



Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

Live long and prosper! Lessons from the TEI

TEI @ Oxford

September 2008

1986 was a long time ago...

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- The first computer virus – Brain – appears, in the USA
- Construction of the channel tunnel begins
- The Soviet Union launches space station Mir
- Disaster at Chernobyl
- Olaf Palme assassinated
- Records of the year: *Raising Hell* (Run DMC)... *Graceland* (Paul Simon)... *Группа крови* (Виктор Цой)

1986 was a long time ago...

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- The first computer virus – Brain – appears, in the USA
- Construction of the channel tunnel begins
- The Soviet Union launches space station Mir
- Disaster at Chernobyl
- Olaf Palme assassinated
- Records of the year: *Raising Hell* (Run DMC)... *Graceland* (Paul Simon)... *Группа крови* (Виктор Цой)

1986 was a long time ago...

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- The first computer virus – Brain – appears, in the USA
- Construction of the channel tunnel begins
- The Soviet Union launches space station Mir
- Disaster at Chernobyl
- Olaf Palme assassinated
- Records of the year: *Raising Hell* (Run DMC)... *Graceland* (Paul Simon)... *Группа крови* (Виктор Цой)

1986 was a long time ago...

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- The first computer virus – Brain – appears, in the USA
- Construction of the channel tunnel begins
- The Soviet Union launches space station Mir
- Disaster at Chernobyl
- Olaf Palme assassinated
- Records of the year: *Raising Hell* (Run DMC)... *Graceland* (Paul Simon)... *Группа крови* (Виктор Цой)

1986 was a long time ago...

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- The first computer virus – Brain – appears, in the USA
- Construction of the channel tunnel begins
- The Soviet Union launches space station Mir
- Disaster at Chernobyl
- Olaf Palme assassinated
- Records of the year: *Raising Hell* (Run DMC)... *Graceland* (Paul Simon)... *Группа крови* (Виктор Цой)

1986 was a long time ago...

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- The first computer virus – Brain – appears, in the USA
- Construction of the channel tunnel begins
- The Soviet Union launches space station Mir
- Disaster at Chernobyl
- Olaf Palme assassinated
- Records of the year: *Raising Hell* (Run DMC)... *Graceland* (Paul Simon)... *Группа крови* (Виктор Цой)

...but we used computers then

Live long and
prosper!
Lessons from
the TEI
TEI @ Oxford

- Corpus linguistics
- Databases on CD ROM
- Large-scale lexical resources already existed (eg TLF, TLG, LASLA...)
- Digital lexicography (e.g. OED)
- Document management systems (e.g. TeX, Scribe, tRoff..)
 - some proprietary (and expensive), some research
- Text archives
- Hypertext theory

But there was no world wide web and not many desktop pcs...

Birth of the Text Encoding Initiative

Live long and
prosper!

Lessons from
the TEI

TEI @ Oxford

- Spring 1987: European workshops on standardisation of historical data (J.P. Genet, M Thaller)
- Autumn 1987: NEH funds an exploratory international workshop on the feasibility of defining "text encoding guidelines"





Today's question:

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- So the TEI is *very old!*
- It comes from a time before the Web, before the DVD, the mobile phone, cable tv, or Microsoft Excel
- Not much in computing survives 5 years, never mind 20
- What relevance can it possibly have today?
- Why is it still here, and how has it survived?

Is the TEI still relevant?

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- With XML everyone can create their own markup system and still share data!
- In the Semantic Web, XML systems will all understand each other's data!

If we have

- historical data marked up with a Historical Markup Language
- linguistic data marked up with a Linguistic Markup Language
- metadata marked up with a Metadata Markup Language

how will we integrate resources or ask interesting questions?

Haven't we been here before?

Relevance 1

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

The TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas and documentation in various languages and formats
- a large and active open source style user community

Relevance 1

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

The TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas and documentation in various languages and formats
- a large and active open source style user community

The TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas and documentation in various languages and formats
- a large and active open source style user community

The TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas and documentation in various languages and formats
- a large and active open source style user community

The TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas and documentation in various languages and formats
- a large and active open source style user community

The TEI provides

- a language-independent framework for defining markup languages
- a very simple consensus-based way of organizing and structuring textual (and other) resources...
- ... which can be enriched and personalized in highly idiosyncratic or specialised ways
- a very rich library of existing specialised components
- an integrated suite of standard stylesheets for delivering schemas and documentation in various languages and formats
- a large and active open source style user community

Relevance 2

Live long and
prosper!
Lessons from
the TEI
TEI @ Oxford

Why would you want those things?

- because we need to interchange resources
 - between people
 - (increasingly) between machines
- because we need to integrate resources
 - of different media types
 - from different technical contexts
- because we need to preserve resources
 - cryogenics is not the answer!
 - we need to preserve metadata as well as data

Why would you want those things?

- because we need to interchange resources
 - between people
 - (increasingly) between machines
- because we need to integrate resources
 - of different media types
 - from different technical contexts
- because we need to preserve resources
 - cryogenics is not the answer!
 - we need to preserve metadata as well as data

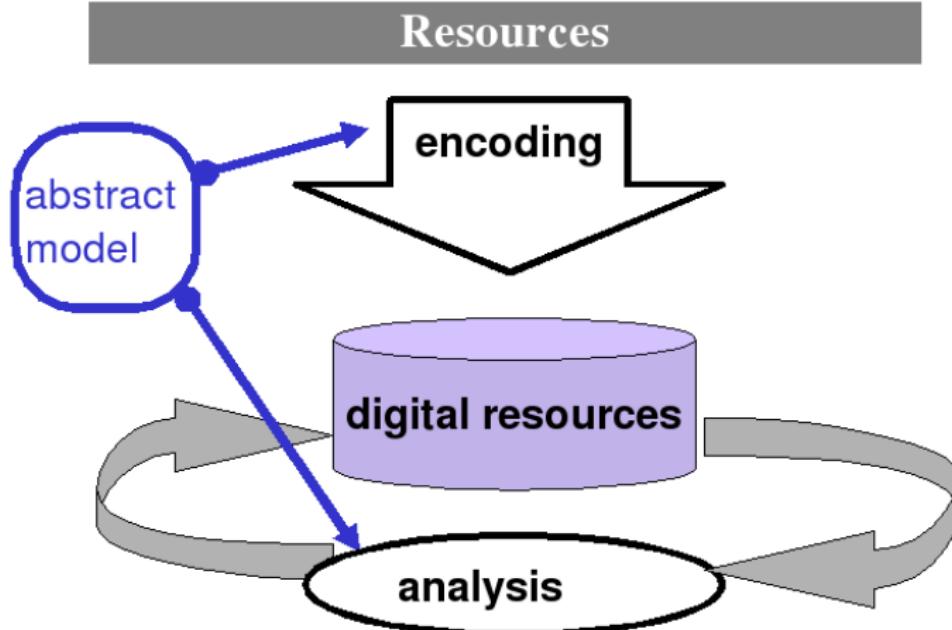
Why would you want those things?

- because we need to interchange resources
 - between people
 - (increasingly) between machines
- because we need to integrate resources
 - of different media types
 - from different technical contexts
- because we need to preserve resources
 - cryogenics is not the answer!
 - we need to preserve metadata as well as data

The virtuous circle of encoding

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford



The scope of intelligent markup

Even within the original scope of the TEI we have

- basic structural and functional components
- diplomatic transcription, images, annotation
- links, correspondence, alignment
- data-like objects such as dates, times, places, persons, events (named entity recognition)
- meta-textual annotations (correction, deletion, etc)
- linguistic analysis at all levels
- contextual metadata of all kinds
- ... and so on and so forth

Is it possible to delimit encyclopaedically all possible kinds of markup?

Reasons for attempting to define a common framework

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- re-usability and repurposing of resources
- modular software development
- lower training costs
- ‘frequently answered questions’ — common technical solutions for different application areas

The TEI was designed to support multiple views of the same resource

Old Skool TEI

Live long and
prosper!

Lessons from
the TEI

TEI @ Oxford

- A traditional (if large) research project with soft funding, driven by academic curiosity
- a codification of best practice, with no formal maintenance method
- uncertain licencing and development practices
- perceived as unmanageably complex except by the priesthood — or simultaneously as too simple for real scholarly work
- lack of specific tools to *do* something with a TEI text
- failure to market the advantages of rich markup

- Proper open source licence, with visible development on Sourceforge
- Architecture rethought to facilitate expansion and integration with other systems
- Self documenting, each release fully validated, delivered using standard mechanisms
- Publicly available processing tools managed together with the Guidelines
- Active developer community, wiki, etc. Test files, exemplars, regular updates...
- New governance structure, new tools, new modules...

Three important things about TEI P5

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- ➊ Being a good digital citizen:
 - Support for multiple schema languages and namespaces
 - Reliance on XML, and hence on Unicode
 - Validation of attributes and datatyping
 - Use of W3C pointers and paths
- ➋ Making it flexible:
 - ODD: a single specification language for developers, users, and teachers, integrating schema and documentation;
 - Verifiable conformance
- ➌ Old annoyances removed and some new topics added

One Specification Language

- A set of TEI documents is described by an ODD, which is itself a TEI document that combines:
 - references to existing declarations
 - formal declarations for elements and attributes
 - documentation and usage notes
- Underlying this:
 - a conceptual model which abstracts from specific elements to generic classes
 - a modular architecture for combining sets of definitions
- specifications are chainable; modifications are written in ODD with ODD as input and output
- Roma is one interface to this: there will be others

For example

An ODD file is a valid TEI document, containing descriptive prose, and a `<schemaSpec>` element to define the schema it documents

```
<div>
  <head>Our Project Manual</head>
  <p>In this project we use the basic TEI structures
    with a few minor modifications to exclude
    elements we do not need</p>
  <schemaSpec ident="TEI-minimal" start="TEI">
    <moduleRef key="tei"/>
    <moduleRef key="header"/>
    <moduleRef key="core"/>
    <moduleRef key="textstructure"/>
    <!-- We don't need these drama elements: -->
    <elementSpec ident="sp" mode="delete" module="core"/>
    <elementSpec ident="speaker" mode="delete" mod-
ule="core"/>
  </schemaSpec>
</div>
```

Support for many schema languages

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- TEI schemas can be generated for
 - XML DTD language
 - ISO RELAX NG language
 - W3C Schema Language
- Content models are defined using RELAX NG syntax
- Datatypes are defined in terms of W3C datatypes
- Some facilities (e.g. alternation, namespaces) cannot be expressed in DTD
- Additional constraints can be expressed in Schematron



Two reasons why standards fail

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- The theory is not yet ripe
- The "not invented here" attitude: the community of users is too diverse

Coping with partially-baked ideas

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

In a TEI ODD, you can ...

- constrain the domain of a value list
- enforce Schematron rules about e.g. co-dependency
- provide new elements in your own namespace
- remove (non-mandatory) child elements

New elements

Live long and
prosper!
Lessons from
the TEI
TEI @ Oxford

A schema is a grammar. How can you add new terminals to an existing syntax?

- All content models are expressed indirectly, by reference to element classes rather than elements
- Hence adding a new element is simply a matter of saying which class/es it belongs to

The TEI schema is also enriched with semantics. How can you explain what a new element means?

- Class membership also conveys some semantics
- ODD includes detailed documentation

Coping with the NIH Syndrome

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- TEI P5 has extensive I18N features for translation of ...
 - schema objects
 - schema documentation
- See *Roma* at <http://www.tei-c.org/Roma/>
- TEI is hospitable to other namespaces
 - You can use SVG for graphics, MathML for math, Word Table markup if you like
 - (but note this doesn't solve the Other Overlap Problem)
- ODD also includes an <equiv> element for mapping to external ontologies

For example

Live long and
prosper!
Lessons from
the TEI
TEI @ Oxford

Embedding SVG within TEI:

```
<figure>
  <svg xmlns="http://www.w3.org/2000/svg"
    width="6cm" height="5cm" viewBox="6 3 6 5">
    <ellipse xmlns="http://www.w3.org/2000/svg"
      style="fill: #ffffff" cx="9.75" cy="6.35" rx="2.75" ry="2.75">
  </svg>
</figure>
```

A user-defined attribute:

```
<div
  xmlns:my="http://www.example.org/ns/nonTEI">
  <p n="12" my:topic="rabbits">Flopsy, Mopsy, Cottontail,
  and Peter...</p>
</div>
```

An NVDL processor can validate a document using multiple namespace schemas

Conformance issues

A document is TEI Conformant if and only if it ...

- is a well-formed XML document
- can be validated against a TEI Schema, that is, a schema derived from the TEI Guidelines
- conforms to the TEI Abstract Model
- uses the TEI Namespace (and other namespaces where relevant) correctly
- is documented by means of a TEI Conformant ODD file which refers to the TEI Guidelines

Or if it can be transformed automatically using some TEI-defined procedures into such a document (it is TEI-conformable)

Standardization should not mean 'Do what I do', but rather
'Explain what you do in terms I can understand'



Evolution works!

Live long and
prosper!
Lessons from
the TEI

TEI @ Oxford

- ① Make modifications in your own namespace
- ② Document them in an ODD
- ③ Propose them to the TEI Council as amendments or feature requests
- ④ TEI P5 now has a 6 month release cycle...

Visit <http://www.tei-c.org> for more background info

Visit <http://tei.sf.net> to download

Using the basic TEI structural elements

TEI @ Oxford

September 2008

TEI Infrastructure

Using the basic
TEI structural
elements

TEI @ Oxford

- The TEI encoding scheme consists of a number of modules
- These declare XML elements and their attributes
- An element's declaration assigns it to one (or more) model classes
- Another part declares its possible content and attributes with reference to these classes
- This indirection allows strength and flexibility
- It makes it easy to add/exclude new elements by referencing existing classes

What is a module?

Using the basic
TEI structural
elements

TEI @ Oxford

- A convenient way of grouping together a number of element declarations
- These are usually on a related topic or specific application
- Most chapters focus on elements drawn from a single module, which that chapter then defines
- A TEI Schema is created by selecting modules and add/removing elements from them as needed

Modules

Using the basic
TEI structural
elements

TEI @ Oxford

Module name	Chapter
analysis	Simple Analytic Mechanisms
certainty	Certainty and Responsibility
core	Elements Available in All TEI Documents
corpus	Language Corpora
dictionaries	Dictionaries
drama	Performance Texts
figures	Tables, Formulae, and Graphics
gaiji	Representation of Non-standard Characters and Glyphs
header	The TEI Header
iso-fs	Feature Structures
linking	Linking, Segmentation, and Alignment
msdescription	Manuscript Description
namesdates	Names, Dates, People, and Places
nets	Graphs, Networks, and Trees
spoken	Transcriptions of Speech
tagdocs	Documentation Elements
tei	The TEI Infrastructure
textcrit	Critical Apparatus
textstructure	Default Text Structure
transcr	Representation of Primary Sources
verse	Verse

The Imaginary Punch Project

Using the basic
TEI structural
elements

TEI @ Oxford

- **Punch** is a famous English humorous journal, published regularly between 1841 and 1992: see <http://www.punch.co.uk/historyofpunch.html>.
- A project plans to make available fully marked up texts of the journal, in conjunction with page images...
 - for social historians
 - for librarians
 - for linguists
- How will the TEI help? And which parts of the TEI will we use?

Example 1

Using the basic
TEI structural elements

TEI @ Oxford

June 15, 1914.]

PUNCH, OR THE LONDON CHARIVARI.

61

CHARIVARIKA.

Two men carrying bombs were arrested last week on the outskirts of Paris, and are suspected of a plot against the French President. They alleged that the bombs were made for the Tsar of Russia, but the Tsar denies that he gave the commission.

The town of Coochinchina, it is reported, has decided to give up gas in favour of electricity. This of course, is not meant as a slight on its most illustrious resident.

Posted at Ventnor, Isle of Wight, on July 14, was a postcard which had been despatched at the Royal Hotel in Gowas. The recipient is said to have expressed the opinion that it would have been quicker, almost, to have telephoned the message.

