

ECHO Schema Usage Guide

version 0.1

Wolfgang Schmidle

Max Planck Institute for the History of Science, Berlin, Germany

29th March 2010

Contents

1	The Modules of the Schema	6
1.1	echo	6
1.2	Overview	6
1.2.1	The File echo.rnc	9
1.3	echo-start	10
1.3.1	The File echo-start.rnc	10
1.3.2	<echo>	10
1.4	echo-metadata	10
1.4.1	The File echo-metadata.rnc	10
1.4.2	<dcterms:identifier>	12
1.4.3	<dcterms:title>, <dcterms:alternative>	12
1.4.4	<dcterms:creator>, <dcterms:contributor>	12
1.4.5	<dcterms:publisher>	12
1.4.6	<dcterms:language>	12
1.4.7	<dcterms:date>	12
1.4.8	<dcterms:description>	12
1.4.9	<dcterms:rights>, <dcterms:license>, <dcterms:accessRights>	12
1.4.10	<dcterms:rightsHolder>	12
1.4.11	<dcterms:provenance>	12
1.4.12	<dcterms:dateCopyrighted>	12
1.4.13	, <font-family>	12
1.4.14	<stylesheet>	12
1.4.15	<echolink>	13
1.5	echo-div	13
1.5.1	The File echo-div.rnc	13
1.5.2	<text>	13
1.5.3	<div>	13

1.6	echo-block	14
1.6.1	The File echo-block.rnc	14
1.6.2	<head>	14
1.6.3	<s>	14
1.6.4	<note>	15
1.6.5	<set-off>	15
1.7	echo-content	15
1.7.1	The File echo-content.rnc	15
1.7.2	<reg>	16
1.7.3	<ref>	17
1.7.4	<sic>	17
1.7.5	<foreign>	17
1.7.6	<emph>	18
1.7.7	<q>, <quote>	18
1.7.8	Avoiding Recursions	18
1.8	echo-gap	19
1.8.1	The File echo-gap.rnc	19
1.8.2	<gap>	19
1.8.3	<unsure>	19
1.9	echo-de	19
1.9.1	The File echo-de.rnc	19
1.9.2	<unknown>	20
1.9.3	<wrong>	20
1.10	echo-figure	20
1.10.1	The File echo-figure.rnc	20
1.10.2	<figure>, <caption>, <description>, <variables>	21
1.10.3	<image>	21
1.11	echo-handwritten	21
1.11.1	The File echo-handwritten.rnc	21
1.11.2	<handwritten>	21
1.12	echo-float	21
1.12.1	The File echo-float.rnc	21
1.12.2	<anchor>	22
1.13	echo-milestone	22
1.13.1	The File echo-milestone.rnc	22
1.13.2	<lb>	23
1.13.3	<cb>	23
1.13.4	<pb>	24
1.14	echo-attribute	24
1.14.1	The File echo-attribute.rnc	24
1.15	echo-datatype	24
1.15.1	The File echo-datatype.rnc	24
1.16	echo-mathematics	25
1.16.1	The File echo-mathematics.rnc	25

1.16.2	<num>	26
1.16.3	<var>	26
1.17	echo-chinese-text	26
1.17.1	The File echo-chinese-text.rnc	26
1.18	echo-gis	28
1.18.1	The File echo-gis.rnc	28
1.18.2	Metadata	28
1.18.3	<place>	29
1.18.4	<time>	29
1.18.5	<person>, <event>	29
1.19	echo-textflows	29
1.19.1	The File echo-textflows.rnc	29
1.20	echo-import-xhtml	30
1.20.1	The File echo-import-xhtml.rnc	30
1.20.2	xhtml-basic-table	30
1.20.3	xhtml-list	31
1.20.4	xhtml-attrs	31
1.20.5	xhtml-datatypes	31
1.20.6	xhtml-text	31
1.20.7	<xhtml:*>	32
1.21	echo-import-mathml	32
1.21.1	The File echo-import-mathml.rnc	32
1.21.2	<mathml:*>	32
2	Offene Punkte	32
2.0.3	, <font-family>	33
2.1	echo-block	33
2.1.1	<head>	33
2.1.2	<p>	33
2.1.3	<s>	33
2.1.4	<note>	33
2.1.5	<p-set-off>	33
2.2	echo-content	33
2.2.1	<reg>	33
2.2.2	<ref>	33
2.2.3	<sic>	33
2.2.4	<foreign>	33
2.2.5	<emph>	33
2.2.6	<q>	33
2.2.7	<quote>	33
2.3	echo-gap	33
2.3.1	<gap>	34
2.3.2	<unsure>	34
2.4	echo-de	34
2.4.1	<unknown>	34

2.4.2	<wrong>	34
2.5	echo-figure	34
2.5.1	<figure>, <caption>, <description>, <variables>	34
2.5.2	<image>	34
2.6	echo-handwritten	34
2.6.1	<handwritten>	34
2.7	echo-float	34
2.7.1	<anchor>	34
2.8	echo-milestone	35
2.8.1	<lb>	35
2.8.2	<cb>	35
2.8.3	<pb>	35
2.9	echo-attribute	35
2.10	echo-datatype	35
2.11	echo-mathematics	35
2.11.1	<num>	36
2.11.2	<var>	36
2.12	echo-chinese-text	36
2.13	echo-gis	36
2.13.1	<place>	36
2.13.2	<time>	36
2.13.3	<person>, <event>	36
2.14	echo-textflows	36
2.15	echo-import-xhtml	36
2.15.1	<xhtml:*>	36
2.16	echo-import-mathml	36
2.16.1	<mml:*>	36
3	Wunschliste für die Darstellung der Elemente	37
3.1	echo	37
3.2	echo-start	37
3.2.1	<echo>	37
3.3	echo-metadata	37
3.3.1	<dcterms:creator>	37
3.3.2	<dcterms:language>	37
3.3.3	<dcterms:accessRights>	37
3.3.4	, <font-family>	37
3.4	echo-div	38
3.4.1	<text>	38
3.4.2	<div>	38
3.5	echo-block	38
3.5.1	<head>	38
3.5.2	<s>	38

3.5.3	<note>	39
3.5.4	<set-off>	39
3.6	echo-content	39
3.6.1	<reg>	39
3.6.2	<ref>	39
3.6.3	<sic>	40
3.6.4	<foreign>	40
3.6.5	<emph>	40
3.6.6	<q>	40
3.6.7	<quote>	40
3.7	echo-gap	40
3.7.1	<gap>	40
3.7.2	<unsure>	41
3.8	echo-de	41
3.8.1	<unknown>	41
3.8.2	<wrong>	41
3.9	echo-figure	41
3.9.1	<figure>, <caption>, <description>, <variables>	41
3.9.2	<image>	41
3.10	echo-handwritten	42
3.10.1	<handwritten>	42
3.11	echo-float	42
3.11.1	<anchor>	42
3.12	echo-milestone	43
3.12.1	<lb>	43
3.12.2	<cb>	43
3.12.3	<pb>	43
3.13	echo-attribute	44
3.14	echo-datatype	44
3.15	echo-mathematics	44
3.15.1	<num>	44
3.15.2	<var>	44
3.16	echo-chinese-text	45
3.17	echo-gis	45
3.17.1	<place>	46
3.17.2	<time>	46
3.17.3	<person>, <event>	46
3.18	echo-textflows	46
3.19	echo-import-xhtml	47
3.19.1	<xhtml:*>	47
3.20	echo-import-mathml	47
3.20.1	<mml:*>	47

1 The Modules of the Schema

Design decisions:

- one schema for all texts
- modules should be independent
- tags in the DESpecs should have some counterpart in the Schema, if possible
- however, do not mimic the DESpecs structure

Standard suffixes as in xhtml:

- .attrib (defined in echo-attribute)
- .datatype (defined in echo-datatype)
- .model (defined in echo-content)
- .class (defined in several modules)

IDs are expected, but XML texts should also validate without IDs

1.1 echo

1.2 Overview

Der Kern des Schemas: "(-)" bedeutet: nicht minimal, aber man bekommt etwas dafür, was über die optische Anzeige des Textes hinausgeht

Ebene	minimal	China	Modul	Element
Grobstruktur	+	-	echo-start	<echo>
	+	-	echo-metadata	<metadata>
	+/-	-	echo-metadata	Metadaten
	+	-	echo-div	<text>
Mittelstruktur	(-)	+	echo-div	<div index>, <div toc>
	(-)	-	echo-div	<div chapter> etc.
Feinstruktur	+	+	echo-block	<head>, <p>, <note>
	+	-	echo-block	<s>
	-	+	echo-block	<quote>
	-	-	echo-block	<set-off>
Text	-	+	echo-content	<emph>
	(-)	-	echo-content	<reg>, <foreign>, <ref>
	-	-	echo-content	<sic>, <set-off>, <q>
Hilfsmodule	+	n/a	echo-attribute	-
	+	n/a	echo-datatype	-

