

Content Models for FEDORA 4

Fedora Users Group Meeting
Zentrum für Sprachkorpora, Universität Hamburg
7th December 2017

Johannes Stigler
Centre for Information Modelling, University of Graz



GAMS: An Infrastructure for Research Projects in the Digital Humanities

- About the name

- **Geisteswissenschaftliches Asset Management System**
(Asset Management System for the Humanities, <http://gams.uni-graz.at>)



- About the name

- **Geisteswissenschaftliches Asset Management System**
(Asset Management System for the Humanities, <http://gams.uni-graz.at>)
- the acronym is the German word for „chamois“



GAMS: An Infrastructure for Research Projects in the Digital Humanities

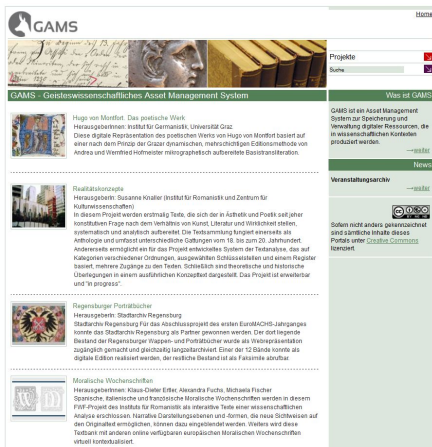
- About the name
 - **Geisteswissenschaftliches Asset Management System**
(Asset Management System for the Humanities, <http://gams.uni-graz.at>)
- Not only a technical infrastructure, but a strategy
 - a repository for publication and long-term archiving of digital resources (from the humanities)
 - both a storage layer and a framework for sustainable application development
 - certified according to the criteria of the Data Seal of Approval
 - built on FEDORA: we use it since version 2.0 (2004) up to 3.5
 - further integral technologies: Blazegraph triplestore, Apache Cocoon, Apache Solr, Loris IIIF image server
 - persistent identification of objects using Handle System
 - based on the XML paradigm of strict separation of content and presentation



GAMS: An Infrastructure for Research Projects in the Digital Humanities

■ About the name

- **Geisteswissenschaftliches Asset Management System**
(Asset Management System for the Humanities, <http://gams.uni-graz.at>)



The screenshot shows the GAMS website interface. At the top left is the GAMS logo. Below it is a banner image featuring a manuscript page, a classical bust, and a bookshelf. The main content area is titled "GAMS - Geisteswissenschaftliches Asset Management System" and lists three featured projects:

- Hugo von Morffort, Das poetische Werk**
HerausgeberInnen: Institut für Germanistik, Universität Graz
Diese digitale Repräsentation des poetischen Werts von Hugo von Morffort basiert auf einer nach dem Prinzip der Grazer dynamischen, mehrschichtigen Editionsmethode von Andrea und Wernfried Hofmeister mikrographisch aufbereitete Basisdigitalisierung.
- Realitätskonzepte**
HerausgeberIn: Susanne Knaller (Institut für Romanistik und Zentrum für Kulturwissenschaften)
In diesem Projekt werden erstmalig Texte, die sich der in Ästhetik und Poetik seit Jahrtausenden verankerten Frage nach dem Verhältnis von Kunst, Literatur und Wirklichkeit stellen, systematisch und analytisch aufbereitet. Die Textsammlung fungiert einerseits als Anthologie und umfasst unterschiedliche Gattungen vom 18. bis zum 20. Jahrhundert. Andererseits ermöglicht sie für das Projekt entscheidendes System der Textanalyse, das auf Kategorien verschiedener Ordnungen, ausgewählten Schlüsselstellen und einem Register basiert, mehrere Zugänge zu den Texten. Schließlich sind theoretische und historische Überlegungen in einem ausführlichen Konzepttext dargestellt. Das Projekt ist einwärtig und "in progress".
- Regensburger Plutarchüber**
HerausgeberIn: Stadtbibliothek Regensburg
Stadtbibliothek Regensburg für das Abschlussprojekt des ersten EuroMACHS-Jahrganges konnte das Stadtbibliothek Regensburg als Partner gewinnen werden. Der dort liegende Bestand der Regensburger Plutarch- und Plutarchüber wurde als Webpräsenz öffentlich zugänglich gemacht und gleichzeitig langzeitarchiviert. Einer der 12 Bände konnte als digitale Edition realisiert werden, der realistische Bestand ist als Faksimile abrufbar.
- Moralische Wochenschriften**
HerausgeberInnen: Klaus-Dieter Ertler, Alexandra Fuchs, Michaela Fischer
Spanische, italienische und französische Moralische Wochenschriften werden in diesem FP7-Projekt des Instituts für Romanistik als interaktive Texte einer wissenschaftlichen Analyse erschlossen. Narrative Darstellungsebenen und -formen, die neue Schlüsse aus den Originalen ermöglichen, können dazu einbezogen werden. Weiters wird diese Textbank mit anderen online verfügbaren europäischen Moralischen Wochenschriften virtuell kontextualisiert.