Miss Nina Boyle, of the Women's Freedom League, went to the paper to list ladies on whom she considers the KKK ought to bestow honours. Among the ladies there are Mrs. Constance Markievicz and Miss Marie Corelli. It is said to be more of an anti-Suffragette than ever.

"NEW THEATRE FOR LONDON,
ALL SEATS IN THE HOUSE TO BE REMOVED."

So the great difficulty has been solved at last! So may theatres fail because the seats are not taken.

A movement is on foot to induce Mr. CHARLES GARNETT to change the name of his play, *A History of Hate*, as so many people have experienced difficulty in pronouncing the title as it stands at present.

In a struggle between a British sailor and a German policeman at Wilhelmshaven, the latter day bonhomie seems to have been fairly even. The German, who used his sword, lost his hand, and the sailor a piece of his nose.

Two men of good position were tried last week before the State Court of Berlin for refusing to address a police man as "Mr." That will surprise no one who knows his Prussia. It is the sequel which takes the breath away. The two men were acquitted!

shows that in the preceding ten years slugs from the Established Church declined from 25,333 to 24,829. "The decrease is accounted for by the lack of young men taking orders." The word is that such orders were not at once snapped up by fatidic Germans.

Miss LAURA WENTWORTH, of Newbury, has had a very big girl. Miss Anna, who was held, sailed from New York in the *Insperito* with a hat which measures 95 inches in diameter. These giantines are justifying themselves,

We are glad that the POSTMASTER-GENERAL has promised a Bill against foreign spies. Only the other day we received a communication "Schmecksy Schmecksy Schmecksy!"

Whilst we are ready to grant that telephone the message.



ONE ADVANTAGE ABOUT THESE ABSOLUTELY BREEZY COUNTRY COATS IS THAT YOU CAN WEAR OUT SOME OF THE CONFURS IN WHICH YOU WANT THEM TO GO

it is not always easy to find the apt quotation, we cannot help thinking that *The Daily Telegraph* would have caused less offence if it had published the following paragraph without any reference to it.

The Mayor and Mayoress of Kensington, Alderman and Mrs. W. H. Davies, held a reception at their home on Friday evening, their guests numbering between 400 and 500.

Oh, how peaceful is their sleep. They who "keeling's" always keep

a cheerful company at all the Cafes, Soups-Chaises I., anonymous advertisements in *The Manchester Guardian*.

We have heard of lively obsolete before, but the chatty soup must be something of a novelty.

"Strawberries are going out," reports *The Evening News*. We are in a position to confirm that statement. We met one out the other evening.

According to *La France Militaire*

the French Navy is about to try the experiment of enclosing black sailors. We should say that they will be found to make the most admirable stokers, not shoving the dirt like the white men.

Describing a recent visit of a party of Congressmen and State officials to one of the textile buildings in New Haven, Connecticut, a newspaper says: "The distinguished guests took water with what gouts they could." Evidently they thought it scarcely worth saying more for.

The statement made last week in the course of a certain trial that "as a man grows older he becomes riper" has caused some amusement. Others are pointing in from the Cannibal Isles for commentaries of eccentricity.

THE PROFESSIONAL ATTITUDE.

(*The warden girl, according to a daily paper, is not to be wax by love-making. She is to be won over and over again by her beauty and unceasing coquetry.*)

Dear Sir I have a devotion

In the manner knights employed.

With feelings with emotion

(Whish I trust have been destroyed);

Now at last, a practised lover,

Bosting conquests not a

bit.

On my sentiments for you.

Cupid's chat is out of fashion;

Sloppy words are never said;

Whispering words with passion

Shake with merriment instead;

Poets qualified to tackle

Lyric metres when inspired

Sleep to make the ladies cackle—

Doublets on whose occupation

Has a dull and solemn trend

Might enjoy, as relaxation,

Jesting with a female friend;

Be contented by the money

The way writing, the being,

How on earth can I be funny

For the pleasure of the thing?

The Daily Chronicle on the latest

development:

"It will also be equipped with a quick-gun, which disposes when the vessel

This is for the best arrangement; it

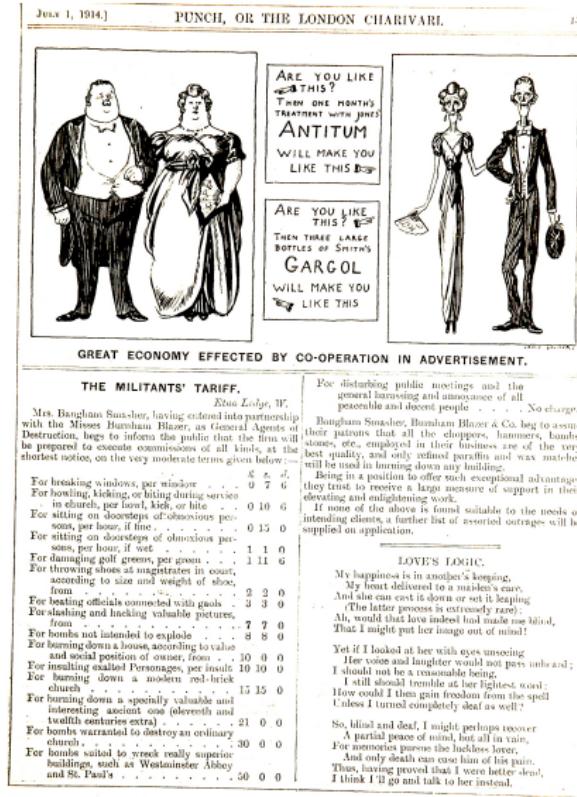
would never do for it to be left floating

where any passer-by could pick it up.

Example 2

Using the basic
TEI structural elements

TEI @ Oxford



THE MILITANTS' TARIFF.

Etta Lodge, W.

Mrs. Bangham Smasher, having entered into partnership with the Misses Burnham Blazer, as General Agents of Destruction, begs to inform the public that the firm will be prepared to execute commissions of all kinds, at the shortest notice, on the very moderate terms given below:

For breaking windows, per window	5 5 d.
For howling, kicking, or biting during service	0 10 6
in church, per foul, kick, or bite	0 10 6
For sitting on dolesteeps of obnoxious persons, per hour	0 15 0
For sitting on dolesteeps of obnoxious persons, per hour, if wet	1 1 0
For darning golf greens, per green	1 11 6
For howling at magistrates in court, according to size and weight of shout, from	2 2 0
For beating officials connected with golf	3 3 0
For smashing and hacking valuable pictures, from	7 2 0
For bombs not intended to explode	3 8 0
For burning down a house, according to value and social position of owner, from	10 0 0
For burning exalted Personages, per insult	10 0 0
For burning down a modest red-brick church	15 15 0
For burning down a specially valuable and indestructible ancient one (eleventh and twelfth centuries), extra	21 0 0
For bombs warranted to destroy an ordinary church	30 0 0
For bombs suited to wreck really superior buildings, such as Westminster Abbey and St. Paul's	50 0 0

For disturbing public meetings and the general harassing and annoyance of all peaceable and decent people No charge.

Bangham Smasher, Burnham Blazer & Co. beg to assure their public that the copper hammers, bombs, stones, etc., employed in their business are of the very best quality, and only refined charcoal and wax matches will be used in burning down any building.

Being in a position to offer such exceptional advantages, they can give a large measure of support in their elevating and enlightening cause.

If none of the above is found suitable to the needs of intending clients, a further list of assorted outrages will be supplied on application.

LOVE'S LOGIC.

My happiness is in another's keeping,

My heart dedicated to a maiden's care,

And yet I can't get it back, nor set it leaping

(The latter process is extremely weary).

Ah, would that love indeed had made me blind,

That I might put her lounge out of mind!

Yet if I looked at her with eyes unseeing

Her voice and laughter would not pass unheeded;

I should not be a reasonable being;

I still should tremble at her slightest word;

How could I then gain freedom from the spell

Unless I turned completely deaf as well?

So blind and deaf, I might perhaps recover

A partial sight and hearing, but all in vain,

For memories pursue the past forever,

And only death can ease him of his pain.

Thus, having proved that I were better dead,

I think I'll go and talk to her instead.

Example 3

Using the basic
TEI structural elements

TEI @ Oxford

JULY 10, 1914.] PUNCH, OR THE LONDON CHARIVARI. 67

THE WORLD'S WORKERS.

Little Girl. "PLEASE, MRS. MURPHY, MUVVER SAYS, IF IT'S FINE TO-MORROW, WILL YOU GO BRIDGES WITH 'EM?"

THE "THORNS OF PRAISE." "HIS PURPLEST SIN." By VANCE BLATHERS (Jack Sheet, 6-3). The Weekly Standard. "... virulent narrative." The Saturday Evening Post. "Replete with up-to-date sentiments . . . knowledge of the heart's trouble . . . very, but never transcending the bounds of decorum." The Buttercup Dispatch. "Passages which seemed to have been written . . . high ideals . . . love interest well sustained . . . careful punctuation." The Nether Walks News. "Mr. Blathers' benefactor . . . reminds me of T. P. O'Connor's Junction-road-painting . . . well-chosen epithets." The Mackintosh Mirror. "Stylish writing . . . Mr. Blathers is evidently a man of great in the most recrue circles." The Chowder Eagle. "Edifying, yet entertaining . . . faultless portraiture, but . . . not in the least like Zola . . . unusually readable." The Criticith Sentinel. "... inside knowledge of Mayfair . . . resolute of humanity at its best . . . fluid and flexible style . . . suitable for a country congregation." The Kilmarnock News. "... cannot remember any book which . . . better than this is." The Publishers' Post. "... redundant with wit . . . The Peabody Advertiser. "Mr. Blathers is a man to be heartily congratulated." The N. Wales Dialect. "... masterly delineation of the smart Set." The Puffin News. "... with a sense of humour." The Birmingham Examiner. "Few books so well worth re- and re-reading." The Popular Courier. "A fine novel." The Slope Spectator. "... marked along the line of . . ." The National Observer. "... mortuous . . ." The Winchester Tribune. "... fond of entertainment. Mr. Blathers next summer . . . awed with impatience." The Isle of Wight Critic. "... clever novel." The Cedar-Idris Atkinson. "... psychology . . . humour . . . passion." The Buckling Post. "... emotional depths . . ." The Sunday Deliverer. "... remarkable book." The Simla Gazette. "... varied . . . professedly enthralling work of fiction." The Grange Trace. "... better than the Common Error." The Cork Pall Mall. "A brilliant first effort." The Huddersfield Examiner. "... well written . . ." The Tooting Express. "... amusing . . ." The Monthly Criterion. "The characters live and move." "Before lunch each session had its annual meeting in private, and at two o'clock the company sat down to a substantial and very welcome meal which was greatly enjoyed by the visitors. After being seated upon by a photographer the party filed . . ." The Daily Guardian. "We were rather afraid they had overdone it."

From a photographic catalogue:—
"This is a most complete little Projector . . . It is quite compact and will make a thirty-inch picture anywhere at a moment's notice." The Buckling Post. "... emotional depths . . ." It should be installed at the Royal Academy without delay.

Looking at Punch, what do we need to mark up?

Using the basic
TEI structural
elements

TEI @ Oxford

- issue information and page number for reference purposes
- "chunks" or divisions of text, which may contain a picture, a poem, some prose, some drama, or a combination
- within the chunks, we can identify formal units such as
 - a picture, a caption
 - stanzas, lines
 - paragraphs
 - speeches and stage-directions
- and more...

TEI tags for the high level structure

Using the basic
TEI structural
elements

TEI @ Oxford

We will treat each issue as a single `<text>` element, and each identifiable chunk within it as a `<div>` element of a particular type (e.g. cartoon, verse, prose)

For example, page 1 has two divisions,

```
<pb n="1"/>
<div type="cartoon">....</div>
<div type="poem">
  <head>Progress</head>....
</div>
```

page 2 also has two, of different types:

```
<pb n="2"/>
<div type="prose">
  <head>The enchanted castle</head>....
</div>
<div type="snippet">
  <head>Correspondence</head>....
</div>
```

Why divisions rather than pages?

Using the basic
TEI structural
elements

TEI @ Oxford

Because a division can start on one page (page 5 for example) and finish on another (page 6)

We use an empty element <pb> to mark the boundary between pages, rather than enclosing each page in a <div type="page">.

```
<pb n="5"/>
<div type="cartoon">...</div>
<div type="review">
  <head>Egypt in Venice</head>...
<pb n="6"/>
...
</div>
<div type="cartoon">...</div>
<div type="verse">
  <head>Enigma</head>...
</div>
<div type="snippets">...</div>
```

The sequence in which divisions appear is rather arbitrary.

Divisions can contain divisions...

Using the basic
TEI structural
elements

TEI @ Oxford

```
<div type="snippets">
  <div type="snippet">Curiously.... Chancellor</div>
  <div type="snippet">Men for the Antarctic... Canadians</div>
</div>
```

- TEI also provides division elements with names that indicate their degree of nesting (<div1>, <div2> etc.) which some people prefer
- Divisions must always tessellate: once "down" a level, you cannot pop "up" again within the same division.

Floating text

Using the basic
TEI structural
elements

TEI @ Oxford

```
<div1>
  <p> ... </p>
  <div2>
    <p> ... </p>
  </div2>
  <div2>
    <p> ... </p>
  </div2>
</div1>
```

is valid **but**

```
<div1>
  <p> ... </p>
  <div2>
    <p> ... </p>
  </div2>
  <p> ... </p>
</div1>
```

is *not valid.*

A special `<floatingText>` element is available for "interruptions"

What are divisions made of?

Using the basic
TEI structural
elements

TEI @ Oxford

(apart from other smaller divisions)

- <head> (heading)
- <p> (paragraph)
- <sp> (speech, contains any of the foregoing, also <stage> and <speaker>)
- <list> (contains <head>, <label>, <item>)
- <table>, (contains <row> containing <cell>) ...
- <l> (verse line) optionally grouped into <lg> (line group)
stanzas
- <figure> (contains <graphic>, <figDesc>, <head>...)

For example....

Using the basic
TEI structural
elements

TEI @ Oxford

Page 3 contains a figure and a dialogue...

```
<div type="cartoon">
  <figure>
    <head>When the ships come home</head>
    <figDesc>A man in Turkish dress lounges on a sofa,
      smoking a cigarette and consulting a book
      labelled "Naval ledger". Another man, in
      traditional Greek costume, stands beside him,
      also reading a notebook.</figDesc>
    <graphic url="Punch/XML/Graphics/003.png"/>
  </figure>
  <sp>
    <speaker>Greece. </speaker>
    <p> Isn't it time we started fighting again?</p>
  </sp>
  <sp>
    <speaker>Turkey. </speaker>
    <p> Yes, I daresay. How soon could you begin?</p>
  </sp>
  <sp>
    <speaker>Greece. </speaker>
    <p> Oh, in a few weeks.</p>
  </sp>
  <sp>
    <speaker>Turkey. </speaker>
    <p> No good for me. Shan't be ready till
      the autumn.</p>
  </sp>
</div>
```

For example...

Using the basic
TEI structural
elements

TEI @ Oxford

The militants' tariff (on Page 15) contains headings, paragraphs, and a table...

```
<div type="prose">
  <head>THE MILITANTS' TARIFF. </head>
  <head rend="right">Etna Lodge, W. </head>
  <p>Mrs. Bangham Smasher, having entered into partnership with the
      Misses Burnham Blazer, as General Agents of Destruction, begs to
      inform the public that the firm will be prepared to execute
      commissions of all kinds, at the shortest notice, on the very
      moderate terms given below: -
  </p>
  <table>
    <row role="label">
      <cell/>
      <cell>f</cell>
      <cell>s. </cell>
      <cell>d. </cell>
    </row>
    <row>
      <cell>For breaking windows, per window ... </cell>
      <cell>0</cell>
      <cell>7</cell>
      <cell>6</cell>
    </row>
    <row>
      <cell>For howling, kicking, or biting during service
          in church, per howl, kick, or bite ... </cell>
      <cell>0</cell>
      <cell>10</cell>
    </row>
  </table>
</div>
```

Global attributes

Using the basic
TEI structural
elements

TEI @ Oxford

Some features (potentially) apply to everything:

- identity
- language
- rendition

TEI provides global attributes for these:

- `@xml:id` provides a unique identifier for any element;
- `@n` provides a name or number for any element
- `@xml:lang` specifies the language of any element, using an ISO standard code
- `@rend` and `@rendition` provide ways of specifying the visual appearance (rendition) of any element

For example...