F: Figures	-	+	echo-figure	<figure>, <caption> etc.
F: Handwritten	-	+	echo-handwritten	<handwritten>
F: Chinese text	-	+	echo-chinese-text	<head ti>, <p @indent>, <p @ics>, <pb @ics>
F: Textflows	-	+	echo-textflows	<head @flow>, <p @flow> <div multifold @flows>
F: Tables	-	+	echo-imp.-xhtml	<xhtml:table>
F: Lists	-	-	echo-imp.-xhtml	<xhtml:ul> etc.
F: Floats	-	-	echo-float	<div float>
F: Images	-	-	echo-figure	<image>
T: Milestones	(-)	+	echo-milestone	<pb> (auch F), <lb>, <cb>
T: Corruptions	-	+	echo-gap	<gap>, <unsure>
	-	+	echo-de	<de:unknown>, <de:wrong>
T: Chinese notes	-	+	echo-chinese-text	<lb halfline>
T: Floats	-	+	echo-float	<anchor>
T: numbers etc.	(-)	-	echo-mathematics	<num>, <var>
T: formulas	(-)	-	echo-imp.-mathml	<mml:math>
T: Verse	-	-	echo-textflows	<lb @label>
T: Images	-	-	echo-figure	<image>
T: Gis	-	-	echo-gis	<place>, <time>
G: Gis	-	-	echo-gis	<dcterms:temporal>, <dcterms:spatial>

- echo.rnc

ist die Haupt-Datei des Schemas. Hier werden alle Module geladen.

- echo-start.rnc
- echo-metadata.rnc
- echo-div.rnc
- echo-block.rnc

definieren die Struktur des XML-Dokuments bis vom root element <echo> bis zum Text-Level:

```

<echo>
  <metadata>      echo-start.rnc
  <text>          echo-metadata.rnc
    <div>         echo-div.rnc
      - " -
      <head>     echo-block.rnc
      <p>        - " -
        <s>      - " -
      <note>    - " -

```

echo-start.rnc: Zum root element <echo> gibt es kein Gegenstück in den DESpecs. Es muss in das Dokument eingefügt werden.

echo-metadata.rnc: Zu <metadata> und allen darin enthaltenen Metadaten gibt es kein Gegenstück in den DESpecs. <metadata> und einige Metadaten sind notwendig und müssen in das Dokument eingefügt werden. Andere Metadaten sind optional.

echo-div.rnc: Zu <text> und <div> gibt es ebenfalls kein Gegenstück in den DESpecs. <text> ist notwendig und muss in das Dokument eingefügt werden. <div> kann weggelassen werden.

echo-block.rnc: <head> und <p> haben direkte Gegenstücke in den DESpecs. <note> hat die Gegenstücke <mgl>, <mgr> und <fn>.

<s> hat kein Gegenstück in den DESpecs. Es muss aber in das Dokument eingefügt werden, um das Dokument valide zu machen.

- echo-content.rnc

definiert das Inline-Modell für Text. Der Text muss nicht weiter ausgezeichnet werden, um ein valides Dokument zu haben. Für die Elemente <reg> <sic> <set-off> <ref> <foreign> <emph>

echo-attribute.rnc

echo-datatype.rnc

definieren Standard-Attribute und -Datentypen. Sie werden von allen Modulen verwendet.

-> diejenigen Elemente auslagern, die zwar auf diese Level gehören, aber nur in Handarbeit eingefügt werden können, wie <set-off> ? Andererseits: auch <metadata> und <div> kommen nicht aus China. Aber die sind wichtig, im Gegensatz zu <set-off>.

echo-gap.rnc echo-de.rnc

entspricht ... im Schema

echo-figure.rnc

entspricht ... im Schema

echo-handwritten.rnc

entspricht ... im Schema

echo-milestone.rnc

entspricht ... im Schema

echo-float.rnc

echo-mathematics.rnc

echo-chinese-text.rnc


```
echo-gis.rnc
echo-textflows.rnc
echo-import-xhtml.rnc
echo-import-mathml.rnc
```

1.2.1 The File echo.rnc

```
# ECHO main file
#
# This RELAX NG compact schema ("ECHO Schema") defines the XML format
# of the MPDL project at the Max Planck Institute for the History of Science
# (see https://itgroup.mpiwg-berlin.mpg.de:8080/tracs/mpdl-project-content)
#
# authors: Wolfgang Schmidle, Malcolm D. Hyman
# Copyright © 2009, 2010 Max Planck Institute for the History of Science, Berlin, Germany
#
# The ECHO Schema is free software: you can redistribute it and/or modify
# it under the terms of the GNU General Public License as published by
# the Free Software Foundation, either version 3 of the License, or
# (at your option) any later version.
#
# This program is distributed in the hope that it will be useful,
# but WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
# GNU General Public License for more details.
#
# You should have received a copy of the GNU General Public License
# along with this program. If not, see <http://www.gnu.org/licenses/>.

default namespace echo = "http://www.mpiwg-berlin.mpg.de/ns/echo/1.0/"

# namespace for data entry structures that will be removed in the course of processing
namespace de = "http://www.mpiwg-berlin.mpg.de/ns/de/1.0/"

# standard W3C namespaces
namespace xhtml = "http://www.w3.org/1999/xhtml"
namespace xlink = "http://www.w3.org/1999/xlink"

# MathML
namespace mml = "http://www.w3.org/1998/Math/MathML"

# Dublin Core metadata
namespace dcterms = "http://purl.org/dc/terms"

namespace xsi = "http://www.w3.org/2001/XMLSchema-instance"

echo.version = "1.0RC"

# core modules
include "modules/echo-start.rnc"
include "modules/echo-metadata.rnc"
include "modules/echo-text.rnc"
include "modules/echo-block.rnc"
include "modules/echo-content.rnc"
include "modules/echo-attribute.rnc"
include "modules/echo-datatype.rnc"
```

```

# text structure
include "modules/echo-div.rnc"
include "modules/echo-milestone.rnc"

# additional text blocks: floats
include "modules/echo-float.rnc"
include "modules/echo-note.rnc"
include "modules/echo-import-xhtml.rnc"
include "modules/echo-figure.rnc"
include "modules/echo-handwritten.rnc"

# additional text blocks: non-floats
include "modules/echo-textflows.rnc"
include "modules/echo-block-scholarly.rnc"

# additions to the inline model
include "modules/echo-gap.rnc"
include "modules/echo-de.rnc"
include "modules/echo-mathematics.rnc"
include "modules/echo-import-mathml.rnc"
include "modules/echo-gis.rnc"
include "modules/echo-content-scholarly.rnc"

# additional attributes for Chinese text
include "modules/echo-chinese-text.rnc"

```

Offen: Das Schema enthält kein

`echo.unicode.version = attribute unicode-version { "5.2" }.`

Would that be relevant? If yes, where would it be it used? Version of raw text and xml separately?

1.3 echo-start

1.3.1 The File echo-start.rnc

```

# echo-start
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

start = echo

echo = element echo { echo.attlist, echo.content }
echo.attlist = attribute version { echo.version }
echo.content = echo.metadata, echo.text

```

1.3.2 <echo>

1.4 echo-metadata

1.4.1 The File echo-metadata.rnc

```

# echo-metadata
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

```

```

namespace dcterms = "http://purl.org/dc/terms"
namespace xsi      = "http://www.w3.org/2001/XMLSchema-instance"

echo.metadata = element metadata { empty, echo.metadata.content }
echo.metadata.content = dcterms.identifier
                      & dcterms.bibliogr
                      & dcterms.licensing
                      & echo.meta

# echo.metadata.content <-- dcterms:spatial, dcterms:temporal (in echo-gis)

dcterms.bibliogr = dcterms.title+ & dcterms.alternative*
                  & dcterms.creator+ & dcterms.contributor* & dcterms.publisher*
                  & dcterms.language+ & dcterms.date?
                  & dcterms.description*
dcterms.licensing = dcterms.rights* & dcterms.license* & dcterms.accessRights
                  & dcterms.rightsHolder* & dcterms.provenance*
                  & dcterms.dateCopyrighted?
echo.meta         = echo.font? & echo.font-family? & echo.stylesheet* & echo.echolink

dcterms.identifier = element dcterms:identifier { echo.identifier }

dcterms.title      = element dcterms:title { echo.language.attrib, text }
dcterms.alternative = element dcterms:alternative { echo.language.attrib, text }
dcterms.creator    = element dcterms:creator { echo.authorityfile?, text }
dcterms.contributor = element dcterms:contributor { echo.authorityfile?, text }
dcterms.publisher  = element dcterms:publisher { text }
dcterms.language   = element dcterms:language { dcterms.languageFormat }
dcterms.date       = element dcterms:date { dcterms.dateFormat }
dcterms.description = element dcterms:description { text }

dcterms.rights     = element dcterms:rights { text }
dcterms.license    = element dcterms:license { echo.license }
dcterms.accessRights = element dcterms:accessRights { echo.accessRights }
dcterms.rightsHolder = element dcterms:rightsHolder { text }
dcterms.provenance = element dcterms:provenance { text }
dcterms.dateCopyrighted = element dcterms:dateCopyrighted { dcterms.dateFormat }

echo.font          = element font { text }
echo.font-family   = element font-family { echo.font-families }
echo.stylesheet    = element stylesheet { echo.URI.datatype }
echo.echolink      = element echolink { echo.URI.datatype }

dcterms.languageFormat = attribute xsi:type { "dcterms:ISO639-3" }, text
dcterms.dateFormat     = attribute xsi:type { "dcterms:W3CDTF" }, text

echo.identifier      = text
echo.authorityfile  = attribute identifier { text }
echo.license         = text
echo.accessRights   = "free" | "restricted"
echo.font-families  = "serif" | "sans-serif" | "monospace"

# echo.font-families <-- "song style" (in echo-chinese-text)

```

There is no counterpart of the metadata in the DESpecs.