On the right side of the page, there is a navigation menu with "Home" at the top, followed by "Projekte" and "Suche". Below that is a section titled "Was ist GAMS" with a brief description of the system and a "News" section with a "Veranstaltungsarchiv" link. At the bottom right, there is a Creative Commons license notice.

GAMS: An Infrastructure for Research Projects in the Digital Humanities

- About the name
 - **Geisteswissenschaftliches Asset Management System**
(Asset Management System for the Humanities, <http://gams.uni-graz.at>)
- Not only a technical infrastructure, but a strategy
 - our projects range from Digital Scholarly Editions, to Image Collections and Language Corpora
 - almost 50 projects with over 100.000 (compound) digital objects handled by one FEDORA instance
 - an in-house developed Java client for ingesting and managing large-scale digital object collections
 - from FEDORA we appreciate the possibility to handle standardized, distributed workflows via Content Models



How do we understand Content Models?

as ...

- a structural definition for a „type“ of a digital object (e.g. manuscript, artifact, ontology etc.)
- as a class definition with
 - a predefined (but growable) set of data streams
 - a set of dissemination methods bound to an object
- a predefined set of workflows triggered during ingest or request of an object content/method
- a set of rules for creating a digital object



Content Models: An example

manuscript of an early modern period account book^a

^a<http://gams.uni-graz.at/o:srbas.1535>

Jahrrechnungen der Stadt Basel 1535 bis 1610 – digital
Beta-Version

HOME | **Jahrrechnungen** | Projekt
Chronologische Liste
Nach Konten gegliedert

Erweiterte Suche | Suche | Suche

Einnahmen

- Einnahmen Stadt
- Weinungeld
- Mehlungsd
- Stadtehezzoll
- Bischofviehzoll
- Pferdzoll
- Torzölle
- Wegzoll Neuer Weg
- Wiesenbrückenzoll
- Glpszoll
- Weinsticherzölse
- Wirtschaftsweinungeld
- Kaufhauszoll
- Pfundzoll
- Hausgeld
- Schultheißenstock Grossbasel
- Lade
- Gewinn Salzhandel
- Schultheißenstock Kleinbasel
- Bräskarren
- Kornausfuhrzoll
- Kaufhauszölse
- Gerberzoll
- Kemmerer Zoll
- Diepfinger Zoll
- Schiffzoll
- Wiesenfässerzoll
- Kaufhauskrän
- Pfundzoll der Fremden- und H.-Unterkauf Blei und Bleigläse
- Augster Zoll
- Rischzoll
- Durchweiszoll Rheinbrücke
- Steinweiszoll Rheinbrücke

Alles in den Datenkorb übernehmen

fol. 1r [?]

Jarrechnung a festo Johannis Baptistae anno xv^o xxxv^o usque ad festum Johannis Baptistae anno xv^o xxxv^o

fol. 1v [?]
fol. 2r [?]
fol. 2v [?]
fol. 3r [?]

