati forniti è alta, tanto meno lo è quella della documentazione circa le scelte editoriali, la grandezza stimata del corpus e la tempistica di pubblicazione dei successivi numeri. Si suggerisce quindi una maggiore attenzione alla stesura di una documentazione dettagliata e di prevedere un archivio di tutti i dati per un download completo dell'attuale collezione.

#### **Note**

- 1. https://web.archive.org/web/20170827081704/http://www.europeananewspapers.eu/.
- 2. Vedi: <a href="http://gams.uni-graz.at/archive/objects/context:mws/methods/sdef:Context/get?">http://gams.uni-graz.at/archive/objects/context:mws/methods/sdef:Context/get?</a> mode=imprint&locale=en (Consultato: 3 aprile 2017).
- 3. Vedi: https://voyant-tools.org/ (Consultato: 3 aprile 2017).
- 4. La consultazione del <teiHeader> delle versioni XML dei testi, dove queste informazioni vengono spesso riportate, non contiene dettagli a riguardo.
- <u>5.</u> La Dublin Core® Metadata Initiative (DCMI) è un'iniziativa volta alla gestione e allo sviluppo di standard di archiviazione digitale. <a href="http://dublincore.org/">http://dublincore.org/</a> (Consultato: 3 aprile 2017).
- <u>6.</u> Disponibile qui: <a href="http://gams.uni-graz.at/mws/schema/mws.xsd">http://gams.uni-graz.at/mws/schema/mws.xsd</a> (Consultato: 3 aprile 2017) conforme all' *ISOcat Data Category Registry*.
- 7. Per esempio, tramite l'interfaccia a linea di comando con 'wget -r'.
- 8. Come quello molto comodo proposto dalla biblioteca digitale digilibLT. Vedi: <a href="http://digiliblt.lett.unipmn.it/g">http://digiliblt.lett.unipmn.it/g</a> bulk opere.php (Consultato: 3 aprile 2017).
- 9. Per esempio, la pagina 'Aiuto'.
- 10. Si consiglia perciò a coloro che desiderino scaricare la risorsa in questo modo di fare ciò su un server o di liberare circa 5GB di spazio sul proprio computer.
- 11. Disponibile qui: <a href="https://web.archive.org/web/20170827082311/http://gams.uni-graz.at/docs">https://web.archive.org/web/20170827082311/http://gams.uni-graz.at/docs</a> (Consultato: 3 aprile 2017).
- 12. Come riportato nella documentazione tecnica di GAMS, una componente dell'infrastruttura del progetto, Cirilo, può essere utilizzata come OAI harvester e quindi essere collegata a risorse quali Europeana. Vedi: <a href="https://web.archive.org/web/20170827082311/http://gams.uni-graz.at/docs#cirilo">https://web.archive.org/web/20170827082311/http://gams.uni-graz.at/docs#cirilo</a> (Consultato: 3 aprile 2017).
- 13. Ogni URL GAMS è prefissato dal codice 11471. Per esempio 'hdl.handle.net/ 11471/513.20.1803'.

## **Bibliografia**

Ertler, Klaus-Dieter. 2012. "Moral Weeklies (Periodical Essays)", *EGO - Europäische Geschichte Online*, available at

http://ieg-ego.eu/en/threads/models-and-stereotypes/anglophilia/klaus-dieter-ertler-moral-weeklies-periodical-essays.

## **Factsheet**

Resource reviewed	
Title	Spectateurs (Moralische Wochenschriften)
Editors	Klaus-Dieter Ertler, Alexandra Fuchs, Michaela Fischer, Elisabeth Hobisch
URI	http://gams.uni-graz.at/context:mws
Publication Date	2011
Date of last access	08.08.2017

Reviewer	
Surname	Franzini
First Name	Greta
Organization	Georg-August-Universität Goettingen
Place	Goettingen, Germany
Email	gfranzini (at) etrap.eu

General Information		
Bibliographic description	Can the text collection be identified in terms similar to traditional bibliographic descriptions (title, responsible editors, institution, date(s) of publication, identifier/address)?  (cf. Catalogue 1.1)	yes
Contributors	Are the contributors (editors, institutions, associates) of the project documented? (cf. Catalogue 1.3)	yes
Contacts	Is contact information given? (cf. Catalogue 1.4)	yes
Aims		
Documentation	Is there a description of the aims and contents of the text collection? (cf. Catalogue 2.1)	yes
Purpose	What is the purpose of the text collection? (cf. Catalogue 2.2)	General purpose

Kind of research	What kind of research does the collection allow to conduct primarily? (cf. Catalogue 3.1.8)	Quantitative research
Self-classification	How does the text collection classify itself (e.g. in its title or documentation)? (cf. Catalogue 2.3)	Database
Field of research	To which field(s) of research does the text collection contribute? (cf. Catalogue 2.2)	Literary studies
Content		
Era	What era(s) do the texts belong to? (cf. Catalogue 2.5)	Early Modern
Language	What languages are the texts in? (cf. Catalogue 2.5)	English, French, German, Italian, Spanish
Types of text	What kind of texts are in the collection? (cf. Catalogue 2.5)	Newspaper/ journal articles
Additional information	What kind of information is published in addition to the texts? (cf. Catalogue 2.5)	Introduction
Composition		
Documentation	Are the principles and decisions regarding the design of the text collection, its composition and the selection of texts documented? (cf. Catalogue 3.1.1-3.1.3)	no
Selection	What selection criteria have been chosen for the text collection? (cf. Catalogue 3.1)	Genre
Size		
Texts/records	How large is the text collection in number of texts/ records? (cf. Catalogue 3.1.4)	51-100
Tokens	How large is the text collection in number of tokens? (cf. Catalogue 3.1.4)	100,000- 1 Mio.
Structure	Does the text collection have identifiable sub- collections or components? (cf. Catalogue 3.1.5)	yes
Data acquisition and	d integration	
Text recording	Does the text collection record or transcribe the textual data for the first time? (cf. Catalogue 3.1.6)	yes