1.4.2 <dcterms:identifier>

1.4.3 <dcterms:title>, <dcterms:alternative>

<dcterms: alternative> refines <dcterms: title>.

1.4.4 <dcterms:creator>, <dcterms:contributor>

<dcterms: contributor> refines <dcterms: creator>.

1.4.5 <dcterms:publisher>

1.4.6 <dcterms:language>

1.4.7 <dcterms:date>

1.4.8 <dcterms:description>

1.4.9 <dcterms:rights>, <dcterms:license>, <dcterms:accessRights>

<dcterms: license> may be text or a URI.

<dcterms: license> and <dcterms: accessRights> refine <dcterms: rights>.

1.4.10 <dcterms:rightsHolder>

1.4.11 <dcterms:provenance>

1.4.12 <dcterms:dateCopyrighted>

<dcterms: dateCopyrighted> refines <dcterms: date>.

1.4.13 , <font-family>

echo.font-families <- "song style" in echo-chinese-text

1.4.14 <stylesheet>

xlink?

1.4.15 <echolink>

xlink?

1.5 echo-div

1.5.1 The File echo-div.rnc

```
# echo-div
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

echo.div = element div { echo.div.attlist, echo.div.content }
echo.div.attlist = echo.core.attrib, echo.level.attrib, echo.n.attrib,
                  echo.div.type, echo.div.type-free?
echo.div.content = echo.div.class+
echo.div.class = echo.div | echo.text.content-blocks.class

echo.div.type      = attribute type { text }
echo.div.type-free = attribute type-free { text }
echo.div.level     = attribute level { echo.count.datatype }

echo.text.logical-structure.class |= echo.div
```

In general, there is no counterpart for <text> or <div> in the DESpecs.

1.5.2 <text>

<text>: The default type is "free"

1.5.3 <div>

<ind> -> <div type="index">

<toc> -> <div type="toc">

Transcribed types:

if the type is in the standard list: type="definition" type-free="界"

if it is not in the standard list: type="other" type-free="界/definition"

1.6 echo-block

1.6.1 The File echo-block.rnc

```
# echo-block
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

echo.head = element head { echo.head.attlist, echo.head.content }
echo.head.attlist = echo.inline.attrib
echo.head.content = echo.inline.model
echo.text.content-blocks.class |= echo.head

# echo.head.attlist <-- type="ti" (in echo-chinese-text), echo.flow

echo.p = element p { echo.p.attlist, echo.p.content }
echo.p.attlist = echo.core.attrib, echo.p.type?
echo.p.content = echo.p.class+
echo.p.type = attribute type { "title" }
echo.p.class = echo.s
echo.text.content-blocks.class |= echo.p

# echo.p.attlist <-- echo.indent, echo.ics (in echo-chinese-text), echo.flow
# echo.p.class <-- echo.p-set-off, echo.quote, echo.milestone.class

echo.s = element s { echo.s.attlist, echo.s.content }
echo.s.attlist = echo.inline.attrib
echo.s.content = echo.inline.model

echo.blockquote = element quote { echo.blockquote.attlist, echo.blockquote.content }
echo.blockquote.attlist = echo.core.attrib, echo.delimiter.attrib?
echo.blockquote.content = echo.p+ | echo.p.content
echo.text.content-blocks.class |= echo.blockquote
```

Offen: can anchored comments be represented by echo.note? (probably yes)

1.6.2 <head>

<h> -> <head>

1.6.3 <s>

semantic units:

(<s> is not in the raw text)

Floating objects in <s> (all <note>, <handwritten>, <table>; most <figure>; some <math>) are replaced by <anchor> and moved to a <div type="float"> at end of the lowest-level <div> that contains the <s>, i.e. in particular after all <p> in the <div>.

Nein, sondern direkt hinter den Absatz!

The new <div type="float"> contains all <note>, <handwritten>, <figure>, <table>, <math> that have been moved in this <div>.

1.6.4 <note>

notes:

```
<mg1> -> <anchor type="note"/>, <note position="left">
```

```
<mgr> -> <anchor type="note"/>, <note position="right">
```

echo.note.content = echo.flexible.model to allow for different kinds of notes

1.6.5 <set-off>

1.7 echo-content

1.7.1 The File echo-content.rnc

```
# echo-content
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

echo.plaintext.model = echo.plaintext.class+
echo.plaintext.class = text

# echo.plaintext.model in: echo.reg, echo.sic, echo.num, echo.var
# echo.plaintext.class <-- echo.gap.class, echo.milestone.class, echo.image, echo.anchor

echo.inline.model = echo.inline.class+
echo.inline.class = echo.plaintext.class

# echo.inline.model in: echo.head, echo.s,
#                       echo.caption, echo.description, echo.variables
#                       echo.s-set-off, echo.ref, echo.foreign, echo.emph, echo.q
# echo.inline.class <-- echo.s-set-off, echo.ref, echo.foreign, echo.emph, echo.q, echo.reg, echo.sic,
#                       echo.num, echo.var, echo.gis.class, mml.math

echo.flexible.content = echo.text.content-blocks
                       | echo.p.content
                       | (echo.space.attrib, echo.inline.model)

echo.flexible.model   = echo.core.attrib, echo.flexible.content

# echo.flexible.content in: Inline.model, Flow.model (in echo-import-xhtml)
# echo.flexible.model   in: echo.note, echo.handwritten

echo.reg = element reg { echo.reg.attlist, echo.reg.content }
echo.reg.attlist = echo.core.attrib, echo.reg.orig, echo.reg.type?, echo.reg.responsible?
echo.reg.content = echo.plaintext.model
echo.reg.orig    = attribute orig { text }
echo.reg.responsible = attribute resp { text }
echo.reg.type    = attribute type { list { echo.reg.types+ } }
echo.reg.types  = "lig"
                 | "abbr" | "simple" | "context" | "wordlist" | "wordlist/context" | "unresolved"
                 | "typo" | "conjecture"
                 | "unknown"
echo.inline.class |= echo.reg
```

```

echo.emph = element emph { echo.emph.attlist, echo.emph.content }
echo.emph.attlist = echo.language.attrib?, echo.style-class.attrib, echo.id.attrib?
echo.emph.content = echo.inline.model
echo.inline.class |= echo.emph

# echo.emph.attlist <-- type="handwritten"

```

The tags `<emph>` (for italics), `<bf>`, `<sc>`, `<u>`, `<sup>`, ``, ``, `<st>`, `<red>`, `<sp>` in the Specs are normally represented by `<emph class="...">`, see "echo-stylesheet.css". The tags can be combined, e.g. `<emph class="it bf">` for bold italics. For whole `<s>` or `<p>`, the class attribute is there (or even higher in the hierarchy).

Conversely, most elements in this module have no counterpart in the Specs and will be added in the post-processing stage.

Only the original text is regularized using echo.reg; typing conventions and additional typos in the transcription are silently resolved.

list of typing conventions in the DESpecs which are silently resolved:

\$ -> f

\' q -> q + combining diacritic (U+0300 etc.) and normalization form C, for example ù

...

Offen: How do we represent Fraktur (`<fr>`)? `xml:lang="de-Latf"`, and if yes, where?

`<foreign xml:lang="de-Latf">`?

``?