Empfangen

Vom winungelt

- Prima angaria [? locii] lb
- Secunda angaria vij^o [sex] lb
- Tertia angaria vij^o [locii] lb
- Quarta angaria [? xxii] lb
- ↳ **Suma iij^o vj^o ix** lb

Vom müllkornungelt

- Prima angaria [? locj] lb
- Secunda angaria [?]^o [sex] lb
- Tertia angaria [? xxxj] lb
- Quarta angaria [?]^o [xvi] lb
- ↳ **Suma iij^o [? xv] lb**

Jahrrechnung Stadt Basel 1535/1536
SIABS Finanz H 93.1
Rechnungslegungszeitraum: 26.6.1535 bis 24.6.1536

Transkription von:
Sotke Calvi
Lukas Meili

Zitiervorschlag: Jahrrechnung Stadt Basel 1535/1536, bearb. v. Sotke Calvi und Lukas Meili, in: Die Jahrrechnungen der Stadt Basel 1535-1610 – digitale Edition, hg. v. Susanna Burgstaller, Basel/Graz 2015 (juletzt verändert am 22.8.2017), permalink über DOI-Handle: 11471/1010.1.1535

Datenkorb [Leeren]

Gesamtsumme: 0,00 lb.

Content Models: An example

- the transcript of the manuscript is coded based on the TEI^a schema
- during the ingest of the TEI document ...
 - facsimiles of pages will also be uploaded
 - the TEI document is enriched with metadata from internal and external sources (geonames.org, GND, SKOS ontologies, etc.)
 - RDF triples are extracted from the TEI document and stored in embedded Blazegraph triplestore
 - descriptive metadata (DC, EDM) is extracted from the TEI document and stored in the object
 - based on the structure information in the TEI document an IIIF manifest is created

^aText Encoding Initiative, <http://tei-c.org>

- on-the-fly during the request to an object disseminator ...
 - HTML, PDF, \LaTeX PDF representations of the object are generated and delivered
 - the facsimile content can be viewed via Mirador
 - the text content can be viewed via the Voyant Tools ^a
 - a data matrix for analyzing (text-) data will be created
 - a query object allows standardized search requests on the textual content of a project

^a<http://voyant-tools.org>



Content Models: An example

HTML and PDF representation of a TEI object ^a ^b

^a<http://gams.uni-graz.at/o:mws-096-334/sdef:TEI/get>

^b<http://gams.uni-graz.at/o:mws-096-334/sdef:TEI/getPDF>

The screenshot shows a digital edition interface for 'Moralische Wochenschriften'. At the top, there are language options: Deutsch, Englisch, Französisch, Italienisch, Spanisch. The main title is 'Periodical Press Spettatori Espectadores Spectateurs Moralische Wochenschriften'. Below the title is a navigation bar with 'ENGLAND', 'FRANKREICH', 'DEUTSCHLAND', 'ITALIEN', 'SPANIEN', 'Gesamtschau', 'Erweiterbare Suche', and a search box. The main content area is divided into two columns. The left column contains the text of the 'Discurso Decimotercio' (Discourse Decimotercio), which discusses the state of the printing press and the decline of the sciences. The right column contains a sidebar with 'Originalansicht', 'Druckversion (Analyseansicht)', 'Darstellungsoptionen', and 'Erstformen'. The 'Darstellungsoptionen' section has checkboxes for 'Ersch. 1', 'Ersch. 2', 'Ersch. 3', 'Ersch. 4', 'Ersch. 5', and 'Ersch. 6'. The 'Erstformen' section has checkboxes for 'Abbildung', 'Autoren/Entwickler', 'Ort', 'Titel', 'Fragepostul', 'Bibliographisch', 'Metadaten/URL', 'URL', 'Angebotsort', 'Traum', 'Länge', and 'Sprache'.

MORALISCHE WOCHENSCHRIFTEN

Institut für Romanistik, Karl-Franzens-Universität Graz

Femlink: <https://gams.uni-graz.at/o:mws-096-334>

Zitierverschlag: Anzonym (García de Cabañal, Luis Perera, Luis Marcelino) (Hrsg.), "Discurso Decimotercio", in *El Cesar*, Vol.13 (1783), S. 109-202, ediert in: *Enlie, Klaus-Dieter / Hübisch, Elisabeth* (Hrsg.): *Die "Spectators" im internationalen Kontext*. Digitale Edition, Graz 2011-2017. hdl.handle.net/1471/5112/1.289 [aufrufen am 28.11.2017]

Discurso Decimotercio

*Namque oblitus modi millesima pagina surgit
Omnibus, & crescit multa damnosa pagyna.*

Juv. Sat. 7. v. 100.

Un libro se hace al punto de este modo
Que de mil hojas, si se quiere, pasa,
Con el qual se molesta al mundo todo
Desperdicando en el papel sin tasa.