Using the basic
TEI structural
elements

TEI @ Oxford

Egypt in Venice (on Page 05) begins with two headings, one in French....

```
<div type="prose" xml:lang="en" xml:id="I1914-07-01_05_02">
  <head>Egypt in Venice. </head>
  <head xml:lang="fr" rend="it">"La Légende de Joseph. " </head>
  <p>Those who know the kind of attractions that the
      Russian ballet offers in so many of its themes .... </p>
</div>
```

Each stanza of the poem on page 10 has a last line which is significantly indented:

```
<lg>
  <l>There were eight pretty walkers who went up a hill; </l>
  <l>They were Jessamine, Joseph and Japhet and Jill, </l>
  <l>And Allie and Sally and Tumbledown Bill, </l>
  <l rend="indent">And Farnaby Fullerton Rigby. </l>
</lg>
```

Macrostructure 1

Using the basic
TEI structural
elements

TEI @ Oxford

All the issues of *Punch* for one year make up a volume. We could regard the volume as a single `<text>`, and each issue as a `<div>` within it. Or we could use the `<group>` element:

```
<text xml:id="v147">
  <front>
    <!-- introductory materials for volume 147 here -->
  </front>
  <group>
    <text xml:id="I1914-07-01">
      <body>
        <!-- first issue (1 July) -->
      </body>
    </text>
    <text xml:id="I1914-07-15">
      <body>
        <!-- second issue (15 July) -->
      </body>
    </text>
    <!-- etc... -->
  </group>
  <back>
    <!-- volume index, appendix etc. -->
  </back>
</text>
```

As well as the texts, we have detailed metadata about each volume, and images of its pages. These are the three parts of a canonical TEI document:

```
<TEI>
  <teiHeader>
    <!-- required; provides metadata -->
  </teiHeader>
  <facsimile>
    <!-- the text, represented in image form -->
  </facsimile>
  <text>
    <!-- the text, transcribed and marked up -->
  </text>
</TEI>
```

Macrostructure 3

Using the basic
TEI structural
elements

TEI @ Oxford

If many such documents are grouped together to form a corpus (rather than a collection), it may be useful to factor out the metadata they have in common:

```
<teiCorpus>
  <teiHeader>
  <!-- shared metadata -->
  </teiHeader>
  <TEI>
    <teiHeader>
    <!-- specific metadata -->
    </teiHeader>
    <text>
    <!-- ... -->
    </text>
  </TEI>
  <TEI>
    <teiHeader>
    <!-- specific metadata -->
    </teiHeader>
    <text>
    <!-- ... -->
    </text>
  </TEI>
</teiCorpus>
```

What kinds of metadata?

Using the basic
TEI structural
elements

TEI @ Oxford

For the *Punch Project* and for any other comparable project, we will need a place for such information as

- identification of the resource itself ("what is this thing?")
- statements of responsibility ("who did what when?")
- indication of source ("what was this derived from?")
- publication statement ("how is this item distributed and by whom?")
- declaration of encoding practice ("what do the codes we added mean?")

The TEI Header supports all these, and more.

The TEI Header

The TEI header was designed with two goals in mind

- needs of bibliographers and librarians trying to document 'electronic books'
- needs of text analysts trying to document 'coding practices' within digital resources

On the one hand, the Librarian's header

- uses standard bibliographic concepts
- respects established mappings to other such records (e.g. MARC)
- has a preference for structured data over loose prose

On the other, Everyman's header

- Supports a (potentially) huge range of very miscellaneous information, organized in fairly ad hoc ways -
- Unpredictable combinations of narrowly encoded documentation systems and loose prose descriptions

TEI Header Structure

The TEI header has four main components:

- <fileDesc> (file description) contains a full bibliographic description of an electronic file.
- <encodingDesc> (encoding description) documents the relationship between an electronic text and the source or sources from which it was derived.
- <revisionDesc> (revision description) summarizes the revision history for a file.
- <profileDesc> (text-profile description) provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sublanguages used, the situation in which it was produced, the participants and their setting. (just about everything not covered in the other header elements)

Only <fileDesc> is required; the others are optional.



Simple TEI Header for Punch Project

Using the basic
TEI structural
elements

TEI @ Oxford

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Punch, or the London Charivari, Vol. 147, July 1, 1914</title>
    </titleStmt>
    <publicationStmt>
      <idno type="gutenberg">24357</idno>
      <availability>
        <p>This text is freely available for re-use
           under US and UK law, consult your local
           legal restrictions if elsewhere. </p>
      </availability>
    </publicationStmt>
    <sourceDesc>
      <p>This text is a TEI version of a Project Gutenberg
         text originally located at <ptr
         target="http://www.gutenberg.org dirs/2/4/3/5/24357/">.
         As per their license agreement we have removed all
         references to the PG trademark. </p>
    </sourceDesc>
  </fileDesc>
  <revisionDesc>
    <change when="2008-07-26T23:49:55.968+01:00"/>
  </revisionDesc>
</teiHeader>
```

Below the paragraph...

Within the elements already introduced, TEI offers plenty of scope for mark-up of smaller components. For example:

- boundaries, such as page, column, or line breaks
- highlighting, emphasis and quotation
- editorial changes such as correction, normalization etc.
- names, numbers, dates, addresses...
- links and cross-references
- notes, annotation, indexing
- graphics
- bibliographic citations
- words and other analyses

Highlighting

Using the basic
TEI structural
elements

TEI @ Oxford

By highlighting we mean any combination of typographic features (font, size, hue, etc.) which distinguishes the highlighted text from its surroundings. This may be for many reasons...

- to mark foreign, archaic, technical usages
- for emphasis when spoken
- to show something is not part of the text.. (e.g. cross references, titles, headings)
- or is attributed to some other agency inside or outside the text (e.g. direct speech, quotation)

TEI provides both a generic `<hi>` tag and a large number of specific ones...

A few highlighting examples

Using the basic
TEI structural
elements

TEI @ Oxford

- **<hi>** (highlighted: reason unknown or unimportant)

```
<p>[ The rest of this communication is  
omitted owing to considerations of  
space. <hi rend="sc">Ed</hi>. ]</p>
```

- **<emph>** (emphasized)

```
<said>'E won't bite yer <emph>if you buy 'im</emph> guv'ner. </said>
```

- **<title>** and **<foreign>**:

```
<p>  
  <foreign xml:lang="fr">À propos</foreign> of Oxford, it is a  
question whether that extremely amusing book  
  <title>Verdant Green</title> is still much read by freshers.  
</p>
```

- **<distinct>** (linguistically marked)

```
But then I remind myself  
that the Russian ballet is nothing if not  
<distinct>bizarre</distinct>
```

Quotation

Using the basic
TEI structural
elements

TEI @ Oxford

Quotation marks can similarly be used to set off text for many reasons:

- <q> (used if the reason is unknown or unimportant)
- <said> (speech or thought)
- <quote> (attributed to an external source)
- <mentioned> and <soCalled> (nuances of narrative status)

```
<p>
< said who="#Celia">I know a lovely tin of potted
grouse, </ said > said Celia, and she went off to cut some sandwiches.
</p>
```

```
<head>How to utilise the art of < soCalled > suggestion </ soCalled >
</head>
<head>The Doctor, six down at the turn,
< soCalled > suggests </ soCalled > to his opponent that
they are playing croquet, and wins by two and one. </ head >
```

Quotation (continued)

Using the basic
TEI structural
elements

TEI @ Oxford

Note that these elements can nest within one another:

```
<p>The poet returned to his work. <said>
    <quote>In
        tooth and claw, </quote>
    </said> he muttered to himself,
<said>
    <quote>In tooth and claw. </quote>
</said>
</p>
```

Editorial intervention

Using the basic
TEI structural
elements

TEI @ Oxford

As a simple example, consider: 'Excuse me sir, but would you like to buy a nice little dawg?' on page 6.

We can:

- use `<orig>` to show that "dawg" is what it says, even though this is a nonstandard spelling
- use `<reg>` to show that "dog" is an editorially-supplied regularisation of what it says
- or provide both within a `<choice>` element to say either is a valid encoding:

```
... a nice little
<choice>
  <orig>dawg</orig>
  <reg>dog</reg>
</choice>?
```

Names of persons, places, things...

Using the basic
TEI structural
elements

TEI @ Oxford

- <name> (a name in the text, contains a proper noun or noun phrase)
- <rs> (a general-purpose name or referencing string)
- <title> (any form of title)

The @type attribute is useful for categorizing these, and they both also have @key, @ref, and @nymRef attributes.

Examples of names

Using the basic
TEI structural
elements

TEI @ Oxford

Using `@type` to distinguish personal from geographic names:

```
<p>The scene opens at a party given by
<name type="person">Potiphar</name> in
<name type="place">Venice</name>. </p>
```

Using `@key` and `@ref` to de-reference names:

```
<p>
  <label>Business done. </label>—The Commons
  still harping on the Budget.
  <name
    type="person"
    ref="http://en.wikipedia.org/wiki/Timothy_Michael_Healy">
    Tim Healy</name> enlivened proceedings by vigorous personal attack
  on <q>the most reckless and incapable
  <rs key="LLG">Chancellor of the Exchequer</rs>
    that ever sat on the Treasury Bench. </q>
  <name key="LLG">Lloyd George's</name>
  retort courteous looked forward to with interest.
</p>
```

Dates

Using the basic
TEI structural
elements

TEI @ Oxford

- <date> contains a date and time in any format
- For processing it is convenient to add a normalized version, using the @when attribute
- Uncertain dates and times, and ranges, can be indicated by other attributes: @notBefore, @notAfter, @from@to

```
<p>House of Commons, <date when="1914-06-22">Monday, June 22,  
1914</date>. </p>  
<p>  
  <date notAfter="1914-06-01" notBefore="1914-03-01">Sunday, a month  
ago, <date> was hot.  
</p>
```

Cross references

Using the basic
TEI structural
elements

TEI @ Oxford

A cross reference is a link from one point in a text (the source) to another (the target).

TEI provides generic elements `<ptr>` and `<ref>` for this purpose. If the linking text can be automatically generated use `<ptr>`; otherwise use `<ref>`.

The source is the location of the `<ptr>` or `<ref>`; the target is specified by the `@target` attribute, in the form of a URI reference.

See `<ref target="#Section12">section 12 on page 34</ref>`.

See `<ptr target="#Section12"/>`.

Bibliographic Citations

Using the basic
TEI structural
elements

TEI @ Oxford

TEI provides special elements for bibliographic citations or references:

- `<bibl>` (loosely structured)
- `<biblStruct>` (standard bibliographic structure)
- `<listBibl>` (encloses a bibliography)

These are typically used in preparing bibliographies, or in footnotes. But even in Punch, there are examples.



Simple <bibl> Example

Using the basic
TEI structural
elements

TEI @ Oxford

In Punch, bibliographic citations are usually associated with a quotation from another paper:
The <cit> element groups the two:

```
<cit>
  <quote>It was the time when Henry III. was
    batting with Simon de Montfort and his
    Barons. </quote>
  <bibl>
    <title>Straits Times. </title>
  </bibl>
</cit>
```

Embedded notes

Using the basic
TEI structural
elements

TEI @ Oxford

Notes, whether appearing in the original source, or added by an editor, can be marked using the `<note>` element.
We might use this to add biographical details to the Punch transcriptions:

```
<p>By-the-by, it is denied that  
Sir <name rend="sc">Joseph Beecham</name>  
    <note>Sir Joseph Beecham, 1st Baronet  
        (8 June 1848 - 23 October 1916)...</note>.  
    was in any way responsible for the Government's  
    "Pills for Earthquakes," by which it was hoped to  
    avert the Irish crisis.</p>
```

`<note>` has attributes `@place` and `@resp`

Linked notes

Using the basic
TEI structural
elements

TEI @ Oxford

Since we have several references to the same person, it might be better to put the notes elsewhere and point to them from the names:

```
<div type="notes">
  <note xml:id="BEECHJ0">Sir Joseph Beecham, 1st Baronet (8 June 1848 -
    23 October 1916) the eldest son of Thomas Beecham (1820-1907) played a
    large part in the growth and expansion of his father's medicinal pill
    business which he joined in 1866....</note>
  <!-- other notes -->
</div>
<div type="snippets">
  <p>... Both Earl <name rend="sc">Beauchamp</name>
    and <name>Sir <ref target="#BEECHJ0">Joseph Beecham</ref>
    </name> appear
    in the recent Honours List.</p>
  <p>By-the-by, it is denied that Sir
    <name rend="sc" ref="#BEECHJ0">Joseph
    Beecham</name> was in any way responsible...</p>
</div>
```

Could also use specialised `<person>` element, in this case.
"Elsewhere" can be *anywhere* on the Internet...

Using TEI in a Philological Context: Reports from work in progress

1. Producing a critical edition
2. Encoding a dictionary
- [3. Compiling translation corpora]



DC 2008 Berlin

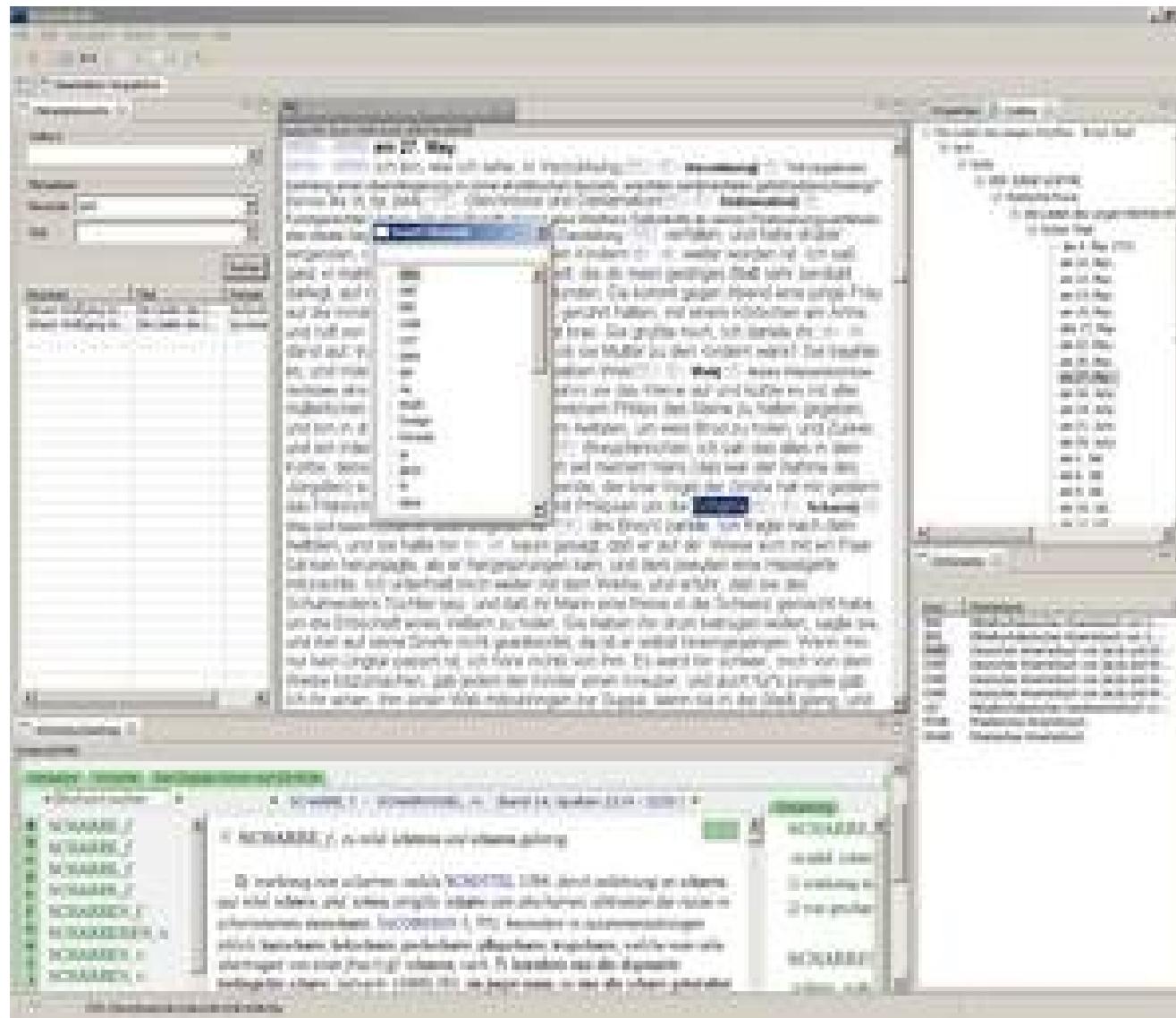


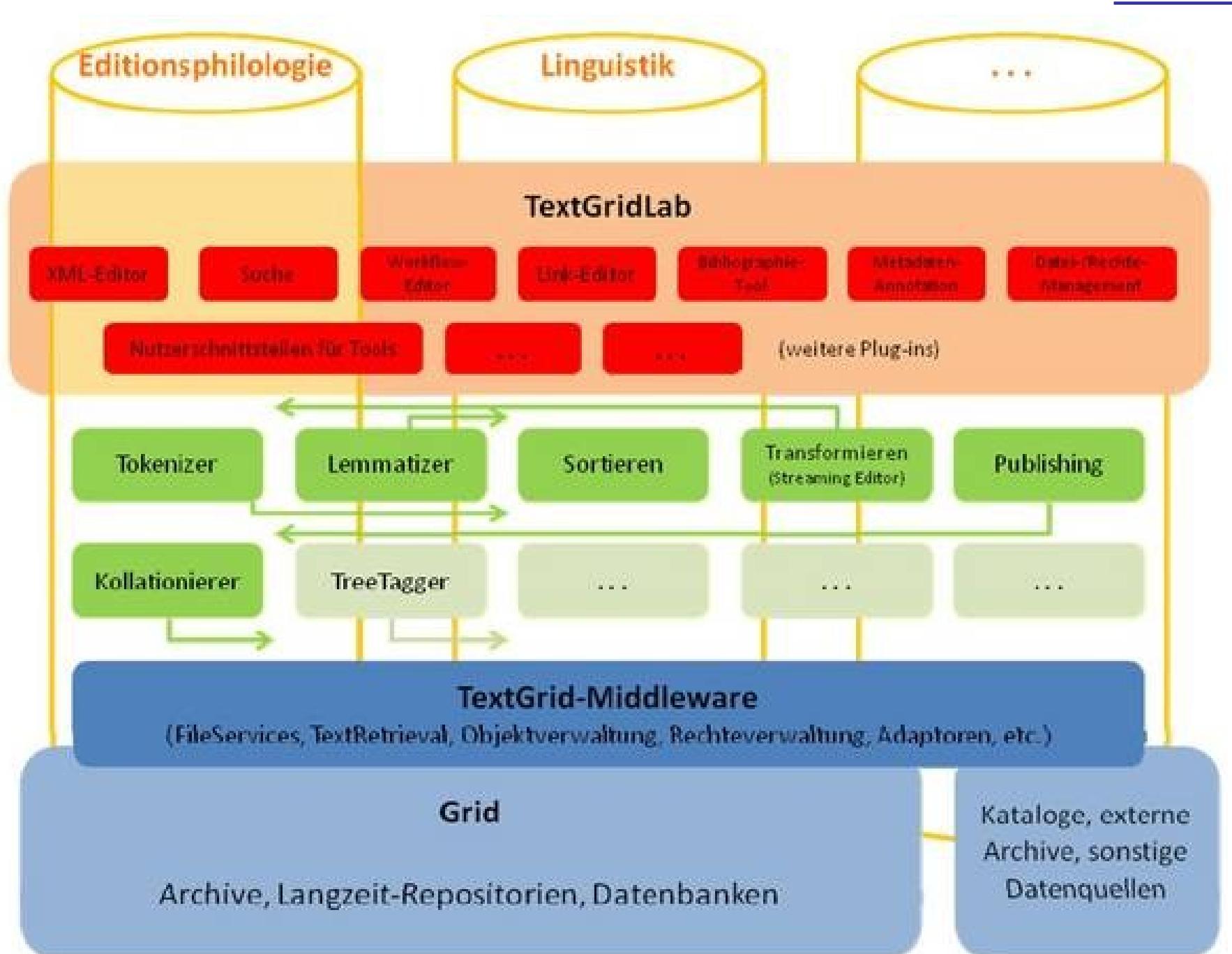
Bundesministerium
für Bildung
und Forschung

The TextGrid project

1. Project
2. Motivation
3. Collaborative design
4. TextGridLab: user toolbox
5. TextGridRep: archiving data

TextGrid: Toolbox





Using TEI in a Philological Context: Reports from work in progress

1. Producing a critical edition
2. Encoding a dictionary
- [3. Compiling translation corpora]



DC 2008 Berlin



Bundesministerium
für Bildung
und Forschung

Producing a critical edition

Jean Paul Arbeitsstelle, University of Würzburg:
The new critical edition of Jean Paul's works
Volume 1: Hesperus, edited by Barbara Hunfeld

[www.jean-paul-portal.de]

Producing a critical edition

Jean Paul edition: texts published by the author
Hesperus versions 1795 - 1798 - 1819
[altogether more than 20,000 pages]

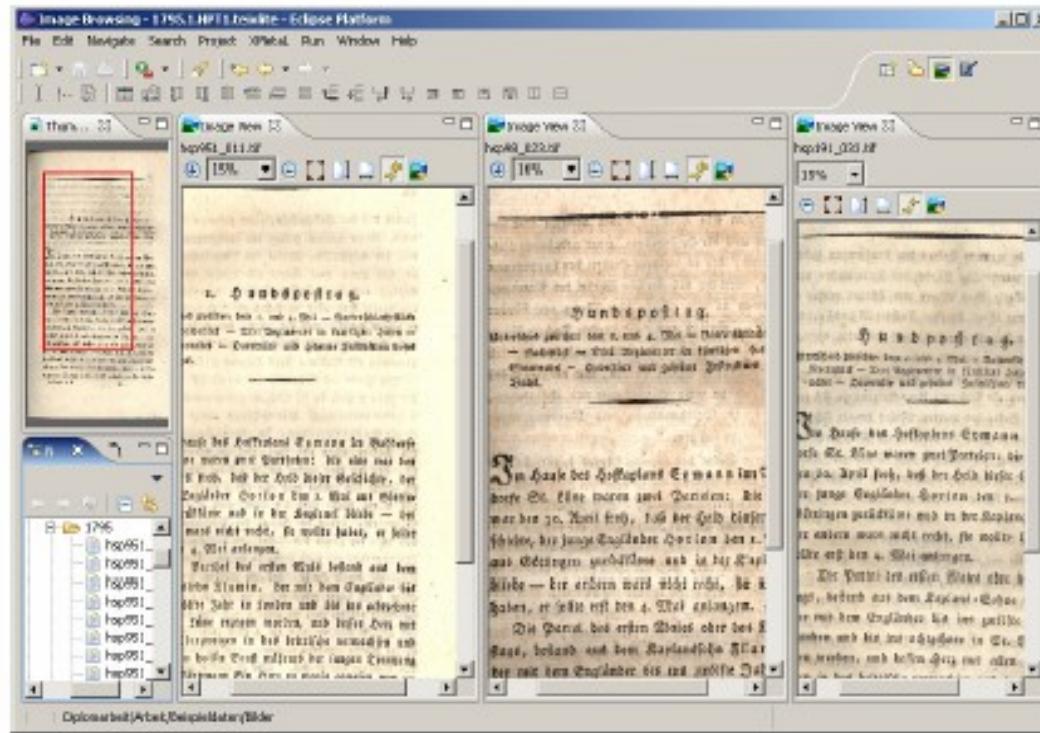
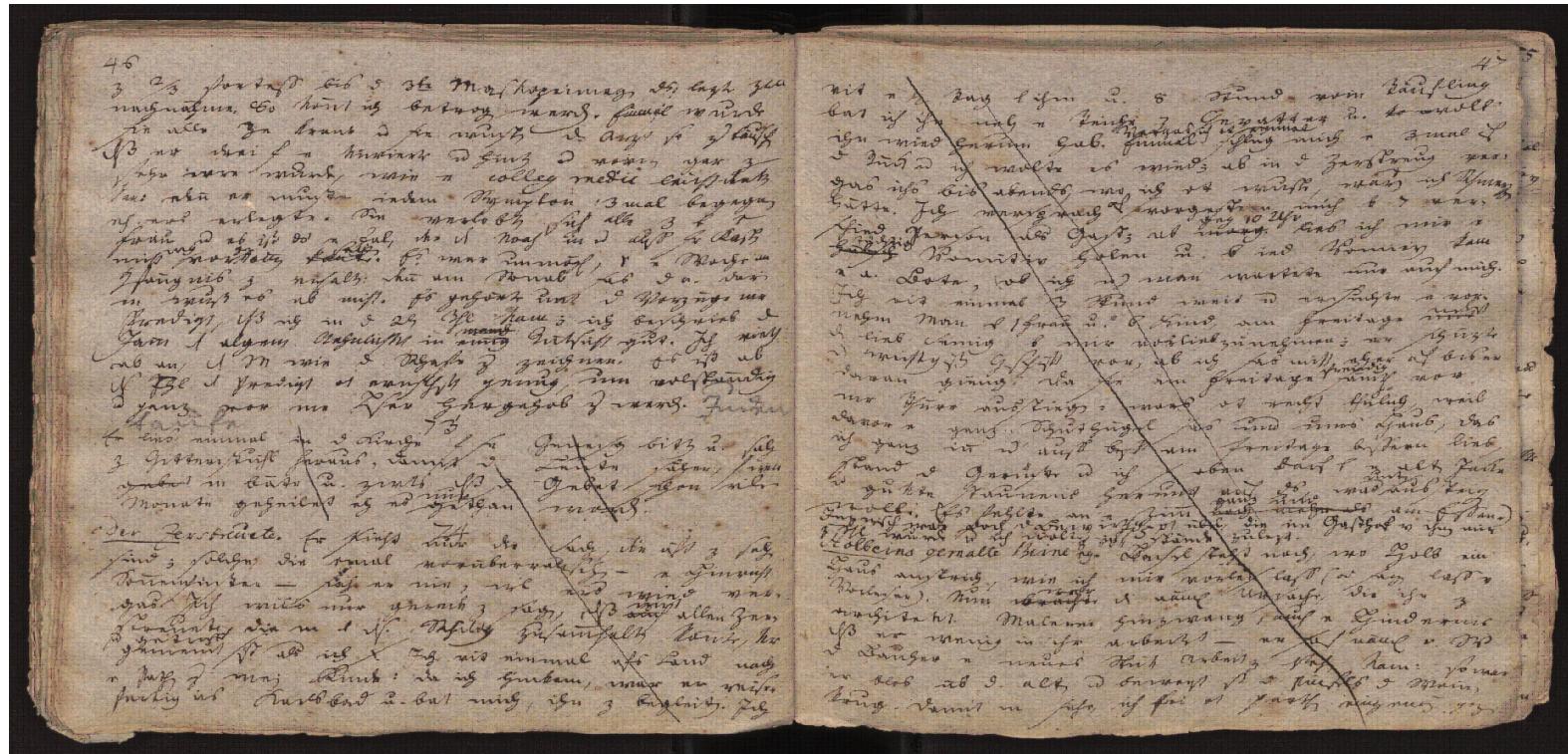


Abb. 12: Perspektive zum Vergleichen von Bildern

Producing a critical edition

Jean Paul, Nachlass, pages used for 'Hesperus'
Staatsbibliothek zu Berlin - Preußischer Kulturbesitz



Producing a critical edition

Jean Paul, Nachlass (Detail: "Holbeins gemalte Beine")
Staatsbibliothek zu Berlin - Preußischer Kulturbesitz

The image shows a single sheet of aged, yellowish-brown paper with handwritten text in black ink. The handwriting is cursive and somewhat faded. A large, hand-drawn 'X' is drawn across the entire page, indicating that the content is not to be reproduced. The text discusses Holbein's painted legs and a visit to Berlin.

Producing a critical edition

Jean Paul, Nachlass, Detail: "Holbeins gemalte Beine"
Staatsbibliothek zu Berlin - Preußischer Kulturbesitz

ganz in w^o auf der M^unchen Freitags bei Paar sind
Kunst & Grafschaft d^o ist eben dann f^ü r g^olt f^ü r
guten F^ürstenkunst gemaing u^ol^o g^olt n^och ein
Möhl. Es fehlt am zim^o g^olt n^och nicht am F^ür
F^ürst^og^olt man dor^o kann i^org^o o^o n^och d^o im G^osthof n^og^olt mit
M^unchen w^o es w^ohlt i^org^o o^o n^och z^ulezt.
Holbeins gemalte Beine pg. D^orf l^oft nach wo Gold in
g^olt a^ork^o, min i^ol^o mit m^orl^o l^oft s^oz^o e^oft,
Berlin). Min m^orl^o d^o n^och M^unchen d^o i^o 3

Producing a critical edition

Test of typesetting, using TEI encoding & TeX

<p>4 Hundespottig Schattenkl-Schneider – Klotildens historische Figur – einige Hoffeste und ein erhabener Mensch.–</p> <p>Eigentlich wollte Klotilde – erfuhr Sebastian am Morgen – bis Johannis im Stoffe blieben; aber da ihre beste Freundin Orlaia vor uns fortgegangen war, nicht zu ihren Eltern, sondern unter die Erde, so mußte sie das verwundete Auge durch eine schnellere Abreise wegziehen von dem Grabestiegel, der wie ein Ruis über den verlorenen Herzen ruhte.</p> <p>Nie wurde eine große Schönheit von einer kleinen unbefangner gelobt als von Agnethen Klotilde. Sonst schützen Mädchen am Mädchen nur das Herz; die zerstörenden Reize eines fremden Gesichts haben so wenig Werth in ihren Augen, daß sie ihrer kaum erwähnen mögen. Jünglinge wirft man richtig vor, daß sie gern schöne Jünglinge zu ihren Freunden auslesen, bei Männern hingegen wollen ihre Lehrmeister viel daraus machen, daß sie die wohliche Schönheit als einen zu lockern und niedrigsten Mordel und Leim der Freundschaft gänzlich verschwinden, und daß daher einer schönen Frau das Herz der allerheiligsten Theurer sey als das Gesicht der Schönsten auf den fünf Erdgärten und Erdscherpen. Agathe war anders; sie ließ schon am Morgen am Schloß, um ihre Freunde anzukleiden.</p>	<p>4 Hundespottig Schattenkl-Schneider – Klotildens historische Figur – einige Hoffeste und ein erhabener Mensch.–</p> <p>Eigentlich wollte Klotilde – erfuhr Sebastian am Morgen – bis nach Johannis im Stoffe bleiben; aber da ihre beste Freundin und Stift-Gemalin Orlaia in uns fortgegangen war, nicht zu ihren Eltern, sondern unter die Erde, so mußte sie das verwundete Auge durch eine schnellere Abreise wegziehen von dem Grabestiegel, der wie ein Ruis über dem verlorenen Herzen ruhte. Ohne Glück war sie dem blumenlosen Golgotha ihrer verwundeten Seele entflohen, und ihr stand noch ein zweiter Anblick daseihen, eine zweite Abreise und die Wiederholung der alten Thränen hervor.</p> <p>Nie wurde eine große Schönheit von einer kleinen unbefangner gelobt als von Agnethen Klotilde. Sonst schützen Mädchen am Mädchen nur das Herz, die zerstörenden Reize eines fremden Gesichts haben so wenig Werth in ihren Augen, daß sie ihrer kaum erwähnen mögen. Jünglinge wirft man richtig vor, daß sie gern schöne Jünglinge zu ihren Freunden auslesen, bei Männern hingegen wollen ihre Lehrmeister viel daraus machen, daß sie die wohliche Schönheit als einen zu lockern und niedrigsten Mordel und Leim der Freundschaft gänzlich verschwinden, und daß daher einer schönen Frau das Herz der allerheiligsten Theurer sey als das Gesicht der Schönsten auf den fünf Erdgärten und Erdscherpen. Agathe war anders; sie ließ schon am Morgen ins Schloß, um ihre Freunde anzukleiden.</p> <p>Plannin macht' es noch anger, er kommt' es nicht erwarten, daß die Wirklichkeit selber Klotildens Madonnenbild in Väikes Gehirnkammern aufging, er kann für mit der Federzeichnung eines Malers zuvor, die wenigstens nicht – kalt ist, denn Maler schreiben im <i>ästhetischen</i> und im <i>kalligraphischen</i> Streichelein.</p> <p>Der Maler hatte, blos um Klotilden zu sehen und zu zeichnen, fast alle Sonntags-Morgen auf einem Berg von Maienthal gelegen, wo er die glänzende Landschaft um das Stoffe auf seine Blätter trug, und den schönen Kopf, der aus dem seihen Fenster herausaß in sein Herz. Sogar Plannin, der sonst die prosaischen Projektionen und Buchdruckstecke über die lebenden Oehlgemälde der Dichtkunst stellte, fand an der folgenden Madonna oder Klotilde des Malers Geschmack:</p> <p>>Wenn mein Ich ein einziger Oeskele ist und brennt, und wenn ich, von Flammen umwelt, die Hand in Farben tauche, um mich darin abzutühlen – wenn dann die hohe Schönheit,¹ die ewig, in mir strahlt, ihr Spiegelbild auf die Wellen, die Himmel und Erde zintend malen, herunterfallen läset, und den klaren Sturm entflammt, wenn alsdann an dem Himmel entzunkenes Palash auf dem Strome ruht, eine Lilienhalle, eines aufgeflogten Engels weggelegte Flügeldecke – eine Gestalt, deren unbefleckte Seele kein Leib, sondern der Schnee umwaltet, der um den Thron Gottes liegt, und aus dem die Engel ihre flüchtigen Resenkörper² bauen – und wenn die zintete Bekleidung zu groß und hart, und ein hölzerner Rahmen um diesen geistigen Hauch auf dem</p> <p>¹ Das Ideal des Schönen ² Wie die Rahmen nach Einschätzung Justinius P. II. 7. glauben</p>
---	---

Encoding a dictionary

-
-
- Joachim Heinrich Campe:
 - „Wörterbuch der deutschen Sprache“, 5 volumes, Braunschweig 1807 – 1811
 - + „Wörterbuch zur Erklärung und Verdeutschung der unserer Sprache aufgedrungenen fremden Ausdrücke. Ein Ergänzungsband zu Adelungs und Campes Wörterbüchern“, Braunschweig 1813
- 6000 pages, about 140.000 entries
- [Testbed for the TextGrid Project]

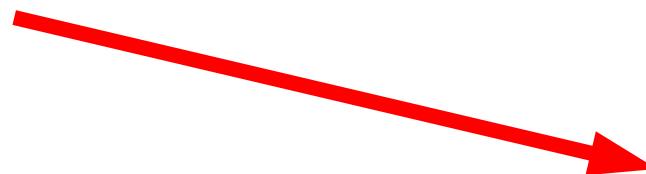
Encoding a dictionary

Workflow

Image digitization
resolution
dynamic range
Type of scanner
image format
quality control
metadata
archiving

Typographical analysis

Text capture
page format
line structure
font information
font size
indentation
special characters
quality control



Encoding a dictionary

Workflow

Typographical analysis: page

Nal	2	Nb	
ser gestellt wird, woran sich Tiere und Weißfische seien und womit sie herausgezogen werden.		der man die höchste Beachtung bezeigten und die man mit einer niedigen Bezeichnung belegen will.	
Die Kaltaupe, <i>Mg.</i> die —n, ein in füßen Wasser lebender Fisch, der eine glatte Haut wie der Zut, aber einen dicken Kopf und Bauch hat, (<i>Gadus lota L.</i>). Zu <i>R. D.</i> die Kaltaupe, in <i>D. D.</i> der Raubaal, der Ruppe, Nutte, Nutten, Trutsch, Trutsch und Frösche.		Die Käslätter, <i>Mg.</i> die —n, eine blättrige Blätter, die schwarz und flinkend ist; die Kaspole.	
Die Kaltente, <i>Mg.</i> die —n, oder der Meerast, eine Kaltart, grüher als der gemeine Zut, mit achtgrauen Rücken, weißem Bauch, weißlich getupften Seitenlinien und linsenförmigen Ringen um die großen Augen. Er hat zwei kurze Bartlappen und die Flossen sind mit kleinen spitzen Zähnen besetzt (<i>Muraena conger L.</i>).		Kasen, v. intrs. Mitteln, der vergangenen Zeit, geasest. <i>R. D.</i> auf eine schwülige und ekelhafte Weise in etwas herumwühlen, und überkaupt eine schwüle Arbeit verrichten. Im Holzmeisterin auch, unerheblich und verschwendert mit etwas zu Werk gehen. Die Leute asen im fetten Graue, sagt man, wenn sie mehrere getreten, als freuen. Das Kasen.	
Die Käsflege, <i>Mg.</i> die —n, eine Art großer Fliegen, die durch das Kas angelockt wird, (<i>Musca cadaverina L.</i>); die Fleischfliege, Schneiffliege.		Käsflege, <i>Mg.</i> die —n, eine Art großer Fliegen, die durch das Kas angelockt wird, von Thieren gefressen.	
+ Der Käsfadel, des —s, <i>Mg.</i> die —n, f. Käsfabel.		Käsfäsig, adj. u. adv. was gern Kas frisst, von Thieren gebräuchlich. <i>S. Käsfessla.</i>	
Der Käsfeder, des —s, <i>Mg.</i> m. d. Gz. + So viel als Kargaschel, s. b. 2) Eine der damals Käfe fücht über finanziert.		Käsfestend, adj. was sich von Käfe a) nähet.	
Der Käsfkreif, des —s, <i>Mg.</i> die —n, oder der Käsfstreifen, des —s, <i>Mg.</i> m. d. Gz. der schwane Streif läng dem Rücken des Käfe; unerheblich, ein schwarzer Streif, welchen besonders sehr Viebe auf dem Rücken tragen, der Käfstrief.		Käsfestig, adj. u. adv. was nicht als füchtig, sondern auch faul es gleich (<i>Kas</i>) frisst, von Thieren gebräuchlich.	
Der Käftrich, des —s, <i>Mg.</i> die —n, f. Käftrich.		Der Käfgeier, des —s, <i>Mg.</i> w. d. Gz., eine kleine Art Geier von der Größe eines Raben, die am Kopfe und Brust nur wenige weiße Dünken haben, wo sie sich in großen Scharen versammeln (<i>Vultur percopterus L.</i>); der Gedreier.	
Der Käftrich, des —s, <i>Mg.</i> die —n, f. Käftrich.		Der Käfgeier, des —s, <i>Mg.</i> die —geräthe, der unangenehme Geruch des Käfes; dann, jeder unangenehme Geruch.	
Der Käftrich, des —s, <i>Mg.</i> die —n, eine Art Vogel von Pfästen in einem Stufe, vor welches man Rege, Käsen ic. stellt, um Käfe uns befürchte Käfe zu fangen.		Der Käfgeist, des —s, <i>Mg.</i> ohne M. Gestank, den das Kas verleiht; dann, ein jeder demselben ähnlicher, überkaupt sehr unangenehme, unausstehliche Geruch.	
Das Käftherchen, des —s; b. <i>Mg.</i> w. d. Gz., bei einigen Käfussothen, das kleine ausfremde Thierchen, welches sich in Eßs, Käftrich, Sauerkrat, erzeugt.		Käfgerig, adj. u. adv. was nach Kas bestegir.	
Das Kälmate, <i>Mg.</i> die —n, ein gewisses Kas zum Käfanganzen.		Das Kägrube, <i>Mg.</i> die —n, eine Grube, in welche das Kas geworfen wird; auch f. Käfslühe.	
Das Kälwär, des —s, <i>Mg.</i> die —n, eine Art Vogel von Pfästen in einem Stufe, vor welches man Rege, Käsen ic. stellt, um Käfe uns befürchte Käfe zu fangen.		Käfslast, adj. u. adv. den Kas ähnlich, die Eigenschaft des Kas haben (<i>cadaverös</i>). Der Käfslaster Geruch. Das edere, von Menschen gebrauchte Wort ist leichenhaft oder leichenartig.	
* Der Kar, des —s, <i>Mg.</i> die —n, oder —m, die alte Benennung alter großer Raubbaul, besonders aber des Adlers, die noch in <i>R. D.</i> holt ist und bei Dichtern vorkommt.		Käfslast, adj. u. adv. 1) So viel als anfaßt. 2) f. Unentiglich. (S. <i>Aas</i>). 3) Häuflich, schwülig. Kasig Wetter. In Preussen auch für kraft, matt, träge, faul.	
Ein häufner Kar heißt mit gewaltigen Schwungen		Der Käfsläfer, des —s, <i>Mg.</i> w. d. Gz., eine Art Käfer, welche auf und das Käfe leben, (<i>Silphia L.</i>).	
Die Käfte, — — — — — Schreider.		Der Käfslkopf, des —s, <i>Mg.</i> die —köpf, eine gewisse Verzierung in der Haupftunk, welche dem von der Haut entblößten Kopfe eines Thieres gleicht.	
Bald wird ihe im Meer der Käfe, am Gesäß		Die Käfslkäfe, <i>Mg.</i> die —n, eine Art Krähen (<i>Corvus corone L.</i>), die an andern Orten Nebelschläge und Nebenkäfe heißt.	
Der Kären, Weite sein. — Ramler.		Die Käfsloste, <i>Mg.</i> die —n, die innwendige Seite des Felle, edler die Fleischfalte.	
Der Aaron, eine Pflanze, f. Krum.		Der Käfslvogel, des —s, <i>Mg.</i> die —vögel, in allgemeinen jeder Vogel, der sich von Käfe nährt.	
Der Ararweile, des —s, v. der Ararweile, <i>Mg.</i> die —n, <i>R. D.</i> eine dem Kar oder Adler ähnliche Art Weibens.		Ab, adv. u. prap. 1) Als Umsatzwort bezeichnet ab eine Trennung, Absonderung, und überkaupt eine Entfernung. Der Detek ist ab. Drei Schritte vom Wegab, Gut ab! Auffund ab, ab und an, ab für uns gefähr, bald mehr bald weniger. Der Preis ist verschieden, einen Thaler auf und ab, auch bloß ab in Rechnungen und Schauspielen, wo es bedeutedet, ziehe ab, tritt ab. Gewöhnlich wird es von einer Bewegung von der in Rede stehenden Sache hinweg, oder von einer Bewegung nach unten	
— — — — — und sond mis — — — — —		Ab, v. rec. 1) Als Umsatzwort bezeichnet ab eine Trennung, Absonderung, und überkaupt eine Entfernung. Der Detek ist ab. Drei Schritte vom Wegab, Gut ab! Auffund ab, ab und an, ab für uns gefähr, bald mehr bald weniger. Der Preis ist verschieden, einen Thaler auf und ab, auch bloß ab in Rechnungen und Schauspielen, wo es bedeutedet, ziehe ab, tritt ab. Gewöhnlich wird es von einer Bewegung von der in Rede stehenden Sache hinweg, oder von einer Bewegung nach unten	
unter zerfummneten Stern		Ab, v. rec. 2) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
Miner Steigtnossen wieder. Schabart.		Ab, v. rec. 3) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
Und wählt die Gläudnoss fort. Voss.		Ab, v. rec. 4) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
Besonders und gewöhnlich aber wird es von toden, in Südlins übergehend und dann flinkenden Körper der Thiere gebraucht. Hieraus entstand 2) die uneigentliche Bedeutung von Kar, nach welcher es überkaupt etwas Stolnzliches, Schwüliges bedeutet. Da sie wird in einigen Gegenden jedes Geschwür Kas genannt, und in der Hobelsprache jede Person, besonders weiblichen Geschlechts,		Ab, v. rec. 5) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 6) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 7) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 8) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 9) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 10) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 11) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 12) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 13) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 14) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 15) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 16) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 17) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 18) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 19) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 20) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 21) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 22) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 23) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 24) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 25) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 26) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 27) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 28) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 29) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 30) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 31) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 32) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 33) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 34) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 35) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 36) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 37) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 38) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 39) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 40) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 41) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 42) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 43) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 44) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 45) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 46) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 47) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 48) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 49) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 50) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 51) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 52) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 53) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 54) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 55) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 56) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 57) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 58) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 59) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 60) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 61) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 62) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 63) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 64) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 65) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 66) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 67) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 68) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 69) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 70) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 71) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 72) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 73) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 74) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 75) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 76) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 77) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 78) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 79) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 80) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 81) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 82) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 83) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 84) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 85) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 86) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 87) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 88) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 89) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 90) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 91) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 92) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 93) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 94) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 95) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 96) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 97) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 98) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 99) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 100) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 101) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 102) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 103) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 104) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 105) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 106) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 107) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 108) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 109) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 110) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 111) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 112) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 113) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 114) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 115) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 116) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 117) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 118) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 119) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 120) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 121) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 122) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 123) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 124) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 125) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 126) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 127) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 128) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 129) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 130) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 131) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 132) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 133) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 134) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 135) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 136) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 137) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 138) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 139) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 140) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 141) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 142) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 143) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 144) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 145) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 146) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 147) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 148) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 149) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 150) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 151) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 152) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 153) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 154) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 155) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 156) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 157) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 158) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 159) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 160) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 161) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 162) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 163) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 164) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 165) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 166) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 167) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 168) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 169) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 170) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 171) Unter zerfummneten Sternen, unter zerfummneten Sternen.	
		Ab, v. rec. 172) Unter zerf	

Encoding a dictionary

Die **Ualschlange**, Mz. die —en, oder der **Meeraal**, eine **Ualart**, grösser als der **gemeine Ual**, mit aschgrauem Rücken, weißem Bauch, weißlich getupfeten Seitenlinien und silberfarbigen Ringen um die großen Augen. Er hat zwei kurze Bartfaden und die Kiefern sind mit kleinen spitzigen Zähnen besetzt (*Muraena conger* L.).

† **Der Ualstachel**, des —s, Mz. die —n, s. **Ualgabel**.

Der Ualstecher, des —s, Mz. w. d. Ez. 1) + So viel als **Ualgabel**, s. d. 2) Einer der damit **Uale** sticht oder fängt.

Der Ualstreif, des —es, Mz. die —e, oder **der Ualstreifen**, des —s, d. Mz. w. d. Ez. der schwarze Streif längs dem Rücken des **Uales**; uneigentlich, ein schwarzer Streif, welchen besonders fahle Pferde auf dem Rückgrath haben; **der Ualstrich**.

Der Ualstrich, des —es, Mz. die —e, s. **Ualstreif**.

Der Ualteich, des —es, Mz. die —e, ein Teich, in welchem **Uale** gehalten werden.

Das Ualthierchen, des —s, d. Mz. w. d. Ez., bei einigen Naturforschern, das kleine aalförmige Thierchen, welches sich in Essig, Kleister, Sauerteige &c. erzeugt.

† **Die Ualwate**, Mz. die —n, ein gewisses Netz zum **Ualfangen**.

Das Ualwehr, des —es, Mz. die —e, eine Art Wehr von Pfählen in einem Flusse, vor welches man Netze, Reusen &c. stellt, um Fische und besonders **Uale** zu fangen.

Der Ualwels, des —es, Mz. die —e, s. **Buntaal**.

Der Ualwurm, des —es, Mz. die —würmer, überhaupt ein an Gestalt dem **Uale** ähnlicher Wurm, verglichen die Essig- und Kleisterälchen sind.

* **Der Uar**, des —es, oder —en, Mz. die —e, oder —en, die alte Benennung aller grosser Raubvögel, besonders aber des **Adlers**, die noch in N. D. üblich ist und bei Dichtern vorkommt.

Ein kühner Uar theilt mit gewalt'gen Schwingen

Die Lüfte, — — — — — Schreiber.

Bald werdet ihr im Meer der Haien, am Gestade

Der Uaren Beute sein. — Rambler.

Der Uaron, eine Pflanze, s. **Arum**.

Der Uarweihe, des —n, oder die **Uarweihe**, Mz. die —n, N. D. eine dem **Uar** oder **Adler** ähnliche Art Weihe.

Workflow
Typographical
analysis:
line & fonts

Encoding a dictionary

Workflow

Guidelines for typing

BUCHSTABENMUSTER: Wörterbuch der deutschen Sprache & → ¶
..... → Wörterbuch zur Erklärung der unserer Sprache aufgedrungenen fremden Ausdrücken ¶

¶ UNICODE-Zeichen (soweit belegt) ¶

komplexere Zeichen in der Codierung wie angegeben (diakritische Zeichen in Anlehnung an TUSTEP) ¶

¶ Mustertranskription gilt nur für den jeweils rot markierten Buchstaben ¶

FRAKTUR	○	ANTIQUA	ANTIQUA
Großbuchstabe	Kleinbuchstabe	Großbuchstabe	Kleinbuchstabe
○	○	○	○
<u>Ansage</u> ,	<u>Fackeln</u> ,	<u>Adverbium</u>	<u>adj. u. adv.</u>
A.....(Ansage)	a.....(Fackeln)	A.....(Adverbium)	a.....(adj. u. adv.)
¶	überdáchen	Álbum	Admiráble
Ao	á.....(Überdáchen)	A.....(Álbum)	á.....(Admiráble)
¶	¶	Áccidenz	Admirál,
○	○	A....(Áccidenz)	á.....(Admirál)
¶	¶		âo
Ürmel	Baumsäge	¶	Pädagög
Á.....(Ürmel)	ää....(Baumsäge)	Ao	ää....(Pädagög)

Encoding a dictionary

Workflow

Guidelines for typing

Kürzelzeichen für Verwendungsangaben (nur im Campe: Deutsches Wörterbuch) mit ihren Unicode-Werten (hex.): ¶

¶	○ Der Felsenfall ☞ ¶25CB; ☞	Kreis ☞ ○ ☞	◎ Der Dichterquell ☞ ¶2609; ☞	Kreis-im-Kreis ☞ □·9737··SUN, miscellaneous symbols ☞
¶	△ Das Felsenherz ☞ ¶25B3; ☞	leeres-Dreieck ☞ △·9651··WHITE-UP-POINTING-TRIANGLE, geometric-shapes ☞	△ Die Felsenbrüst ☞ ¶25EC; ☞	Dreieck-mit-Innenkreis ☞ □·9708··WHITE-UP-POINTING-TRIANGLE-WITH-DOT, geometric-shapes ☞
¶	+ Die Überart ☞ ¶03EE; ☞	Kreuz ☞ +·1006··COPTIC-CAPITAL-LETTER-DEL, greek-and-coptic ☞	‡ Der Walz ☞ (DK) ☞	Doppelkreuz ☞ ‡ ☞
¶	✗ Evangelisch ☞ {DHK} ☞	Doppelhalbkreis,- seitlich-offen ☞ ✗ ☞	✗ Der Beterjunge ☞ {DHKG} ☞	Doppelhalbkreis,- durchgestrichen ☞ ✗ ☞
¶	* Wältigen ☞ * ☞	Stern ☞ * ☞	** Wältig ☞ ** ☞	zwei-Sterne ☞ ** ☞
¶	¤ Überallmächtig ☞ ¶2641; ☞	umgekehrtes-Venuszeichen ☞ ¤·9793··EARTH ☞		

Encoding a dictionary

Workflow Text capture: result

\$0001.107 ^@^@0^@<filename:cwds1 c0002.jpg/>

@^@1^@Aal<S2>Ab

\$0002.01 <duolan 2><-P>ser gestellt wird, woran sich Aale und Weißfische setzen und womit

\$0002.02 sie herausgezogen werden.</negEZ>

\$0002.03 <negEZ>† <W+2>Die Aalraupe</W+2>, <W+1>Mz. die -- n</W+1>, ein in süßem Wasser lebender Fisch,

\$0002.04 der eine glatte Haut wie der Aal, aber einen dicken Kopf und

\$0002.05 Bauch hat, (<A>Gadus Iota L.). Jn N. D. die <W+1>Aalquappe</W+1>, in

\$0002.06 O. D. <W+1>der Raubaal, der Ruppe, Rutte, Rutten, Trusch, Trusche</W+1>

\$0002.07 und <W+1>Trüsche</W+1>.</negEZ>

\$0002.08 <P><W+2>Die Aalreuse</W+2>, <W+1>Mz. die -- n</W+1>, eine Reuse zum Aalfangen.</P>

\$0002.09 <negEZ><W+2>Die Aalschlange</W+2>, <W+1>Mz. die -- en</W+1>, oder <W+1>der Meeraal</W+1>, eine Aalart,

\$0002.10 größer als der gemeine Aal, mit aschgrauem Rücken, weißem Bau=

\$0002.11 che, weißlich getüpfelten Seitenlinien und silberfarbigen Ringen um

\$0002.12 die großen Augen. Er hat zwei kurze Bartfaden und die Kiefern

\$0002.13 sind mit kleinen spitzigen Zähnen besetzt (<A>Muraena conger L.).</negEZ>

Encoding a dictionary

Workflow Text capture: result 2

\$0002.35 <negEZ>* <W+2>Der Aar</W+2>, <W+1>des -- es</W+1>, oder -- <W+1>en, Mz. die -- e</W+1>, oder -- <W+1>en</W+1>, die alte

\$0002.36 Benennung aller großer Raubvögel, besonders aber des Adlers, die

\$0002.37 noch in N. D. üblich ist und bei Dichtern vorkommt.</negEZ>

\$0002.38 <shi>Ein kühner <W+1>Aar</W+1> theilt mit gewalt'gen Schwingen

\$0002.39 Die Lüfte, -- - - - - - ^#s+Schreiber^#s-.

\$0002.40 Bald werdet ihr im Meer der Haien, am Gestade

\$0002.41 Der <W+1>Aaren</W+1> Beute sein. -- ^#s+Ramlar^#s-.

</shi>

\$0002.42 <P><W+2>Der Aaron</W+2>, eine Pflanze,
s. <W+1>Arum</W+1>. </P>

Encoding a dictionary

Workflow

Lexicographical analysis

Entry structures

Example: Der Aar

* Der Aar, des —es, oder —en, Mz. die —e, oder —en, die alte Benennung aller großer Raubvögel, besonders aber des Adlers, die noch in N. D. üblich ist und bei Dichtern vorkommt.

Ein kühner Aar theilt mit gewalt'gen Schwingen
Die Lüfte, — — — — — Schreiber.
Bald werdet ihr im Meer der Haien, am Gestade
Der Aaren Beute sein. — Rambler.

Encoding a dictionary

Workflow

Lexicographical analysis

Entry structures: marker + keyword

Example: Der Aar

* Der Aar, des —es, oder —en, Mz. die —e, oder —en, die alte Benennung aller großer Raubvögel, besonders aber des Adlers, die noch in N. D. üblich ist und bei Dichtern vorkommt.

Ein führner Aar theilt mit gewalt'gen Schwingen
Die Lüfte, — — — — — Schreiber.
Bald werdet ihr im Meer der Haien, am Gestade
Der Aaren Beute sein. — Ramler.

Encoding a dictionary

Workflow

Lexicographical analysis

Entry structures: grammatical forms

Example: Der Aar

* Der Aar, des —es, oder —en, Mz. die —e, oder —en, die alte

Benennung aller großer Raubvögel, besonders aber des Adlers, die noch in R. D. üblich ist und bei Dichtern vorkommt.

Ein füher Aar theilt mit gewalt'gen Schwingen

Die Lüste, — — — — — Schreiber.

Bald werdet ihr im Meer der Haien, am Gestade

Der Aaren Beute sein. — Ramler.

Encoding a dictionary

Workflow

Lexicographical analysis

Entry structures: definition

Example: Der Aar

* Der Aar, des —es, oder —en, Mz. die —e, oder —en, die alte Benennung aller großer Raubvögel, besonders aber des Adlers, die noch in N. D. üblich ist und bei Dichtern vorkommt.

Ein führner Aar theilt mit gewalt'gen Schwingen
Die Lüste, — — — — — Schreiber.
Bald werdet ihr im Meer der Haien, am Gestade
Der Aaren Beute sein. — Ramler.

Encoding a dictionary

Workflow

Lexicographical analysis

Entry structures: context examples
(literary quotations)

Example: Der Aar

* Der Aar, des —es, oder —en, Mz. die —e, oder —en, die alte
Benennung aller großer Raubvögel, besonders aber des Adlers, die
noch in N. D. üblich ist und bei Dichtern vorkommt.

Ein führner Aar theilt mit gewalt'gen Schwingen
Die Lüfte, — — — — — Schreiber.
Bald werdet ihr im Meer der Haien, am Gestade
Der Aaren Beute sein. — Ramler.

Encoding a dictionary

- Workflow TEI Encoding structures