1.7.2 `<reg>`

echo.reg.types: no type "other"!

examples:


```

<reg orig="ijs" type="lig">ijs</reg>
<reg orig="spherae" type="simple">sphaerae</reg>
<reg orig="spherae">sphaerae</reg>
<reg orig="sphē- rae" type="simple">sphae-  
<lb/>rae</reg>
<reg orig="eiufdē" type="context">eiufdem</reg>
<reg orig="efsēt" type="context">efsent</reg>
<reg orig="lib." type="context">liber</reg>
in <reg orig="lib." type="context">libro</reg>
<reg orig="qñ" type="wordlist">quando</reg>
<reg orig="tm̃" type="wordlist/context">tamen</reg>
<reg orig="tm̃" type="wordlist/context">tantum</reg>
<reg orig="Arist." type="unresolved">Arist.</reg>
<reg orig="inrerrogas" type="typo" resp="paul">interrogas</reg>
<reg orig="quem" type="conjecture" resp="paul">quam</reg>
<reg orig="re ferre" type="conjecture" resp="paul">re-  
<lb/>ferre</reg>
<reg orig="pp" type="unknown">pp</reg>
<reg orig="pp" type="conjecture" resp="paul">prope</reg>

```

note:

- the default type is "simple", e.g. `<reg orig="spherae">sphaerae</reg>`
- the first example `ijs` applies only if `ij` is not silently resolved
- missing hyphens are indicated by a soft hyphen "-" rather than `<reg>`; however, you may use "conjecture" in non-trivial cases
- the generic "abbr" may be used for any abbreviation
- abbreviations are not resolved within `<ref>`, e.g.
ex `<ref standard="Euclid: 5:19">.19. lib. quinti Eu-
<lb/>clid.</ref>`

1.7.3 `<ref>`

1.7.4 `<sic>`

mistakes in the original text:

`o<!>` -> `o<de:wrong/>` -> `<sic comment="n missing">o</sic>` (see the discussion in `echo-de`)

1.7.5 `<foreign>`

Foreign text is not be marked in the transcription, i.e. `<foreign>` cannot be inserted automatically without additional linguistic knowledge.

Exception: `<rom>sentence</rom>` -> `<foreign xml:lang="la">sentence</foreign>`
with language "la" as a first guess

(`echo.foreign` has `echo.core.attrib`, but `echo.language.attrib` is obligatory)

1.7.6 <emph>

emphasis (should be used only when something is not tagged otherwise)

1.7.7 <q>, <quote>

echo.q: short inline quotes. Note that echo.delimiter-attr is optional; however, please use it if possible

echo.quote: longer inline quotes (one-sentence quotes are <quote><s>Sentence.</s></quote> and not <s><q>Sentence.</q></s>)

echo.blockquote: blockquotes

quotes: avoid recursions:

q: contained in s; cannot contain p or s

quote in p: contained in p (but may be set-off); cannot contain p; on the same level as s; contains s+

quote in div: on the same level as p; contains p+ or s+

1.7.8 Avoiding Recursions

Wie ist die Inline-model-Hierarchie?

- inline anfangen können: s head cap desc vari, evtl. note handwritten xhtml
- in inline, und Inhalt inline (mit Rekursionsgefahr): s-set-off, ref, foreign, emph, q
- in inline, und Inhalt plaintext: reg, sic, num, var, place, time
- in inline, und Inhalt es selbst: mml.math
- in plaintext, und Inhalt plaintext: gibt es nicht
- in plaintext, Inhalt text: unsure
- in plaintext, leer: milestones, image, anchor, gap, unknown, wrong

-> Schematron-Regel, die Rekursionen aufspürt, d.h.

1. z.B. <ref> in <ref>

2. z.B. <ref> in <foreign> in ... in <ref>

also zusammen: kein Element aus dieser Gruppe darf sich selbst als ancestor haben.

1.8 echo-gap

1.8.1 The File echo-gap.rnc

```
# echo-gap
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".
```

```
echo.plaintext.class |= echo.gap.class
```

```
echo.gap = element gap { echo.gap.attlist, empty }
echo.gap.attlist = echo.extent?
echo.extent = attribute extent { echo.count.datatype }
echo.gap.class |= echo.gap
```

```
echo.undef = element undef { empty, echo.undef.content }
echo.undef.content = text
echo.gap.class |= echo.undef
```

```
@@ -> <gap extent="2"/>
```

```
<gap> -> <gap/>
```

x<?> -> x<undef/> or <undef>x</undef> (this can not be fully automated)

1.8.2 <gap>

1.8.3 <undef>

1.9 echo-de

1.9.1 The File echo-de.rnc

```
# de.rnc
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".
```

```
namespace de = "http://www.mpiwg-berlin.mpg.de/ns/de/1.0/"
```

```
echo.gap.class |= de.class
```

```
de.unknown = element de:unknown { de.unknown.attlist, empty }
de.unknown.attlist = attribute code { echo.count.datatype }
de.class |= de.unknown
```

```
de.wrong = element de:wrong { empty }
de.class |= de.wrong
```

de.class soll sich selbst in plaintext einhängen!? oder zusammen in ein Modul?
oder so lassen? immerhin haben wir uns vor einem Jahr etwas dabei gedacht

*

This module contains tags from the DESpecs that will be removed in the course of processing. We use the namespace "de" for the corresponding elements in the xml:

<001> -> <de:unknown code="001"/>

<!> -> <de:wrong/> -> remove or <sic>

1.9.2 <unknown>

1.9.3 <wrong>

1.10 echo-figure

1.10.1 The File echo-figure.rnc

```
# echo-figure
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

echo.figure = element figure { echo.figure.attlist, echo.figure.content }
echo.figure.attlist = echo.figure.position?,
                    echo.file.attrib?, echo.link.attrib?, echo.id.attrib?
echo.figure.content = echo.image*, echo.caption?, echo.description*, echo.variables?
echo.figure.position = attribute position { echo.figure.positions }
echo.figure.positions = "here"
echo.float.class |= echo.figure
echo.anchor.types |= "figure"

echo.caption = element caption { echo.caption.attlist, echo.caption.content }
echo.caption.attlist = echo.inline.attrib
echo.caption.content = echo.inline.model

echo.description = element description { echo.description.attlist, echo.description.content }
echo.description.attlist = echo.inline.attrib
echo.description.content = echo.inline.model

echo.variables = element variables { echo.variables.attlist, echo.variables.content }
echo.variables.attlist = echo.inline.attrib
echo.variables.content = echo.inline.model

echo.image = element image { echo.image.attlist, empty }
echo.image.attlist = echo.image.description?,
                  echo.file.attrib?, echo.link.attrib?, echo.id.attrib?
echo.image.description = attribute desc { text }
echo.plaintext.class |= echo.image
```

<fig> -> <anchor/>, <figure>

<cap> -> <caption>

<desc> -> <description>

<var> -> <variables>

<fig/> -> <image> (no direct counterpart of <image> in the specs)

1.10.2 <figure>, <caption>, <description>, <variables>

1.10.3 <image>

1.11 echo-handwritten

1.11.1 The File echo-handwritten.rnc

```
# echo-handwritten
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

echo.handwritten = element handwritten { echo.handwritten.attlist, echo.handwritten.content }
echo.handwritten.attlist = echo.link.attrib?
echo.handwritten.content = echo.flexible.model?
echo.float.class |= echo.handwritten
echo.anchor.types |= "handwritten"
echo.emph.attlist &= attribute type { "handwritten" }?
```

<hd> -> <anchor/>, <handwritten/>

In its simplest form, <handwritten> is just an empty tag. Nonetheless, within <s> it is replaced by <anchor> and moved to <div type="float"> to cater for scholarly additions, i.e. it is part of echo.float.class and not of echo.inline.class

1.11.2 <handwritten>

1.12 echo-float

1.12.1 The File echo-float.rnc

```
# echo-float
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

echo.float.class = notAllowed
echo.text.content-blocks.class |= echo.float.class

# echo.float.class <-- echo.note, echo.figure, echo.handwritten,
#                       xhtml.table, xhtml.list.class,
#                       mmml.math

echo.anchor = element anchor { echo.anchor.attlist, empty }
echo.anchor.attlist = attribute type { echo.anchor.types }, echo.symbol.attrib?, echo.link.attrib
echo.anchor.types = notAllowed
echo.plaintext.class |= echo.anchor

# echo.anchor.types <-- "note", "figure", "handwritten",
#                       "table", "ul", "ol", "dl",
#                       "math"

echo.float-div = element div { echo.float-div.attlist, echo.float-div.content }
echo.float-div.attlist = echo.core.attrib, echo.level.attrib, echo.n.attrib,
```

```

                                echo.float-div.type
echo.float-div.content = echo.float.class+
echo.float-div.type    = attribute type { "float" }
echo.text.content-blocks.class |= echo.float-div
#echo.p.class |= echo.float-div
#echo.p.class |= echo.float.class

```

Liste aller div-Typen, die auf besondere Weise behandelt werden:

u.a. float

aber auch multiframe, parallel; chapter, section; ...

1.12.2 <anchor>

Verschieben der Floats in <s>: Möglichst einheitlich bleiben, also entweder immer nach <s> oder immer nach <p>. Beides erlaubt. Besser wäre nach <p>.

Erstellung des XML-Textes: Erst <s> markieren, erst dann floats verschieben. Oder erst hinter den <p>, und wenn die <s> erstellt sind, kann man das <div float> einfach an die richtige Stelle schieben.