Se lamentaba el otro día un Letrado, de que no se escriben ahora en nuestra España tantos Tratados de Jurisprudencia, como en los dos siglos antecedentes. No vemos, decía, casi con las lágrimas en los ojos, salir en nuestros días de las prensas de estos volúminos Comentarios, de estas copiosas glosas, de estas colecciones inmensas de Decisiones de la Rota, Questiones, Disputaciones, Consejo, Allocations, &c. que llenan nuestros estantes, y son la honra de nuestros Estudios. ¡Tanta es como todo esto el decadencia de las Letras! ¡Tan grande el abandono en que yace la mas noble, y mas importante de las Ciencias humanas! Mas iba á decir: pero lo vivo de su dolor atajo aqui su Discurso.

A la verdad sus quejas no pueden ser mas bien fundadas, y todo buen Ciudadano debiera acompañarle en ellas, y llorar con lágrimas de sangre este triste efecto de nuestra indolencia, y aversion al trabajo. Es cierto, que no faltan algunos que quisieran reducir el Estudio de toda la Jurisprudencia Civil, al texto solo de las Leyes, y desterrar por consiguiente semejantes Obras. Tal fue en otro tiempo la idea del Emperador Justiniano, y tal la que en este mismo siglo tubo entre otros Soberanos Victor Amadéo de Saboya, Ahuelo del actual Rey de Cerdeña, que prohibió su uso en todos los Tribunales de sus Estados. Pero estos Príncipes (sea dicho con la venia de sus difuntas Magestades) no han procedido en esto, con la madurez que de ellos debia esperarse.

Son muchas las utilidades, que traigan las Obras que tenemos de esta clase á la República, y no menores las que producian otras del mismo gusto, si hubiese ahora Sugeros de bastante zelo, para emprenderlas. En primer lugar, una ininidad de trapos que ahora quedan abandonados, y sin la mas remota esperanza de salir jamás de entre la basura, y el estiercol, se harian entonces utiles; y despues de haber servido en los usos mas viles, tendrian con todo eso seguro un destino mas brillante, y una fortuna sin comparacion mas elevada, que toda la que antes

Content Models: An example

TEI object viewed via Mirador^a

^a<http://gams.uni-graz.at/o:mws-096-334/sdef:IIIF/getMirador>

Discurso Decimotercio

Layout ändern Vollbild

Discurso Decimotercio

190 EL CENSOR. 191

nuestros días de las prensas de estos voluminosos Comentarios, de estas corpulentas glosas, de estas colecciones inmensas de Decisiones de la Rota, Questiones, Disputaciones, Consejos, Alegaciones, &c. que llenan nuestros estantes, y son la honra de nuestros Estudios; ¡Tanta es como todo esto la decadencia de las Letras! ¡Tan grande el abandono en que yace la mas noble, y mas importante de las Ciencias humanas! Mas iba à decir: pero lo vivo de su dolor atajo aqui su Discurso.

A la verdad sus quejas no pueden ser mas bien fundadas, y todo buen Ciudadano debiera acompañarle en ellas, y llorar con lágrimas de sangre este triste efecto de nuestra indolencia, y aversión al trabajo. Es cierto, que no faltan algunos que quisieran reducir el Estudio de toda la Jurisprudencia Civil, al texto solo de las Leyes, y desterrar por consiguiente semejante

DISCURSO XIII. 191

la idea del Emperador Justiniano, y tal la que en este mismo siglo tubo entre otros Soberanos Victor Amadeo de Saboya, Abuelo del actual Rey de Cerdeña, que prohibió su uso en todos los Tribunales de sus Estados. Pero estos Príncipes (sea dicho con la venia de sus difuntas Magestades) no han procedido en esto, con la madurez que de ellos debía esperarse.