```
<entry>
  <form>
    <form> ... </form>
  </form>
  <sense>
    <def> ... </def>
    <note> ... </note>
    <etym> ... </etym>
    <cit> ... </cit>
    <re> ... </re>
  </sense>
</entry>
```

Encoding a dictionary

Workflow



1_form.pdf

Zitate werden durch <quote>-Tags ausgezeichnet. Die zugehörige Quelle wird in <bibl></bibl> angegeben. Die Einheit aus <quote> und <bibl> wird durch <cit type="quote">...</cit> geklammert.



2_sense.pdf

<cit>

Ein kühner Aar theilt mit gewalt'gen Schwingen
Die Lüfte, — — — — — Schreiber.



6_re.pdf

Bsp:

<sense> ...

<cit type="quote">

<quote> Ein kühner Aar theilt mit ...

</quote>

<bibl>

<author n="#Schreiber">

<hi rend="spaced">Schreiber</hi>

</author>

</bibl>

<c type="\$.">.</c>

</cit>

</sense>



7_usg.pdf



3_cit.pdf

Quelle <author n="#Name">Name</author>
<title>Titel</title>



8_Allgemeines.pdf



4_xr.pdf

in <bibl></bibl>

</cit>



5_note.pdf

<author> bezeichnet den Autor eines Zitates. Sein Name ist gesperrt gedruckt. Das Attribut n="#Name" wird vergeben, um immer eine eindeutige Zuordnung von Zitat und Autor zu gewährleisten, z.B. auch dann, wenn der Name abgekürzt ist oder der Verweis „derselbe“ auf das Vorausgegangene verweist

Encoding a dictionary

Workflow

Defining a TEI Relax NG Schema
and customize it according to the
Guidelines using Roma

Programming a converter to
produce valid TEI files for every
entry of the dictionary

Using the TEI header do keep track
of the workflow

Workflow

Encoding a dictionary

Converter output: valid TEI file