1.13 echo-milestone

1.13.1 The File echo-milestone.rnc

```

# echo-milestone
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

echo.milestone.class = notAllowed
echo.p.class         |= echo.milestone.class
echo.plaintext.class |= echo.milestone.class

echo.lb = element lb { echo.lb.attlist, empty }
echo.lb.attlist = empty
echo.milestone.class |= echo.lb

# echo.lb.attlist <-- type="halflines" (in echo-chinese-text),
#                               echo.label (in echo-textflows)

echo.cb = element cb { echo.cb.attlist, empty }
echo.cb.attlist = echo.cb.thead?
echo.cb.thead = attribute thead { text }
echo.milestone.class |= echo.cb
echo.text.content-blocks.class |= echo.cb

# echo.cb.attlist <-- echo.label (in echo-textflows)

echo.pb = element pb { echo.pb.attlist, empty }
echo.pb.attlist = echo.n.attrib, echo.pb.o?, echo.pb.o-norm?,
                  echo.pb.rthead?,
                  echo.file.attrib?, echo.link.attrib?
echo.pb.o      = attribute o { text }
echo.pb.o-norm = attribute o-norm { text }

```

```

echo.pb.rhead = attribute rhead { text }
                | attribute chead { text }
echo.milestone.class |= echo.pb
echo.text.content-blocks.class |= echo.pb

# echo.pb.attlist <-- ics="yes", echo.flow

```

1.13.2 <lb>

line breaks: <lb/> can be in plaintext (<s>, <head>, some <note>, all members of echo.inline.class) and <p>

in <p>: since a paragraph is split into <s>, most line breaks are actually in <s>. However:

- <lb/></s><s> and </s><s><lb/> shouldn't occur (-> </s><lb/><s> [and space before </s>?])

- <lb/></s></p> shouldn't occur at all

in <s> (and similarly for <head> and the members of echo.inline.class):

- line break -> <lb/>; no space before <lb/>; no line break after <lb/>; space after <lb/> if there is a hyphen before <lb/> (no automated space if the hyphen is missing)

examples:

```
<s>□<lb/>□□□□□□</s>
```

```
<note>Plutar <lb/>chus in <lb/>commẽ <lb/>tario de <lb/>dæmo-<lb/>nio So-
<lb/>cratis.</note>
```

-> das ist sicher veraltet!

We use the normal hyphen U+002D instead of the soft hyphen U+00AD because the soft hyphen is not displayed in the xhtml.

1.13.3 <cb>

column breaks:

```
<col 1>...</col><col 2>...</col> -> ...<cb/>...
```

1.13.4 <pb>

page breaks: <pb/> can occur wherever <lb/> occurs (although it will be rare in <head>), and <div>

```
<pb vii><rh>xyz</rh> -> <pb n="10" o="vii" o-norm="7" recho.head="xyz"
xlink:href="URI"/>
```

```
<pb 一六七a> -> <pb n="..." o="一六七a" o-norm="167a" xlink:href="URI"/>
```

(echo.pb.type = attribute type { text } hat sich mit flow erledigt)

1.14 echo-attribute

1.14.1 The File echo-attribute.rnc

```
# echo-attribute
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

namespace xlink = "http://www.w3.org/1999/xlink"

echo.core.attrib      = echo.language.attrib?, echo.style-class.attrib?, echo.id.attrib?
echo.inline.attrib   = echo.core.attrib, echo.space.attrib

echo.language.attrib = attribute xml:lang { echo.language.datatype }
echo.style-class.attrib = echo.style.attrib
                        | echo.class.attrib
echo.style.attrib     = attribute style { echo.css.datatype }
echo.class.attrib     = attribute class { echo.class.datatype }
echo.id.attrib        = attribute xml:id { echo.id.datatype }
echo.space.attrib     = attribute xml:space { "preserve" }

echo.n.attrib         = attribute n { echo.count.datatype }
echo.level.attrib    = attribute level { echo.count.datatype }
echo.symbol.attrib    = attribute symbol { text }

echo.file.attrib     = attribute file { text }
echo.link.attrib     = attribute xlink:href { xsd:anyURI },
                        attribute xlink:label { xsd:anyURI }?,
                        attribute xlink:type { text }?
echo.delimiter.attrib = attribute open { text },
                        attribute close { text }
```

1.15 echo-datatype

1.15.1 The File echo-datatype.rnc

```
# echo-datatype
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

echo.language.datatype = text
echo.css.datatype      = text
```



```

echo.class.datatype      = xsd:NMTOKENS
echo.id.datatype         = xsd:NCName
echo.URI.datatype        = xsd:anyURI
echo.integer.datatype    = xsd:integer
echo.real.datatype       = xsd:float
echo.count.datatype      = xsd:nonNegativeInteger
echo.measure.datatype    = text

```

1.16 echo-mathematics

1.16.1 The File echo-mathematics.rnc

```

# echo-mathematics
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

echo.num = element num { echo.num.attlist, echo.num.content }
echo.num.attlist = echo.num.value, echo.style-class.attrib?
echo.num.content = echo.plaintext.model
echo.num.value = attribute value { echo.real.datatype }
echo.inline.class |= echo.num

echo.var = element var { echo.var.attlist, echo.var.content }
echo.var.attlist = echo.style-class.attrib?, attribute type { echo.var.types }?
echo.var.content = echo.plaintext.model
echo.inline.class |= echo.var

echo.var.types = (
  "point" | "line" | "area"
| "angle"
| "triangle" | "square" | "circle" | "rectangle" | "polygon"
| "number"
| "gnomon"
| "other" )

```

entfernt:

```

echo.math.types = "math:definition" | "math:axiom" | "math:postulate" | "math:proposition"
| "math:theorem" | "math:lemma" | "math:corollary" | "math:proof" | "math:example"
| "math:exercise"

```

```

echo.body-div.types |= echo.math.types

```

*

Pseudo-namespaces in order to avoid conflicts with other modules, even though the schema would still be valid.

Note: The scope of echo.num and echo.var is very limited. More complex mathematics is expressed with MathML.

*

echo.mathematics.class ? oder verwirrt das nur, wenn die Elemente nicht explizit genannt werden?

Man könnte andersrum sogar sagen, dass `<num>` und `<var>` wieder nach echo-content sollten, denn der Hauptzweck von echo-mathematics war, die echo.math.types zu definieren.

1.16.2 `<num>`

number:

"vii" -> `<num value="7">vii</num>`

"½" -> `<num value="0.5">½</num>`

1.16.3 `<var>`

variable:

"AB" -> `<var type="line">AB</var>`

line ist optional

1.17 echo-chinese-text

1.17.1 The File echo-chinese-text.rnc

```
# echo-chinese-text
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".
```

```
echo.font-families |= "song style"
```

```
echo.head.attlist &= attribute type { "ti" }?
```

```
echo.indent = attribute indent { echo.measure.datatype }
echo.p.attlist &= echo.indent?
```

```
echo.ics = attribute ics { "yes" }
echo.p.attlist &= echo.ics?
echo.pb.attlist &= echo.ics?
```

```
echo.lb.attlist &= attribute type { "halflines" }?
```

```
echo.font-families |= "song style"
```

etc.; Ask Martina to provide a list. Problem: der Unterschied font/font-families ist nicht völlig klar, bzw. es sind wirklich Familien gemeint im Gegensatz zu font-family in html.

*

```
<ti> -> <head type="ti">
```

*

indentations in Chinese text:

<p ii> -> <p indent="2char">

<p xx> -> <p indent="-2char">

(indent is deliberately not defined as style="valid css" because it may be semantically meaningful)

*

paragraphs or pages in "individualized character style":

Or should the post-processing move "ics" from <pb> to all <h>, <h type="ti">, <p> on that page? -> wahrscheinlich. Entscheide beim ersten Text, wo es relevant ist.

*

<sm> in Chinese text:

in <p>: <sm> -> <anchor type="note"/>, <note>

<emph style="sm"> (everywhere else: <h>, rhead, <ti>, <toc>, etc.)

(plus some manual corrections where this simple distinction doesn't fit, e.g. <sm>chen</sm>)

(If a note is spread over more than one page in the original text, all parts will be moved to the first page)

\bs\bs -> <lb type="halfline"/>

*

<sl> -> <emph class="sl">

<dl> -> <emph class="dl">

<wl> -> <emph class="wl">

<cl> -> <emph class="cl">

*

Find out whether <list> in Chinese text can simply be represented by <xhtml:ol> etc. (unlike <toc> and <index>, <list> is not a semantic category, i.e. it does not need a <div>).