Text integration	What kind of material has been taken over from other sources? (cf. Catalogue 3.1.6)	none
Quality assurance	Has the quality of the data (transcriptions, metadata, annotations, etc.) been checked? (cf. Catalogue 3.1.7)	yes
Typology	Considering aims and methods of the text collection, how would you classify it further? For definitions please consider the help-texts. (cf. Catalogue 3.1.8)	Corpus, Reference corpus, Parallel corpus
Data Modelling		
Text treatment	How are the textual sources represented in the digital collection? (cf. Catalogue 3.2.1)	Diplomatic transcription
Basic format	In which basic format are the texts encoded? (cf. Catalogue 3.2.4)	XML
Annotations		
Annotation type	With what information are the texts further enriched? (cf. Catalogue 3.2.2)	Linguistic annotations, Structural information
Annotation integration	How are the annotations linked to the texts themselves? (cf. Catalogue 3.2.2)	Embedded
Metadata		
Metadata type	What kind of metadata are included in the text collection? (cf. Catalogue 3.2.3)	Descriptive
Metadata level	On which level are the metadata included? (cf. Catalogue 3.2.2)	Individual texts
Data schemas and st	andards	
Schemas	What kind of data/metadata/annotation schemas are used for the text collection? (cf. Catalogue 3.2.4)	Project specific schema
Standards	Which standards for text encoding, metadata and annotation are used in the text collection? (cf. Catalogue 3.2.4)	TEI, Dublin Core
Provision		
Accessability of the basic data	Is the textual data accessible in a source format (e.g. XML, TXT)? (cf. Catalogue 4.1)	yes

Download	Can the entire raw data of the project be downloaded (as a whole)? (cf. Catalogue 4.2)	no
Technical interfaces	Are there technical interfaces which allow the reuse of the data of the text collection in other contexts? (cf. Catalogue 4.2)	none
Analytical data	Besides the textual data, does the project provide analytical data (e.g. statistics) to download or harvest?  (cf. Catalogue 4.3)	no
Reuse	Can you use the data with other tools useful for this kind of content? (cf. Catalogue 4.4)	yes
User Interface		
Interface provision	Does the text collection have a dedicated user interface designed for the collection at hand in which the texts of the collection are represented and/or in which the data is analyzable? (cf. Catalogue 5.1)	yes
User Interface quest	ions	
Usability	From your point of view, is the interface of the text collection clearly arranged and easy to navigate so that the user can quickly identify the purpose, the content and the main access methods of the resource?  (cf. Catalogue 5.3)	yes
Acces modes		
Browsing	Does the project offer the possibility to browse the contents by simple browsing options or advanced structured access via indices (e.g. by author, year, genre)?  (cf. Catalogue 5.4)	yes
Fulltext search	Does the project offer a fulltext search? (cf. Catalogue 5.4)	yes
Advanced search	Does the project offer an advanced search? (cf. Catalogue 5.4)	yes
Analysis		
Tools	Does the text collection integrate tools for analyses of the data? (cf. Catalogue 5.5)	yes

Customization	Can the user alter the interface in order to affect the outcomes of representation and analysis of the text collection (besides basic search functionalities), e.g. by applying his or her own queries or by choosing analysis parameters?  (cf. Catalogue 5.5)	no
Visualization	Does the text collection provide particular visualizations of the data? (cf. Catalogue 5.6)	no visualization
Personalization	Is there a personalisation mode that enables the users e.g. to create their own sub-collections of the existing text collection?  (cf. Catalogue 5.7)	no
Preservation		
Documentation	Does the text collection provide sufficient documentation about the project in general as well as about the aims, contents and methods of the text collection?  (cf. Catalogue 6.1)	no
Open Access	Is the text collection Open Access? (cf. Catalogue 6.2)	yes
Rights		
Declared	Are the rights to (re)use the content declared? (cf. Catalogue 6.2)	yes
License	Under what license are the contents released? (cf. Catalogue 6.2)	CC-BY-NC, CC- BY-NC-SA
Persistent identification and addressing	Are there persistent identifiers and an addressing system for the text collection and/or parts/objects of it and which mechanism is used to that end? (cf. Catalogue 6.3)	Persistent URLs
Citation	Does the text collection supply citation guidelines? (cf. Catalogue 6.3)	yes
Archiving of the data	Does the documentation include information about the long term sustainability of the basic data (archiving of the data)?  (cf. Catalogue 6.4)	no
Institutional curation	Does the project provide information about institutional support for the curation and sustainability of the project?  (cf. Catalogue 6.4)	yes
Completion	Is the text collection completed? (cf. Catalogue 6.4)	no
Personnel		

Contributors	Klaus-Dieter Ertler Alexandra Fuchs Elisabeth Hobisch Michaela Fischer Hubert Stigler Martina Semlak Alexandra Fuchs Ulrike Rieger
--------------	--