```
<?xml version="1.0" encoding="UTF-8" ?>
<?oxygen RNGSchema="../../dictionary_schemas/campe_dictionary.rnc" type="compact"?>
- <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="de">
- <teiHeader xmlns="">
- <fileDesc>
- <titleStmt>
<title>Joachim Heinrich Campe, Wörterbuch der Deutschen Sprache, Band 1 mit TEI-Auszeichnung</title>
<author>Kompetenzzentrum für EDV-Philologie, Universität Würzburg</author>
</titleStmt>
- <publicationStmt>
- <availability>
<p>für die Kodierungs- und Korrekturphase</p>
</availability>
<authority>Kompetenzzentrum für EDV-Philologie, Universität Würzburg</authority>
</publicationStmt>
- <sourceDesc>
<p>Wörtbucheintrag Aar</p>
- <biblStruct xml:lang="de" type="dictionary">
- <monogr>
<author>Joachim Heinrich Campe</author>
<title>Wörterbuch der Deutschen Sprache</title>
<title type="sub">Veranstaltet und herausgegeben von Joachim Heinrich Campe. Erster Theil. A - bis E.</title>
- <imprint>
<pubPlace>Braunschweig</pubPlace>
<publisher>In der Schulbuchhandlung.</publisher>
<date when="1807">1807</date>
| 
</imprint>
```

Workflow

Encoding a dictionary

Converter output: valid TEI file 2

```
</fileDesc>
- <encodingDesc>
- <editorialDecl>
- <normalization>
<p>Formaterfassung der Lemmata: Alle Lemmata (Stichwort ggf. mit Artikel) sind ausnahmslos in einem ca. 0,5  
Punkt größeren Font gesetzt; dieses wird nur an dieser Stelle notiert. Der negative Einzug (Linksaurückung)  
wird nicht kodiert</p>
<p>Typographische Besonderheiten: Alle übrigen typographischen Veränderungen sind mit dem Element <hi/>  
und spezifizierendem Attribut kodiert.</p>
</normalization>
</editorialDecl>
</encodingDesc>
- <revisionDesc>
<change who="#CG">2008-09-22 - automatisch generiert durch CPM</change>
<change who="#--">Überarbeitung</change>
<change who="#--">Textkontrolle</change>
<change who="#--">Übernahme in Bandkodierung</change>
</revisionDesc>
</teiHeader>
```

Workflow

Encoding a dictionary

Converter output: valid TEI file 3

```
<text xmlns="">
  <body>
    <div>
      <entry xml:id="cwds1_00045_aar">
        <form>
          <form type="lemma">
            <usg type="time"> <g>*</g> </usg>
            <gramGrp> <pos value="noun" /> <gen value="m" /> </gramGrp>
            <form type="determiner"> <orth>Der</orth> </form>
            <form type="headword"> <orth>Aar</orth> </form> <c type="$,">, </c>
          </form>
          <form type="inflected">
            <gramGrp> <case value="genitive" /> <number value="singular" /> </gramGrp>
            <form type="determiner"> <orth>des</orth> </form>
            <form type="headword"> <orth> <oVar> <oRef /> -- es </oVar> </orth>
            </form>
          </form>
          <form type="inflected">
            <gramGrp><case value="nominative" /><number value="plural">Mz.</number> </gramGrp>
            <form type="determiner"> <orth>die</orth> </form>
            <form type="headword"> <orth> <oVar> <oRef /> -- e </oVar> </orth> </form>
          </form>
        </form>
        <sense> oder <c type="$(">--</c> <hi rend="large">en</hi> <c type="$,">, </c>
          die alte <lb n="0002.036" /> Benennung aller großer Raubvögel <c type="$,">, </c>
          besonders aber des Adlers <c type="$,">, </c> die <lb n="0002.037" />
          noch <usg type="geo"> in N <c type="$.">. </c> D <c type="$.">. </c> </usg>
          üblich ist und <usg type="style">bei Dichtern</usg> vorkommt <c type="$.">. </c>
```

Encoding a dictionary

Workflow

Correcting the encoding manually

Final control of the encoding

Compiling larger units

Transform into baseline encoding
for efficient search

Encoding a dictionary

Workflow

Baseline encoding

```
<revisionDesc>
  <change who="#CG">Neuerstellung</change>
  <change who="#MB">20080430 Textkontrolle</change>
</revisionDesc> </teiHeader>