1.18 echo-gis

1.18.1 The File echo-gis.rnc

```
# echo-gis
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

# Note: this module is experimental.

namespace dcterms = "http://purl.org/dc/terms"
namespace xsi      = "http://www.w3.org/2001/XMLSchema-instance"

echo.metadata.content &= echo.spatial* & echo.temporal*

echo.spatial = element dcterms:spatial { empty, echo.spatial.content }
echo.spatial.content = dcterms.pointFormat
                    | dcterms.boxFormat
dcterms.pointFormat = attribute xsi:type { "dcterms:Point" }, text
dcterms.boxFormat   = attribute xsi:type { "dcterms:Box" }, text

echo.temporal = element dcterms:temporal { empty, echo.temporal.content }
echo.temporal.content = dcterms.dateFormat
                    | dcterms.periodFormat
dcterms.periodFormat = attribute xsi:type { "dcterms:Period" }, text

echo.gis.class = notAllowed
echo.inline.class |= echo.gis.class

echo.place = element place { echo.place.attlist, echo.place.content }
echo.place.attlist = echo.place.standard?, echo.place.type?, echo.place.position?,
                    echo.link.attrib?, echo.style-class.attrib?
echo.place.content = echo.plaintext.model
echo.place.standard = attribute standard { text }
echo.place.type = attribute type { echo.place.types }
echo.place.types = "city"
                 | "river"
echo.place.position = attribute point { text } | attribute box { text }
echo.gis.class |= echo.place

echo.time = element time { echo.time.attlist, echo.time.content }
echo.time.attlist = echo.time.standard?,
                    echo.style-class.attrib?
echo.time.content = echo.plaintext.model
echo.time.standard = attribute w3cdtf { text } | attribute period { text }
echo.gis.class |= echo.time
```

Note: this module is experimental.

1.18.2 Metadata

(dcterms.dateFormat is defined in echo-metadata)

spatial/temporal: encode e.g. "Song Dynasty"?

Zusammenhang mit authority files klären

1.18.3 <place>

```
echo.place.types = "city" | "river"
```

-> etc. etc., oder alternativ ganz weglassen. Wieder: authority file, das diese Info hat?

1.18.4 <time>

"standard" means different things in echo.place and echo.time!?

1.18.5 <person>, <event>

Es wird auch noch <person> und <event> geben, aber die Details sind noch unklar.

1.19 echo-textflows

1.19.1 The File echo-textflows.rnc

```
# echo-textflows
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

echo.multiflow-div = element div { echo.multiflow-div.attlist, echo.multiflow-div.content }
echo.multiflow-div.attlist = echo.core.attrib, echo.level.attrib, echo.n.attrib,
                           echo.multiflow-div.type, echo.multiflow-div.flows
echo.multiflow-div.content = echo.multiflow-div.content.class+
echo.multiflow-div.content.class = echo.text.content-blocks.class | echo.parallel-div
echo.multiflow-div.type = attribute type { "multiflow" }
echo.multiflow-div.flows = attribute flows { echo.integer.datatype }
echo.text.content-blocks.class |= echo.multiflow-div

echo.flow = attribute flow { text }
echo.pb.attlist &= echo.flow?
echo.p.attlist &= echo.flow?
echo.head.attlist &= echo.flow?

echo.label = attribute label { text }
echo.lb.attlist &= echo.label?
echo.cb.attlist &= echo.label?

echo.parallel-div = element div { echo.parallel-div.attlist, echo.parallel-div.content }
echo.parallel-div.attlist = echo.core.attrib, echo.level.attrib, echo.n.attrib,
                           echo.parallel-div.type
echo.parallel-div.content = echo.text.content-blocks
echo.parallel-div.type = attribute type { "parallel" }
```

vor der Verschiebung der div-Typen in das Schematron:

```
echo.body.types |= "multiflow" | "parallel"
```

```
echo.body-div.attlist &= echo.flows?
```

wieder: Liste der div-Typen, die besonders behandelt werden.

*

```
@flow is normally a number, or "footnote"
```

1.20 echo-import-xhtml

1.20.1 The File echo-import-xhtml.rnc

```
# echo-import-xhtml
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

default namespace xhtml = "http://www.w3.org/1999/xhtml"

Inline.model = echo.flexible.content
Flow.model   = echo.flexible.content

include "../xhtml/xhtml-basic-table.rnc"
xhtml.table = table
table.attlist &= echo.link.attrib?
echo.float.class |= xhtml.table
echo.anchor.types |= "table"

include "../xhtml/xhtml-list.rnc"
xhtml.list.class = ul | ol | dl
echo.float.class |= xhtml.list.class
echo.anchor.types |= "ul" | "ol" | "dl"

include "../xhtml/xhtml-attribs.rnc"
Common.attrib &= echo.id.attrib?

include "../xhtml/xhtml-datatypes.rnc"
```

1.20.2 xhtml-basic-table

We ignore Block.class in xhtml-basic-table: The following lines replace

"Block.class |= table" in xhtml-basic-table

```
echo.float.class |= xhtml.table
echo.anchor.types |= "table"
```

*

```
include "../xhtml/xhtml-basic-table.rnc"
```

-> we don't need the additions in xhtml-table

The attribute `border` is defined in `xhtml-table`. Is this a reason to include `xhtml-table` rather than `xhtml-basic-table`?

1.20.3 xhtml-list

We ignore `Block.class` in `xhtml-list`: The following lines replace

```
"List.class = ul | ol | dl",
```

```
"Block.class |= List.class"
```

```
echo.float.class |= xhtml.ul | xhtml.ol | xhtml.dl  
echo.anchor.types |= "ul" | "ol" | "dl"
```

*

Note that `xhtml:ul`, `xhtml:ol`, `xhtml:dl` do not correspond to ``, ``, `<dl>` in the DESpecs!

-> extract a list of tags and their xhtml counterparts from "echo-stylesheet.xml"

1.20.4 xhtml-attrs

We use the following xhtml elements: `table`, `caption`, `tr`, `th`, `td`; `dl`, `dt`, `dd`; `ol`, `ul`, `li`. All these elements have `Common.attrib` (`th` and `td` via `Cell.attrib`). `xml:lang` and `class` are already in `Common.attrib`, and we add `xml:id { xsd:NCName }` to it (however, `Common.attrib` already includes the attribute `id { xsd:ID }`)

1.20.5 xhtml-datatypes

1.20.6 xhtml-text

Dieses Modul wird nicht importiert. Stattdessen werden `Inline.model` und `Flow.model` durch `echo.flexible.content` ersetzt:

```
Inline.model = echo.flexible.content
```

```
Flow.model = echo.flexible.content
```

ursprüngliche Definitionen:

```
Inline.model = (text | Inline.class)*
```

Block.mix = Block.class

Block.model = Block.mix+

Flow.model = (text | Inline.class | Block.class)*

1.20.7 <xhtml:*>

1.21 echo-import-mathml

1.21.1 The File echo-import-mathml.rnc

```
# echo-import-MathML
#
# This file is part of the ECHO Schema. See the main file "echo.rnc".

default namespace mml = "http://www.w3.org/1998/Math/MathML"

mml.math = element math { mml.math.attlist, mml.math.content }
mml.math.attlist = attribute mode { "display" }?, attribute display { "inline" }?
mml.math.content = element mml:* { (attribute * { text } | text | mml.math.content)* }

echo.inline.class    |= mml.math
echo.p-set-off.class |= mml.math
echo.float.class     |= mml.math
echo.anchor.types    |= "math"
```

Note: Simple mathematical terms, i.e. numbers and variables, are marked using echo.num and echo.var (defined in echo-mathematics).

*

mml.math.content erlaubt beliebige Elemente <mml: *> innerhalb von <mml: math>.

This placeholder code above is good enough for the moment. We simply assume that the MathML parts are well-formed. This is plausible since the MathML code is created from a LaTeX formula by a MathML-converter.

In addition, Oxygen uses a separate validation engine for MathML.

1.21.2 <mml:*>

2 Offene Punkte

Zielgruppe?

2.0.3 , <font-family>

2.1 echo-block

2.1.1 <head>

Markierung von Überschriften

2.1.2 <p>

2.1.3 <s>

2.1.4 <note>

2.1.5 <p-set-off>

p+ ist sinnvoll; trotzdem echo.div.content, um das Schema einheitlicher zu machen?

2.2 echo-content

flexible.model: inlinemodel statt s.content, und genauso bei p-setoff oder so, aber nicht xml:space zu inlinemodel dazu wegen ref etc.

2.2.1 <reg>

2.2.2 <ref>

2.2.3 <sic>

2.2.4 <foreign>

2.2.5 <emph>

2.2.6 <q>

2.2.7 <quote>

2.3 echo-gap

x

2.3.1 <gap>

2.3.2 <unsure>

2.4 echo-de

2.4.1 <unknown>

2.4.2 <wrong>

2.5 echo-figure

x

2.5.1 <figure>, <caption>, <description>, <variables>

2.5.2 <image>

2.6 echo-handwritten

x

2.6.1 <handwritten>

2.7 echo-float

2.7.1 <anchor>

Wenn ich es recht verstehe, ist es bei Arboreal, wie es jetzt ist, nicht klar, ob man Marginalien an der entsprechenden Stelle sehen möchte oder nicht.

Ist es möglich, Arboreal dazu zu bekommen, bestimmte Floats wie z.B. Marginalien dort anzuzeigen, wo ihr anchor ist?

Ansonsten gibt es mehrere Möglichkeiten: - Verschiebe Marginalien weiterhin aus dem <s> hinter den <p> und akzeptiere, dass Peter unzufrieden ist. - Lasse Marginalien immer in <s> und akzeptiere, dass Arboreal den Text dann eventuell schlecht lesbar anzeigt, und dass die Text-Datei dadurch schlecht lesbar wird. - Verschiebe die Marginalie nur so viel, dass sie auf der richtigen Seite bleibt. Vorschlag war, sie hinter das <s> zu schieben.