Son muchas las utilidades, que trajeron las Obras que tenemos de esta clase à la Republica, y no menos las que producirian otras del mismo gusto, si hubiese ahora Sugetos de bastante zelo, para emprenderlas. En primer lugar, una infinidad de trapos que ahora quedan abandonados, y sin la mas remota esperanza de salir jamás de entre la basura, y el estiercol, se harian entonces utiles; y despues de haber servido en los usos mas viles, tendrian con todo eso seguro un destino mas brillante, y una sin comparacion mas elevada, que

Page 188 Page 190 Page 191 Page 192 Page 193 Page 194 Page 195 Page 196 Page 197 Page 198 Page 199 Page 200 Page 201

Content Models: An example

TEI object viewed via Voyant Tools^a

^a<http://gams.uni-graz.at/o:mws-096-334/sdef:Voyant/get>

The screenshot displays the Voyant Tools interface for a TEI document. On the left, a word cloud highlights the most frequent words, with 'mas' being the largest. The main text area shows the beginning of a document titled 'Discurso Decimotercio; El Censor (Anonym (García de Cañuelo, Luis; Pereira, Luis Marcelino))'. The text discusses the state of printing in the 16th century. On the right, a line graph plots 'Rare Frequencies' across 'Document Segments' (1 to 20), with five lines representing different terms: 'a', 'no', 'mas', 'no', and 'estas'. Below the graph is a concordance table for the selected term 'mas'.

Discurso Decimotercio; El Censor (Anonym (García de Cañuelo, Luis; Pereira, Luis Marcelino))

Discurso Decimotercio

Namque oblitia modi millesima pagina surgit Omnibus, & crescit multa damnosa papayo.

Juv. Sat. 7. v. 100.

Un libro se hace al punto de este modo Que de mil hojas, si se quiere, pasa Con el qual se molesta al mundo todo Desperdiçando en el papel sin tasa.

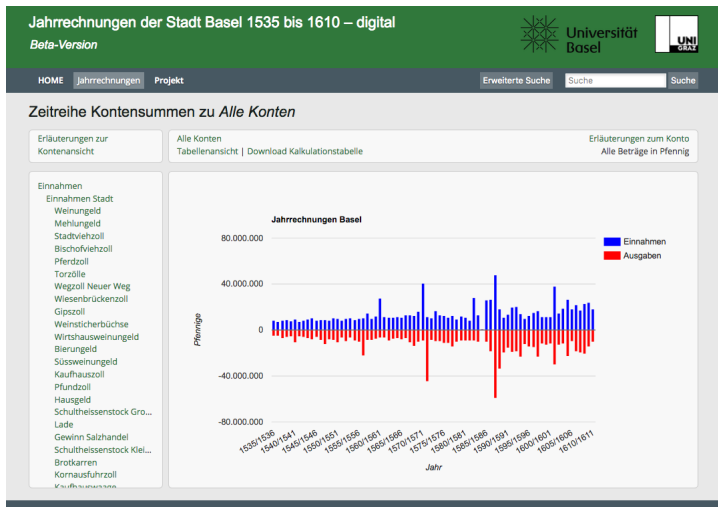
Se lamentaba el otro dia un Letrado, de que no se escribian ahora en nuestra Espanya tanto. Tratados de Jurisprudencia, como en los dias siglos antecedentes. No vemos, decia, casi con las lagrimas en los ojos, salir en nuestros dias de las prensas de estos voluminosos Comentarios, de estas corpulentas glosas, de estas colocaciones inmensas de Dicciones de la Rota, Questiones, Disputaciones, Consejos, & Memorandums. En su lugar se venian en aquellos dias, si bien la lengua ha mudado...

mas

Document	Left	Term	Right
[1] Dis...	las Ciencias humanas! Mas	a	decir: pero lo vivo de
[1] Dis...	que tenemos de esta clase	a	la Republica, y no memora
[1] Dis...	Qué hombre que tenga ley	a	la camisa que viste, dejará
[1] Dis...	y que se vea precisado	a	separarla de su cuerpo podrá
[1] Dis...	su primera blanca, y llegar	a	ser uno de los adomos
[1] Dis...	nuestros basureros, se aumentaria con...	a	favor de semejantes Obras, y
[1] Dis...	la Republica debería por tanto	a	sus Autores la subsistencia de
[1] Dis...	ocupacion, y se ven reducidos	a	la mendiguitz, ó en una
[1] Dis...	ocasion muy fuerte de recurrir	a	medios delinquentes, para tener que
[1] Dis...	teniendo los ricos menos acreedores	a	sus limosnas, y los verdaderamente

Content Models: An example

Query object showing a summary report of the account books



An incomplete list of ...