<text> <body> <div n="a" type="letter" xml:id="letter-a"> <lb/>
  <entry xml:id="cwds1_000045_Aar">
    <form type="lemma"> <gramGrp> <pos value="noun"/> <gen value="m"/> </gramGrp>
      <usg type="time">*</usg> Der <form type="headword">Aar</form>, des -es, oder -en,
      Mz. die -e, oder -en,</form>
  <sense>
    <def>die alte<lb/>Benennung aller großer Raubvögel, besonders aber des
    Adlers,</def> die<lb/>noch <usg type="geo">in N. D.</usg> üblich ist und <usg
    type="style">bei Dichtern</usg> vorkommt.<lb/>
    <cit type="example"> <cit type="verse">
      <quote>Ein kühner Aar theilt mit gewalt'gen Schwingen<lb/>Die Lüfte, - - - - -</quote>
    <bibl> <author n="#Schreiber">Schreiber.</author> </bibl> </cit>
      <lb/> <cit type="verse"> <quote>Bald werdet ihr im Meer der Haien, am Gestade<lb/>Der Aaren
      Beute sein. -</quote> <bibl> <author n="#Ramler">Ramler.</author></bibl> </cit>
```

Baseline Encoding for the compilation of dictionaries in TEI P5*

Mirjam Blümm, Stefan Büdenbender, Christian Graiger, Werner Wegstein

June 7th 2008

Abstract

This paper is concerned with the structuring of dictionary entries and gives an overview of the XML elements and element classes of the modules `core`, `dictionaries`, `textstructure` and `gaiji` of TEI P5, which were specified for the baseline encoding.

Open issues

- Treatment of obvious errata
- Ontology for attribute values (z.B. for quotations)
- Placement of meta information within an element `<entry>`

Contents

1 Baseline elements of TEI P5	2
1.1 XML elements	2
1.2 XML-Attribute	4
1.3 Examples	5
1.3.1 Dictionary entry "Der Aar"	5
1.3.2 Dictionary entry "Abändern"	7
2 Hierarchy	10
A Description of the schema	10

Encoding a dictionary

Guidelines to Baseline encoding

1.1 XML elements

Table 1 shows those XML elements of the module `dictionaries`, which are included in the baseline encoding. The marker *optional* refers to elements, which are used only if demanded by the structures of the dictionary entry. If an entry contains a definition, for example, the appropriate tag `<def>` has to be used. If not, it is omitted.

Table 1: elements of the baseline encoding (modul `dictionaries`)

Element	Function	Status
<code><entry></code>	contains a well-structured dictionary entry	required
<code><form></code>	groups all the information on the written and spoken forms of one headword	required
<code><sense></code>	groups together all information relating to one word sense in a dictionary entry, for example definitions, examples, translation equivalents, etymology ...	required
<code><gramGrp></code>	groups morpho-syntactic information about a lexical item, groups the elements <code><pos></code> , <code><gen></code> and <code><subc></code>	required
<code><pos></code>	indicates the part of speech	required
<code><subc></code>	contains subcategorization information ("transitiv" / "intransitiv")	optional
<code><gen></code>	identifies the morphological gender	optional
<code><def></code>	contains definition text in a dictionary entry	optional
<code><etym></code>	encloses etymological information	optional
<code><usg></code>	contains usage information	optional
<code><re></code>	contains a dictionary entry for a lexical item related to the headword, such as a compound phrase or derived form, embedded inside a larger entry	optional
<code><superEntry></code>	groups successive entries for a set of homographs	optional
<code><xr></code>	cross-reference phrase: contains a phrase, sentence, or icon referring the reader to some other location in this or another text	optional

The elements `<entry>`, `<form>` and `<sense>` are necessary prerequisites for search enquiries and thus represent the obligatory components of the baseline encoding. The element `<form>` groups all the information on the lemma, the element `<sense>` all information relating to the word sense. This applies also to entries, which consist of

Encoding a dictionary

Wer ein Wörterbuch schreibt,
baut recht eigentlich am Wege
und setzt sein Werk dem Tadel aus.

Matthias von Lexer

Encoding a dictionary

Thank you.



Overview — what should be remembered in this section

- Concepts related to the TEI architecture
 - Modules, classes
 - ODD, Roma
- Bibliographical elements in the TEI
 - <bibl>
 - <biblStruct>



Following the TEI spirit

Conformance to the TEI means:

- Sharing a common text encoding culture
- Sharing the same vocabulary (when applicable)
- Allowing user autonomy in defining modifications (extensions, customization), but sharing the mechanisms to do so

The TEI gives you a lot of help in following these rules.



Colourful TEI: the bibliographic elements

- <bibl> freely annotated in-line element
- <biblStruct> highly structured bibliographical entry
- <biblFull> complex structure mimicking the TEI header



<bibl> - examples

<bibl>Blain, Clements and Grundy: Feminist Companion to Literature in English (Yale, 1990)</bibl>

<bibl><title>Dictionnaire des difficultés de la langue française</title> (<author>V. Thomas</author>, <publisher>Larousse</publisher>)</bibl>.</item>

<bibl><idno type="ISO">ISO/IEC 2382 (all parts)</idno>, <title type="introductory">Information technology</title><title type="main">Vocabulary</title></bibl>



<biblStruct> - example

[Export from WALS online - <http://wals.info>]

```
<biblStruct type="incollection">
  <analytic>
    <author>
      <forename>Pliny Earle</forename>
      <surname>Goddard</surname>
    </author>
    <title type="main" level="a">Athapascan (Hupa)</title>
  </analytic>
  <monogr>
    <editor>Boas, Franz</editor>
    <title type="main" level="m">Handbook of American Indian Languages</title>
    <imprint>
      <pubPlace>Washington, D. C.</pubPlace>
      <publisher>Government Printing Office</publisher>
      <date>1911</date>
      <biblScope type="pp">85-158</biblScope>
    </imprint>
  </monogr>
</biblStruct>
```



Defining a TEI Schema

- A schema helps you know a document is valid in addition to being well-formed
- A TEI schema is a combination of TEI modules, optionally including customizations of the elements/attributes/classes that they contain
- This schema is defined in an application-independent manner with a TEI ODD (One Document Does it all) file which allows for:
 - creation of schemas using DTD, RELAX NG or W3C Schema languages
 - internationalized documentation which reflects your customization of the TEI



Important ODD concepts

The TEI's literate programming with ODD (One Document Does it all) provides:

- Schema specification
- User oriented documentation
- Modularity: all specifications pertaining to a coherent sub-domain of the TEI
- Classes: identifying shared behaviours or semantics
- Extensibility: a consequence of the above mechanisms



The TEI Class System

- The TEI distinguishes over 500 elements,
 - Having these organised into classes aids comprehension, modularity, and modification.
- **Attribute class:** the members share common attributes
- **Model class:** they can appear in the same locations (and often are structurally or semantically related)
- Classes may contain other classes
 - Elements inherit the properties from any classes of which they are members



Attribute Classes

- Attribute classes are given (usually adjectival) names beginning with att.:
 - e.g. members of the *att.canonical* class get a *key* attribute rather than have them define it individually
 - If another element needs a *key* attribute then the easiest way to provide it is to add it to the *att.naming* class
- Classes can be grouped together into a super classes



att.global

- All elements are members of att.global; this provides, among others:
 - xml:id
 - a unique identifier
 - xml:lang
 - the language of the element content
 - n
 - a number or name for an element
 - rend
 - how the element in question was rendered or presented in the source text.



Attribute classes — E.g. <biblStruct>/att.typed

<biblStruct> (structured bibliographic citation) contains a structured bibliographic citation, in which only bibliographic sub-elements appear and in a specified order. [3.11.1 Elements of Bibliographic References](#) [2.2.7 The Source Description](#) [15.3.2 Declarable Elements](#)

Module	core — 3 Elements Available in All TEI Documents
Attributes	att.declarable (@default) att.typed (@type, @subtype)
Used by	model.biblLike
May contain	core: analytic monogr note relatedItem series header: idno textcrit: witDetail
Declaration	<div style="border: 1px solid #ccc; padding: 10px;"><pre>element biblStruct { att.global.attributes, att.declarable.attributes, att.typed.attributes, (analytic*, (monogr, series*)+, (model.noteLike idno relatedItem)*) }</pre>Compact to XML format</div>



att.typed — definition

att.typed

att.typed provides attributes which can be used to classify or subclassify elements in any way. [1.3.1 Attribute Classes](#)

Module	tei — 1 The TEI Infrastructure
Attributes	<p>In addition to global attributes</p> <p>@type characterizes the element in some sense, using any convenient classification scheme or typology.</p> <p>@subtype provides more specific information about the type of element.</p> <pre><biblStruct type="incollection"> <analytic> <author>...</author> <title type="main" level="a">Athapascan (Hupa)</title> </analytic> <monogr> <editor>...</editor> <title type="main" level="m">Handbook of ...</title> <imprint> <biblScope type="pp">85-158</biblScope> </imprint> </monogr> </biblStruct></pre>



Model Classes

- Model classes contain groups of elements allowed in the same place.
 - e.g. if you are adding an element which is wanted wherever the <bibl> is allowed, add it to the model.biblLike class
- Model classes are usually named with a Like or Part suffix:
 - model.divLike: structural class grouping elements for divisions
 - model.divPart: structural class grouping elements used inside divisions
 - model.nameLike: semantic class grouping name elements
 - model.persNamePart: semantic sub-class grouping elements that are part of a personal name



Model classes — E.g. <bibl>/model.biblPart

```
element bibl{  
    att.global.attributes,  
    att.declarable.attributes,  
    att.typed.attributes,  
    (  
        text  
        | model.gLike  
        | model.highlighted  
        | model.pPart.data  
        | model.pPart.edit  
        | model.segLike  
        | model.ptrLike  
        | model.biblPart  
        | model.global  
    )*  
}
```



P5: Guidelines for Electronic Text Encoding and Interchange

model.biblPart

[Home](#) | [Table of contents](#)
[A Model Classes](#)

model.biblPart groups elements which represent components of a bibliographic description.

[3.11 Bibliographic Citations and References](#)

Module	tei — 1 The TEI Infrastructure
Used by	bibl
Members	model.imprintPart [biblScope distributor pubPlace publisher] model.respLike [author editor respStmt] edition extent idno meeting msIdentifier relatedItem series

[\[English\]](#) [\[Deutsch\]](#) [\[Español\]](#) [\[Italiano\]](#) [\[Français\]](#) [\[日本語\]](#) [\[中文\]](#)



The TEI ODD in practice

The TEI Guidelines, its schema, and its schema fragments, are all produced from a single XML resource containing:

1. Descriptive prose (lots of it)
2. Examples of usage (plenty)
3. Formal declarations for components of the TEI Abstract Model:
 - elements and attributes
 - Modules
 - classes and macros



Possibilities of customizing the TEI

The TEI has over 20 modules. A working project will:

- Choose the modules they need
- Probably narrow the set of elements within a module
- Probably add local datatype constraints
- Possibly add new elements
- Possibly localize the names of elements



Real life TEI customization

We aim to support a range of interactions with the TEI:

Easy TEI

Simple access to the TEI through Roma

Subsetting the TEI

Making the TEI even easier to use

Enlarging the application profile

Using modules

Modifying the TEI objects

First insights into extensibility

Behind the scene - ODD

Starting to use the actual specification language



Roma

- The TEI knows you don't want to necessarily have to write TEI code in order to customize the TEI. So it has provided Roma, which is a command-line script, and corresponding web front-end to help you do this.
- The people behind Roma are:
 - Arno Mittelbach
 - Initial programming
 - Sebastian Rahtz
 - Maintenance and frequent improvements
 - Iain Bernevig
 - A 'Sanity Checker' addition



Quick and simple access to the TEI

- Go to Roma at <http://tei.oucs.ox.ac.uk/Roma/>
- Generate a schema [Schema]
- Make a trial with the editor, creating a simple document

- That's all!



Scientific bibliographies

- additional needs
 - precise identification of persons and their affiliations
 - e.g. organisation, location, addresses