Soll man das auch von der Länge der Marginalie abhängig machen?

Mit der dritten Möglichkeit könnte ich vermutlich leben. Frage ist dann, ob man dann trotzdem einen `<anchor>` vergibt, und ob man ein `<div float>` einführt, was die Lesbarkeit wieder verschlechtern würde.

Außerdem ist damit das Problem nur halb gelöst, denn die `<note>` kann damit trotzdem noch auf die nächste Seite rutschen. Man müsste also einen Algorithmus haben: - Wenn es keinen `<pb>` in dem `<s>` gibt, verschiebe alle Marginalien hinter den `<s>` (auch wenn es der letzte `<s>` im `<p>` ist). - Wenn es einen `<pb>` gibt, verschiebe alle Marginalien vor dem `<pb>` vor das `<s>` und alle Marginalien hinter dem `<pb>` hinter das `<s>`. - Wenn es zwei `<pb>` gibt: Der Fall ist sehr selten und muss dann von Hand entschieden werden.

*

Unterschied machen zwischen `<note>` und zum Beispiel `<figure>`? Oder nicht?

Wird die Datei schlechter lesbar, wenn man hinter `<s>` tut statt hinter `<p>`?

Paul: keinen komplizierten Algorithmus. Einfach hinter `<s>`. Wenn es dann hinter einen `<pb>` rutscht, ist das eben so.

2.8 echo-milestone

x

2.8.1 `<lb>`

2.8.2 `<cb>`

2.8.3 `<pb>`

2.9 echo-attribute

x

2.10 echo-datatype

x

2.11 echo-mathematics

x

2.11.1 <num>

2.11.2 <var>

2.12 echo-chinese-text

x

2.13 echo-gis

x

2.13.1 <place>

2.13.2 <time>

2.13.3 <person>, <event>

2.14 echo-textflows

x

2.15 echo-import-xhtml

x

2.15.1 <xhtml:*>

2.16 echo-import-mathml

x

2.16.1 <mml:*>

3 Wunschliste für die Darstellung der Elemente

3.1 echo

-

3.2 echo-start

3.2.1 <echo>

Es sollten die zu @version passenden XSLT-Skripte verwendet werden.

3.3 echo-metadata

Die Metadaten-Suche sollte die hierarchische Struktur der dcterms-Metadaten berücksichtigen: Wenn "A refines B", dann findet eine Such nach einem Inhalt von B auch Einträge in A finden, aber nicht umgekehrt.

3.3.1 <dcterms:creator>

@identifier: Die Einträge sollten in links umgewandelt werden.

3.3.2 <dcterms:language>

Alle angegebenen Sprachen sind Haupt-Sprachen des Textes. Text in einer der Haupt-Sprachen sollte korrekt mit Donatus/Pollux analysiert und indiziert werden. Textabschnitte in anderen Sprachen müssen nicht unbedingt morphologisch analysiert werden.

3.3.3 <dcterms:accessRights>

Texte mit <dcterms:accessRights>restricted</dcterms:accessRights> sollten wohl nicht in eXist aufgenommen werden.

3.3.4 , <font-family>

Umsetzen als Anweisung für den Browser? Beispiel "Fraktur". Noch nicht ausgereift.

3.4 echo-div

3.4.1 <text>

@xml: lang siehe <dcterms: language> (section 3.3.2)

3.4.2 <div>

Erzeuge ein TOC aus den <div>. Berücksichtige dabei nur die <div> mit folgenden Werten von @type: volume, front, body, back, chapter, section, ...

```
body
  chapter 1: Kapitel-Überschrift
    section 1.1: Abschnitt-Überschrift
      section 1.1.1: Unterabschnitt-Überschrift
```

Alternativ kann man auch eine Liste von div-Typen ausschließen:
multiflow, parallel, other, ...

Wahrscheinlich ist der opt-in-Ansatz aber besser als der opt-out-Ansatz. Als grobe Orientierung könnte jedenfalls dienen, dass Typen, die in anderen Modulen als echo-div definiert sind, nicht im TOC angezeigt werden.

Das TOC sollte Seitennummern sowie links zu den entsprechenden Abschnitten haben.

(Für die Numerierung und Seitennummern habe ich ein XSL-Skript. Das kann ich noch um das Einfügen der Überschrift erweitern. Schematron stellt sicher, dass alle <div chapter> und <div section> wirklich mit einer Überschrift anfangen.)

Zusammenspiel mit eSciDoc-TOC ?

3.5 echo-block

3.5.1 <head>

CSS: fett

(<pb> in <head> sollten nur innerhalb von Flows vorkommen; siehe echo-textflows)

3.5.2 <s>

Kann ignoriert werden. Die Attribute sollen natürlich beachtet werden. (Ausnahme sind eventuell <s> mit implizitem <lb/> in <div lines>)

xml: <s> in einer Zeile lassen, Fließtext, wird vom Browser umgebrochen

3.5.3 <note>

siehe echo-float (section 3.11)

3.5.4 <set-off>

sollte in eigenen Zeilen und eingerückt dargestellt werden

3.6 echo-content

3.6.1 <reg>

In der Standard-Anzeige wird der Inhalt des reg-Tags angezeigt. Der User kann aber wählen, ob er stattdessen jeweils den Inhalt des orig-Attributs angezeigt bekommen möchte. Zusätzlich sollte es eine dritte Version geben, nämlich wo die Wortformen noch orthographisch normalisiert sind. Diese Darstellung ist sprachspezifisch; im Lateinischen sind das u/v, i/j, f/s, ß/ss, æ/ae, œ/oe. Beispiel:

XML:	totum	<reg orig="relatiuũ" type="context">relatium</reg>	ipfius
Standard:	totum	relatium	ipfius
Original:	totum	relatiuũ	ipfius
normalisiert:	totum	relativum	ipsius

In der Original-Darstellung kann es Zeichen geben, die nur mit speziellen Fonts darstellbar sind. Zeichen, die in Unicode enthalten sind, werden normalerweise vom Browser korrekt angezeigt, wenn es irgendeinen Font auf dem Computer gibt, der das Zeichen enthält. Manchmal muss man Zeichen aus der Private Use Area nehmen, zum Beispiel U+E8BF für die Ligatur q̅ von q und 3 (U+A76B). Bei dieser Darstellung sollte der User also darüber aufgeklärt werden, dass er eventuell spezielle Fonts braucht.

3.6.2 <ref>

CSS: z.B. kursiv

Innerhalb von <ref> werden Abkürzungen nicht aufgelöst. Innerhalb von <ref> werden daher normale Wörter morphologisch analysiert (d.h. an Donatus/Pollux weitergeleitet), aber Wörter mit Punkt nicht. Zum Beispiel wird "lib." in

ex <ref>.7. propofitione lib. Archimedis de infidentibus aquæ</ref>

nicht analysiert.

Wenn es in `<ref>` ein `@standard` gibt, kann der ganze Inhalt unanalysiert bleiben, und stattdessen gibt es einen `link`, der zu der reference passt. Beispiel:

```
... ex <ref standard="Eucl:11:1">prima <lb/>vndecimi Euclid.</ref> pofitæ erunt.
```

Frage ist, wohin der link führen soll.

3.6.3 `<sic>`

kann hier ignoriert werden; wird in Arboreal gebraucht

3.6.4 `<foreign>`

CSS: Der Übergang von Fraktur zu Antiqua oder von A. zu F. sollte durch einen passenden Font ausgedrückt werden. `@xml:lang` sollte ausgewertet werden.

3.6.5 `<emph>`

`@style` sollte als `@class` und ein CSS-Stylesheet an den Browser weitergegeben werden. `@class` sollte nur in Ausnahmefällen vorkommen, z.B. in xhtml. Falls es im Text vorkommt, kann es direkt weitergegeben werden.

3.6.6 `<q>`

CSS: Sollte optisch erkennbar sein.

Für kurze quotes innerhalb eines `<s>`. Beachte `@open` und `@close`.

3.6.7 `<quote>`

ähnlich wie `<q>`; als blockquote wie `<set-off>`

3.7 echo-gap

3.7.1 `<gap>`

Darstellung als [...] ?

3.7.2 <unsure>

Ignorieren?

3.8 echo-de

3.8.1 <unknown>

Ignorieren? Sollte aber möglichst nicht mehr im Text enthalten sein.

3.8.2 <wrong>

Ignorieren? Sollte aber möglichst nicht mehr im Text enthalten sein.

3.9 echo-figure

3.9.1 <figure>, <caption>, <description>, <variables>

Figures, die nicht in <div float> sind, sollte auf alle Fälle an der richtigen Stelle angezeigt werden. Ansonsten siehe Floats (section 3.11). Beispiel:

```
<figure>
  <caption>Geometrische Formen</caption>
  <despription>Kreis</description>
  <despription>Dreieck</description>
  <despription>Viereck</description>
  <variables>a, b, c</variables>
</figure>
```

wird zu:

[figure 156]

Geometrische Formen

Kreis

Dreieck

Viereck

a, b, c

Beachte: "[figure 156]" und nicht "Figure 156 [remark: ...]"