- CMs supporting specific XML schemata
 - TEI
 - LIDO
 - MEI
 - METS
 - SKOS
 - OWL
 - BIBTEX
- CMs for aggregated content
 - Context
 - Query
 - R
- CMs for static content
 - PDF
 - Resource

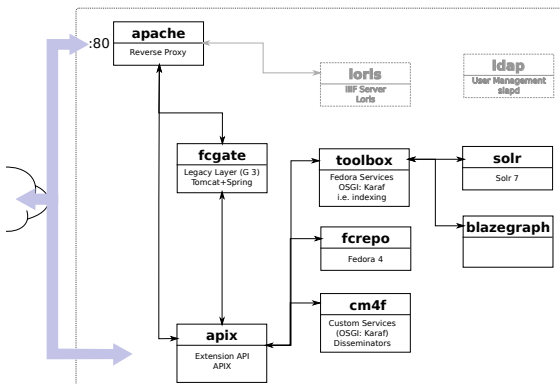
Goals of the project

- we want to migrate the philosophy and the REST API of FEDORA 3.x (eg. the access path conventions of 3.x) to FEDORA 4
- ... holding our mind open for new possibilities arising with FEDORA 4
- (*Eclipse Rich Client Platform*) client with generic access and management layer for *FEDORA 3.x* and *4.x* repositories
- tool for generating API-X OSGI modules from WSDL data stream of FEDORA 3 content model objects
- Docker based system architecture with a single point of access to all services



GAMS goes FEDORA 4

GAMS system architecture under FEDORA 4



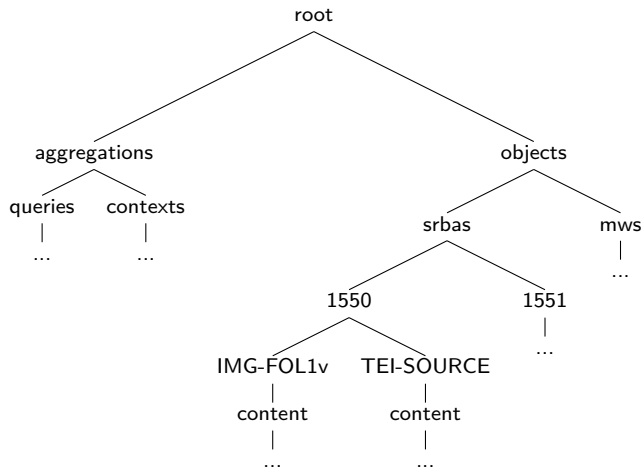
A new data model

- premis is of our data model are oriented on the philosophy of PCDM (see CM4F, <http://cm4f.org>)
- for modeling and storing content data, we use well-established standards like TEI, LIDO, MEI, SKOS, OWL etc.
- for metadata DCMI, EDM, IIIF, METS, MODS etc.
- we prefer human readable path names
- levels of our data model are structured by . . .
 - content model class from FEDORA 3 bound to an object
 - persistent identifier of the object



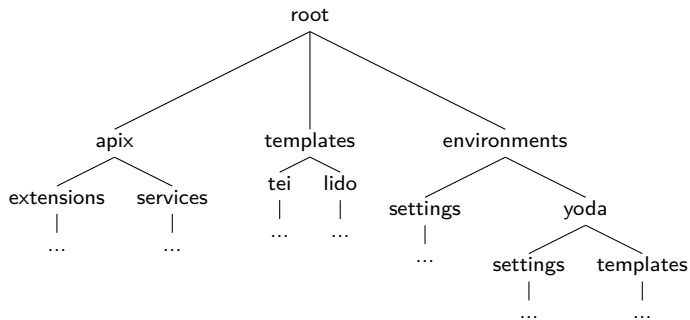
A new data model

Content objects



A new data model

Environment objects



Questions to the community

- How could YOU imagine such research data repository in a FEDORA 4 environment?
- What is the status of the API Extension Architecture subproject?
- Is there a suggested backup strategy for huge data stores?
- Any plans for a replication mechanism to support clustered deployments?



Thank you for your attention!

johannes.stigler@uni-graz.at
<http://gams.uni-graz.at>
<http://informationsmodellierung.uni-graz.at>