```
<biblStruct>
  <monogr>
    <author>
      <forename>Laurent</forename>
      <surname>Romary</surname>
      <affiliation>
        <orgName>Max Planck Digital Library</orgName>
        <address>
          <street>Invalidenstr. 35</street>
          <postCode>10115</postCode>
          <settlement>Berlin</settlement>
          <country>Deutschland</country>
        </address>
        <email>...@...</email>
      </affiliation>
    </author>
    <title>...</title>
    <imprint>
      <date></date>
      <publisher></publisher>
    </imprint>
  </monogr>
</biblStruct>
```



Roma: Start



Roma: generating validators for the TEI

These pages will help you design your own TEI validator, as a DTD, RELAXNG or W3C Schema.

Create a new or upload existing customization

- Build schema (Create a new customisation by adding elements and modules to the smallest recommended schema)
- Reduce schema (Create a new customisation by removing elements and modules from the largest possible schema)
- Create customization from template

TEI Absolutely Bare

TEI Absolutely Bare

TEI Lite

TEI for Linguistic Corpora

TEI for Manuscript Description

TEI with Drama

TEI for Speech Representation

TEI for authoring ODD

TEI with SVG

TEI with MathML

TEI with XInclude (experimental)

TEI for Dictionaries (experimental)

Submit

Search TEI Guide



Roma: Schema Select

 **TEI Roma: generating validators for the TEI**

Time to give you a schema

New Customize Language Modules Add Elements Change Classes **Schema** Documentation Save Customization

Creating a schema

Which format do you prefer?

Relax NG schema (compact syntax) ▾
Relax NG schema (compact syntax)
Relax NG schema (XML syntax) **(selected)**
W3C schema
DTD

Submit

Search TEI database

Roma was written by Arno Mittelbach and is maintained by Sebastian Rahtz. Please direct queries to the [TEI @ Oxford](#) project.
This is Roma version 2.7, last updated 2006-10-29.



Roma: Generate Doc

TEI Roma: generating validators for the TEI Documentation?

New Customize Language Modules Add Elements Change Classes Schema **Documentation** Save Customization

Getting some nice documentation

Which output would you prefer?

html

html

PDF

TEI Lite

Tei

Roma was written by Arno Mittelbach and is maintained by Sebastian Rahtz. Please direct queries to the [TEI @ Oxford](#) project.
This is Roma version 2.7, last updated 2006-10-29.



Subsetting the TEI

Suppose you now feel you want to use some more of the TEI, but not all of it

- Go to Roma...
- Look at [Modules]
- Explore default modules by pointing to main elements (by order of interest). You can throw away most things, but
 - In **textstructure**, you should really keep <TEI>, <text>, <body> and <div>
 - In **core**, most people need <p>, <q>, <list>, <pb/> and <head>
 - From **header**, keep everything unless you really understand the details
- Start checking out elements
- Make editorial choices (numbered vs. unnumbered divs)



Roma: Customize



Roma: generating validators for the TEI

Set your parameters

[New](#) [Customize](#) [Language](#) [Modules](#) [Add Elements](#) [Change Classes](#) [Schema](#) [Documentation](#) [Save Customization](#)

Set your parameters

Title

My TEI Extension

Filename

myTei

Prefix for TEI
pattern names
in schema

Language

- English
- Deutsch
- Français
- Russian
- Svenska
- 日本語
- 中文

Author name

generated by Roma 2.7

Description

My TEI Customization
starts with modules tei, core, header, and
textstructure



Roma: Modules

TEI Roma: generating validators for the TEI Modules

New Customize Language **Modules** Add Elements Change Classes Schema Documentation Save Customization

List of TEI Modules

Module name	A short description	Changes
add analysis	Simple analytic mechanisms	
add certainty	Certainty and uncertainty	
add core	Elements common to all TEI documents	
add corpus	Header extensions for corpus texts	
add declarefs	Feature system declarations	
add dictionaries	Printed dictionaries	
add drama	Performance texts	
add figures	Tables, formulae, and figures	
add gaiji	Character and glyph documentation	
add header	The TEI Header	
add iso-fs	Feature structures	
add linking	Linking, segmentation and alignment	
add msdescription	Manuscript Description	
add namesdates	Names and dates	
add nets	Graphs, networks and trees	

List of selected Modules

[remove core](#)
[remove tei](#)
[remove header](#)
[remove textstructure](#)



Roma: Change Module

TEI Roma: generating validators for the TEI

Change module

[New](#) [Customize](#) [Language](#) [Modules](#) [Add Elements](#) [Change Classes](#) [Schema](#) [Documentation](#) [Save Customization](#)

[back](#)

List of elements in module: core

	Include	Exclude	Tag name	Description	Attributes
abbr	<input checked="" type="radio"/>	<input type="radio"/>	abbr	contains an abbreviation of any sort.	Change attributes
add	<input checked="" type="radio"/>	<input type="radio"/>	add	contains letters, words, or phrases inserted in the text by an author, scribe, annotator, or corrector.	Change attributes
addrLine	<input checked="" type="radio"/>	<input type="radio"/>	addrLine	contains one line of a postal or other address.	Change attributes
address	<input checked="" type="radio"/>	<input type="radio"/>	address	contains a postal or other address, for example of a publisher, an organization, or an individual.	Change attributes
altIdent	<input checked="" type="radio"/>	<input type="radio"/>	altIdent	supplies the recommended XML name for an element, class, attribute, etc. in some language.	Change attributes
analytic	<input checked="" type="radio"/>	<input type="radio"/>	analytic	contains bibliographic elements describing an item (e.g. an article or poem) published within a monograph or journal and not as an independent publication.	Change attributes
author	<input checked="" type="radio"/>	<input type="radio"/>	author	in a bibliographic reference, contains the name of the author(s), personal or corporate, of a work; the primary statement of responsibility for any bibliographic item.	Change attributes
bibl	<input checked="" type="radio"/>	<input type="radio"/>	bibl	contains a loosely-structured bibliographic citation of	Change



Roma: Show Class



att.divLike	provides a set of attributes common to all elements which behave in the same way as divisions.
Class	att.metrical
Attributes	<p>(In addition to global attributes and those inherited from [att.metrical])</p> <p>type specifies a name conventionally used for this level of subdivision, e.g. act, volume, book, section, canto, etc.</p> <p>Status: Recommended</p> <p>Datatype: data.enumerated</p> <p>Values: any string of characters</p> <p>org specifies how the content of the division is organized.</p> <p>Status: Optional</p> <p>Default: uniform</p> <p>Legal values are:</p> <p>composite (composite content: i.e. no claim is made about the sequence in which the immediate contents of this division are to be processed, or their inter-relationships.)</p> <p>uniform (uniform content: i.e. the immediate contents of this element are regarded as forming a logical unit, to be processed in sequence.)</p> <p>sample indicates whether this division is a sample of the original source and if so, from which part.</p> <p>Status: Optional</p> <p>Default: complete</p> <p>Legal values are:</p> <p>initial (division lacks material present at end in source.)</p> <p>medial (division lacks material at start and end.)</p>



Roma: Change Class Attributes

TEI Roma: generating validators for the TEI

Added Attributes

New Customize Language Modules Add Elements Change Classes Schema Documentation Save Customization

List of attributes

Add new attributes

Change attribute	Include	Exclude	Tag name	Description	Delete
<u>type</u>	<input checked="" type="radio"/>	<input type="radio"/>	type	specifies a name conventionally used for this level of subdivision, e.g. act, volume, book, section, canto, etc.	Delete
<u>org</u>	<input checked="" type="radio"/>	<input type="radio"/>	org	specifies how the content of the division is organized.	Delete
<u>sample</u>	<input checked="" type="radio"/>	<input type="radio"/>	sample	indicates whether this division is a sample of the original source and if so, from which part.	Delete
<u>part</u>	<input checked="" type="radio"/>	<input type="radio"/>	part	specifies whether or not the division is fragmented by some other structural element, for example a speech which is divided between two or more verse stanzas.	Delete

[Submit Query](#)

[Search TEI database](#)

Roma was written by Arno Mittelbach and is maintained by Sebastian Rahtz. Please direct queries to the [TEI @ Oxford](#) project.
This is Roma version 2.7, last updated 2006-10-29.



Adding TEI objects

You can add your own elements and attributes. But

- make very sure you are not just making something which is syntactic sugar for an existing TEI concept
- do not rename existing elements - you can do that directly in ODD
- if you want facilities from a very different field of discourse, such as maths or vector graphics, use the existing standards in that area
- consider interoperability



Roma: Add Element

TEI Roma: generating validators for the TEI

Add Element

[New](#) [Customize](#) [Language](#) [Modules](#) [Add Elements](#) [Change Classes](#) [Schema](#) [Documentation](#) [Save Customization](#)

[go back to list](#)

Defining a new element

Name

Description

Model classes

- model.addrPart
- model.assertLike
- model.biblLike
- model.castItemPart
- model.catDescPart
- model.choicePart
- model.common
- model.complexVal
- model.dateLike
- model.datePart
- model.divPart
- model.divPart.spoken
- model.divPart.stage
- model.divPart.verse
- model.divWrapper
- model.divWrapper.bottom
- model.editorialDeclPart
- model.encodingPart
- model.entryLike
- model.entryParts
- model.entryParts.top
- model.featureVal
- model.formPart
- model.frontPart
- model.global
- model.global.edit
- model.gLike
- model.glossLike
- model.global.meta
- model.headerPart
- model.hiLike
- model.graphicLike
- model.inter
- model.gramPart
- model.imprintPart
- model.ILike



Roma: language Selection

TEI Roma: generating validators for the TEI

Choose a different language

New Customize Language Modules Add Elements Change Classes Schema Documentation Save Customization

Do you want the output schema to use a different language?

Language for element and attribute names

- English
- Deutsch
- Español
- Français
- 日本語
- 中文

Language for documentation

- English
- Deutsch
- Español
- Français
- 日本語
- 中文

Submit Query

Search TEI database

Roma was written by Arno Mittelbach and is maintained by Sebastian Rahtz. Please direct queries to the [TEI @ Oxford](#) project.
This is Roma version 2.7, last updated 2006-10-29.



Roma: In French (sorry...)

TEI Roma: generating validators for the TEI

Change module

	New	Customize	Language	Modules	Add Elements	Change Classes	Schema	Documentation	Save Customization																																				
back																																													
Liste des éléments compris dans ce module header																																													
<table border="1"><thead><tr><th></th><th>Inclure</th><th>Exclude</th><th>Nom de balise</th><th>Description</th><th>Attributs</th></tr></thead><tbody><tr><td>authority</td><td><input type="radio"/></td><td><input checked="" type="radio"/></td><td>authority</td><td><p>L'élément Responsable de la mise à disposition authority fournit le nom d'une personne ou d'une institution responsable de la mise à disposition d'un fichier électronique, autre qu'un éditeur ou un distributeur.</p></td><td>Changer les attributs</td></tr><tr><td>availability</td><td><input type="radio"/></td><td><input checked="" type="radio"/></td><td>availability</td><td><p>L'élément Disponibilité availability fournit des informations sur la disponibilité du texte, par exemple toutes restrictions sur son usage ou sa diffusion, son statut de copyright, etc.</p></td><td>Changer les attributs</td></tr><tr><td>broadcast</td><td><input type="radio"/></td><td><input checked="" type="radio"/></td><td>broadcast</td><td><p>L'élément Diffusion broadcast décrit une émission utilisée comme source de la parole.</p></td><td>Changer les attributs</td></tr><tr><td>cRefPattern</td><td><input type="radio"/></td><td><input checked="" type="radio"/></td><td>cRefPattern</td><td><p>L'élément Adressage de fragments fragmentPattern spécifie un modèle d'expression et des règles de remplacement pour transformer une référence canonique en URI.</p></td><td>Changer les attributs</td></tr><tr><td>catDesc</td><td><input type="radio"/></td><td><input checked="" type="radio"/></td><td>catDesc</td><td><p>L'élément Description de catégorie catDesc décrit une certaine catégorie à l'intérieur d'une taxinomie ou d'une typologie de texte, sous forme d'une brève description en prose suivie ou en termes de paramètres contextuels</p></td><td>Changer les attributs</td></tr></tbody></table>											Inclure	Exclude	Nom de balise	Description	Attributs	authority	<input type="radio"/>	<input checked="" type="radio"/>	authority	<p>L'élément Responsable de la mise à disposition authority fournit le nom d'une personne ou d'une institution responsable de la mise à disposition d'un fichier électronique, autre qu'un éditeur ou un distributeur.</p>	Changer les attributs	availability	<input type="radio"/>	<input checked="" type="radio"/>	availability	<p>L'élément Disponibilité availability fournit des informations sur la disponibilité du texte, par exemple toutes restrictions sur son usage ou sa diffusion, son statut de copyright, etc.</p>	Changer les attributs	broadcast	<input type="radio"/>	<input checked="" type="radio"/>	broadcast	<p>L'élément Diffusion broadcast décrit une émission utilisée comme source de la parole.</p>	Changer les attributs	cRefPattern	<input type="radio"/>	<input checked="" type="radio"/>	cRefPattern	<p>L'élément Adressage de fragments fragmentPattern spécifie un modèle d'expression et des règles de remplacement pour transformer une référence canonique en URI.</p>	Changer les attributs	catDesc	<input type="radio"/>	<input checked="" type="radio"/>	catDesc	<p>L'élément Description de catégorie catDesc décrit une certaine catégorie à l'intérieur d'une taxinomie ou d'une typologie de texte, sous forme d'une brève description en prose suivie ou en termes de paramètres contextuels</p>	Changer les attributs
	Inclure	Exclude	Nom de balise	Description	Attributs																																								
authority	<input type="radio"/>	<input checked="" type="radio"/>	authority	<p>L'élément Responsable de la mise à disposition authority fournit le nom d'une personne ou d'une institution responsable de la mise à disposition d'un fichier électronique, autre qu'un éditeur ou un distributeur.</p>	Changer les attributs																																								
availability	<input type="radio"/>	<input checked="" type="radio"/>	availability	<p>L'élément Disponibilité availability fournit des informations sur la disponibilité du texte, par exemple toutes restrictions sur son usage ou sa diffusion, son statut de copyright, etc.</p>	Changer les attributs																																								
broadcast	<input type="radio"/>	<input checked="" type="radio"/>	broadcast	<p>L'élément Diffusion broadcast décrit une émission utilisée comme source de la parole.</p>	Changer les attributs																																								
cRefPattern	<input type="radio"/>	<input checked="" type="radio"/>	cRefPattern	<p>L'élément Adressage de fragments fragmentPattern spécifie un modèle d'expression et des règles de remplacement pour transformer une référence canonique en URI.</p>	Changer les attributs																																								
catDesc	<input type="radio"/>	<input checked="" type="radio"/>	catDesc	<p>L'élément Description de catégorie catDesc décrit une certaine catégorie à l'intérieur d'une taxinomie ou d'une typologie de texte, sous forme d'une brève description en prose suivie ou en termes de paramètres contextuels</p>	Changer les attributs																																								



Under the hood

TEI customizations are themselves expressed in TEI XML, using elements from the **tagdocs** module.

For example:

```
<schemaSpec ident="myTEILite">
    <desc>This is TEI Lite with simplified heads</desc>
    <moduleRef key="tei"/>
    <moduleRef key="core"/>
    <moduleRef key="textstructure"/>
    <moduleRef key="header"/>
    <moduleRef key="linking"/>
    <elementSpec ident="head" mode="change">
        <content>
            <rng:text/>
        </content>
    </elementSpec>
</schemaSpec>
```

produces something like TEI Lite, with a slight change



ODD processors

- The TEI maintains a library of XSLT scripts that can generate
 - The TEI Guidelines in canonical TEI XML format
 - The Guidelines in HTML or PDF
 - RELAXNG, DTD, or W3C schema fragments
- The same library is used by the customization layer to generate
 - project-specific documentation
 - project-specific schemas
 - translations into other (human) languages



The TEI abstract model

- The TEI abstract model sees a markup scheme (a schema) as consisting of a number of discrete modules, which can be combined more or less as required.
- A *schema* is made by combining references to *modules* and optional element over-rides or additions
- Each *element* declares the module it belongs to: elements cannot appear in more than one module.
- Each module extends the range of elements and attributes available by adding new members to existing *classes of elements*, or by defining new classes.



Expression of TEI content models

TEI elements have to be defined using some language notation;
choices include:

1. using XML DTD language (as in older versions of the TEI)
2. using W3C schema language
3. using the RELAXNG schema language
4. inventing an entirely new abstract language for later transformation to specific schema language

We chose a combination of 3 and 4 - using our abstract language, but switching to RELAXNG for content modelling.



What does an ODD look like?

```
<elementSpec module="spoken" ident="pause">
  <classes>
    <memberOf key="model.divPart.spoken"/>
    <memberOf key="att.timed"/>
    <memberOf key="att.typed"/>
  </classes>
  <content>
    <rng:empty/>
  </content>
  <attList>
    <attDef ident="who" usage="opt">
      <gloss>A unique identifier</gloss>
      <desc>supplies the identifier of the person or group pausing. Its value is the identifier of a
      <gi>person</gi> or <gi>persGrp</gi> element in the TEI header.
    </desc>
    <datatype>
      <rng:ref name="data.pointer"/>
    </datatype>
  </attDef>
  </attList>
  <desc>a pause either between or within utterances.</desc>
</elementSpec>
```



... from which we generate

```
element pause { pause.content, pause.attributes }
  pause.content = empty
  pause.attributes =
    att.global.attributes,
    att.timed.attributes,
    att.typed.attributes,
    att.ascribed.attributes,
  model.divPart.spoken |= pause
  att.timed |= pause
  att.typed |= pause
  att.ascribed |= pause
```



.. or

```
<!ELEMENT %n.pause; %om.RR; EMPTY>
<!ATTLIST %n.pause;
  %att.global.attributes;
  %att.timed.attributes;
  %att.typed.attributes;
  %att.ascribed.attributes;>
<!ENTITY % model.divPart.spoken
  "%x.model.divPart.spoken; %n.event; | %n.kinesic;
  | %n.pause; | %n.shift; | %n.u;
  | %n.vocal; | %n.writing;">
```



... and, indeed

<pause/>

[Home](#) | [Table of Contents](#)
[Element catalogue](#)

<pause/> a pause either between or within utterances. [8.3.2 Pausing](#)

Module	spoken — 8 Transcriptions of Speech
Attributes	[att.timed att.typed att.ascribed]
Declaration	<div style="border: 1px solid #ccc; padding: 10px;"><pre>element pause { att.global.attributes, att.timed.attributes, att.duration.w3c.attributes, att.typed.attributes, att.ascribed.attributes, empty }</pre></div> <div style="text-align: right; margin-top: -10px;">Compact to XML format</div>
Example	<pre><pause dur="PT42S" type="pregnant"/></pre>
Contained by	model.global.spoken
May contain	Empty element



Adding a new element

```
<schemaSpec>
  <moduleRef key="header"/>
  <moduleRef key="core"/>
  <moduleRef key="text"/>
  <moduleRef key="textstructure"/>
  <elementSpec ident="soundClip" mode="add">
    <classes>
      <memberOf key="model.graphicLike"/>
    </classes>
    <attList>
      <attDef ident="location">
        <desc>supplies the location of the clip</desc>
        <datatype>
          <rng:ref name="data.pointer"/>
        </datatype>
      </attDef>
    </attList>
    <desc>includes an audio object in a document.</desc>
  </elementSpec>
</schemaSpec>
```



Uniformity of description

- modules, elements, attributes, value-lists are treated uniformly
 - each has an identifier, a gloss, a description, and one or more equivalents
 - each can be added, changed, replaced, deleted within a given context
- Example
 - membership in the att.typed class gives you a generic *type* attribute, which can be over-ridden for specific class members



Overriding a value-list

```
<elementSpec ident="list" module="core">
    <classes>
        <memberOf key="att.typed"/>
    </classes>
    <attDef ident="type" mode="replace">
        <valList type="closed">
            <valItem ident="ordered">
                <gloss>Items are ordered</gloss>
            </valItem>
            <valItem ident="bulleted">
                <gloss>Items are bulleted</gloss>
            </valItem>
            <valItem ident="frabjous">
                <gloss>Items are frabjous</gloss>
            </valItem>
        </valList>
    </attDef>
</elementSpec>
```