3.9.2 <image>

Falls es @xlink:href gibt, sollte das Bild an dieser Stelle angezeigt werden.

3.10 echo-handwritten

3.10.1 <handwritten>

gehört nicht zum Textflow und wird erst auf Anforderung gezeigt. Meistens ist <handwritten> leer.

3.11 echo-float

3.11.1 <anchor>

Wenn es @symbol gibt, sollte das Symbol angezeigt werden, denn es steht ja im Text. Hochgestellt, wenn kein space davor ist, sonst nicht hochgestellt, aber kursiv. Beispiel:

Anmerkung² und anchor *a* eines anchored comment

Wenn es auch @xlink:href gibt, sollte das Symbol ein link zum zugehörigen Float sein.

Floats sind entweder als Teil des normalen Textflows, oder sie sind getrennt. Floats als Teil des normalen Textflows sollten, wie Bilder auf einer Webseite, an der richtigen Stelle angezeigt werden, d.h. dort, wo ihr <anchor> ist. Getrennte Floats sollten erst angezeigt werden, wenn der User sie anfordert.

normaler Textflow:

- <figure>
- <xhtml: *> (Tabellen, Listen)
- <mml: *> (MathML)

getrennt:

- <note> mit @position=foot (zumindest sollten Fußnoten, die über mehr als eine Seite gehen, auf Wunsch vollständig angezeigt werden)
- <handwritten>
- <note> mit @position=end (sollten bei ihrem <div endnotes> angezeigt werden, d.h. auf der Seite, auf der sie sind)

unklar:

- <note> mit @position=left/right: Peter hätte sie gerne im normalen Textflow

3.12 echo-milestone

3.12.1 <lb>

Zeilenumbruch: <lb/> ist so weit wie möglich/sinnvoll im XML-Baum verschwunden, also zum Beispiel:

```
<s>Bla Bla. <lb/><s>
```

statt

```
<s>Bla Bla. <s>  
<lb/>
```

Bei der Anzeige des XML-Textes für eine Seite sollte <lb/> innerhalb von <s> nicht in einer eigenen Zeile stehen. In anderen Worten: <lb> sollte nur dann in eigener Zeile stehen, wenn es auch im XML-Text in eigener Zeile steht.

3.12.2 <cb>

Im Prinzip wie <lb>. Aus dem <cb> ist nicht mehr automatisch rekonstruierbar, wo die erste Spalte anfängt und die letzte Spalte aufhört. Daher kann man es wie <lb> behandeln. Allerdings kann eine Spalte einen "running head" haben, der dargestellt wird wie ein running head einer Seite.

3.12.3 <pb>

CSS: Optisch klar vom "echten" Text abgrenzen: gesperrt?

Werden running heads dargestellt? Oder kann der User sie ein- und ausblenden?

Worttrennungen: Das System muss umgehen können mit Fällen wie

```
eue-<lb/>  
<anchor type="figure" xlink:href="..." /><lb/>  
<anchor type="figure" xlink:href="..." />  
<pb ... />  
niet
```

3.13 echo-attribute

`@xml:lang` siehe `<dcterms:language>` und `<foreign>`. Druch `@xml:lang` ist im Prinzip für jeden Textabschnitt klar, welche Sprache es ist, insbesondere welche Schreibrichtung es ist. Zeigt der Browser zum Beispiel Arabisch problemlos korrekt an, oder muss man dem Browser auf der Grundlage von `@xml:lang` explizit sagen, was er tun soll?

`@style` und `@class` siehe `<emph>` (section 3.6.5)

`@xml:id` kann ignoriert werden, wird aber für die links im TOC gebraucht. Floats verwenden `@xlink:label` statt `@xml:id`.

`@xml:space` sollte für die Ausgabe des XML-Textes beachtet werden.

3.14 echo-datatype

-

3.15 echo-mathematics

Ob div-Typen wie "math:definition" im TOC angezeigt werden sollen, ist unklar.

3.15.1 <num>

Der Inhalt von `<num>` wird nicht morphologisch analysiert. (Wird er in den Index aufgenommen?)

Falls es `@value` gibt, könnte man dem User anbieten, `@value` statt dem Inhalt anzuzeigen, also zum Beispiel "0.5" statt "1/2".

3.15.2 <var>

Der Inhalt von `<var>` wird nicht morphologisch analysiert und nicht in den Index aufgenommen.

Außerdem sollte `@type` berücksichtigt werden, wenn vorhanden. Beispielsweise könnte

```
<type="segment">. A. B. </var>
```

mit einer Linie über dem Namen dargestellt werden.

xml: auch `<var>` nicht in eigener Zeile anzeigen! sondern `<s>` in einer Zeile lassen, Fließtext, wird vom Browser umgebrochen

3.16 echo-chinese-text

Senkrechte Schrift im Original ist im XML bereits waagrecht und wird daher auch waagrecht angezeigt.

`<head>` mit `@ti`: CSS: z.B. in größerer Schrift

`<p>` und `<pb>` mit `@ics`: CSS: z.B. kursiv (allerdings sehen kursive chinesische Zeichen etwas merkwürdig aus)

`<p>` mit `@indent`: Chinesische Schrift ist monospaced (d.h. hat eine feste Laufweite). Rücke die erste Zeile um entsprechen viele Zeichen ein. Wie stark die weiteren Zeilen eingerückt werden, ist unklar. Die wichtige Information war die Einrückung der ersten Zeile, und `@indent` bezieht sich deshalb auf die erste Zeile eines Absatzes. Die restlichen Zeilen können im Original entweder genauso eingerückt sein oder auch nicht.

1. rücke nur die erste Zeile ein
2. rücke alle Zeilen des Absatzes gleich weit ein
3. erste Zeile eingerückt, alle Zeilen rechtsbündig passend, letzte Zeile linksbündig mit der Zeile davor.

Beachte: Der Wert von `@indent` kann auch negativ sein. Die Grundlinie sollte auf allen Seiten gleich angezeigt werden.

Notes: Die Entscheidung, ob eine Note (d.h. kleinere Schrift im Text) zum Textflow dazugehört oder nicht, ist im XML bereits gefallen. Dementsprechend gibt es `<anchor>` und `<note>` (mit `type="inline"`, der nicht angibt, wo sie angezeigt werden soll, sondern wo sie im Original stand), oder `<emph style="sm">`.

`<lb type="halfline">` innerhalb von `<emph style="sm">`: setze die beiden Zeilen untereinander und markiere die Note farblich. Beispiel:

```
<head>鑄造大小守銃尺量法<emph style="sm">近者多<lb type="halfline"/>遠者少</emph></head>
```

wird zu:

鑄造大小守銃尺量法近者多
遠者少

3.17 echo-gjs

Dieses Modul ist noch experimentell.

3.17.1 <place>

CSS: Sollte optisch erkennbar sein, z.B. kursiv

Am besten ein link zum Beispiel zum entsprechenden Eintrag in einem Authority file. Sollte an das GIS-System weitergereicht werden können. (Wie bei <num> könnte der User wählen, den standardisierten Namen angezeigt zu bekommen.)

3.17.2 <time>

CSS: Sollte optisch erkennbar sein, z.B. kursiv

Sollte in irgendeiner Form an das GIS-System weitergereicht werden können.

3.17.3 <person>, <event>

Es wird auch noch <person> und <event> geben, aber die Details sind noch unklar.

3.18 echo-textflows

Wie die div-Typen "multiflow" und "parallel" angezeigt werden könnten: User kann wählen zwischen

1. Seitendarstellung mit allen Flows auf dieser Seite
2. Seitendarstellung mit einem bestimmten Flow
3. Darstellung in der Reihenfolge wie im XML, ohne auf die Seiteneinteilung Rücksicht zu nehmen

Wenn es @flow in <pb> gibt, gilt <pb> nur für diesen Flow. Beispiele sind Eipo, Conimbricenses, Fußnoten. Wie mehrere Flows, insbesondere der Haupttext und die Fußnoten, gleichzeitig angezeigt werden können, müssen wir uns noch überlegen.

(<pb> in Fußnoten sind mit flow="footnote" gekennzeichnet. <pb> in Endnoten haben kein @flow.)

3.19 echo-import-xhtml

3.19.1 <xhtml:*>

<xhtml:*> kann direkt an den Browser weitergereicht werden. Beachte allerdings, dass wir nicht das content model von xhtml verwenden; man kann also nicht einfach die ganze Tabelle ohne Änderungen übergeben, sondern nur die Elemente mit dem Namespace `xhtml`.

3.20 echo-import-mathml

3.20.1 <mml:*>

<mml:*> kann direkt und ohne Änderungen an den Browser weitergereicht werden. Zumindest Firefox kann Formeln annähernd korrekt darstellen.