

I22 Stigler

**I22-1#Interviewer: Inwiefern arbeiten Sie im Bereich eHumanities? Bitte beschreiben Sie ihre Tätigkeit, ihr Forschungsfeld in dem Bereich und ihre Erfahrung.**

Befragte/r: Also, ich bin [Anonymisierung:] Leiter/in eines Zentrums für Informationsmodellierung, aus einem Aspekt einer angewandten Forschung mit Fragen digitaler Editionstechnik, digitalen Archiven und wir sind auch gerade dabei, das Kölner EUROMACHS-Modell, nicht nur Kölner, sind ja vier, fünf Hochschulen..., sozusagen an diesem Modell zu partizipieren. Wir haben vorgestern die Genehmigung unserer Universität bekommen und hoffen, dass wir im Herbst dieses Jahres einsteigen können in das Masterstudium, Master Degree Euromax, haben bei uns in unserem Umfeld an der geisteswissenschaftlichen Fakultät jetzt seit dreieinhalb Jahren ein so genanntes Wahlfachmodul „Informationsmodellierung“, wo wir uns mit Fragen digitaler Edition, digitaler Editionstechnik angewandt beschäftigen.

**I22-2#I: Ja, soweit ich informiert bin, hieß das vorher „Informationsverarbeitung“, jetzt „Informationsmodellierung“. Warum diese Namensänderung?**

B: Ich würde sagen: Das ist der ganz zentrale Aspekt des Einsatzes von Informationstechnologien im Bereich der Geisteswissenschaften. Ich sage auch immer: In Zukunft werden sich Geisteswissenschaftler verstärkt sozusagen mit Fragen digitaler Repräsentation von Wissen und Wissensprodukten, die sie produzieren, auseinandersetzen müssen. Wir haben heute gerade über das Thema gesprochen, sozusagen ein „unendlicher“ Ansatz in diesem Kontext sagt, es ist wichtig, sich zu überlegen, mit welchen Technologien man an Dinge herangeht. Ich sage: Technologien werden immer unwichtiger, die ändern sich ständig. Morgen gibt es eine performante XML-Datenbank, dann kann ich..., dann stellt sich die Frage ganz anders plötzlich, was aber nachhaltig ist, ist sozusagen die Qualität der Daten. Wenn diese Daten modelliert vorliegen, wenn der Text, der semantisiert ist mit unterschiedlichen..., Bezug nehmend auf unterschiedliche Fachdomänen, unterschiedliches Fachwissen, dann macht das Ganze Sinn. Dann kann man so ...Textdaten in unterschiedlichsten Frameworks weiterverwenden. Deswegen haben wir uns ganz bewusst Zentrum für Informationsmodellierung genannt, um deutlich zu machen, das ist der eigentliche Aspekt, wenn es in der Geisteswissenschaft um die Frage von Datenverarbeitung geht. Mit welchen Frameworks und Technologien das weiterverarbeitet ist, ist Aufgabe von Informationstechnikern im weitesten Sinne, damit muss sich ein Geisteswissenschaftler oder eine Geisteswissenschaftlerin gar nicht auseinandersetzen. **I: Das heißt: Es geht also um die Verarbeitung von fachspezifischen Themen hin zu Modellen, die dann mit**

**unterschiedlichen Technologien dargestellt werden?** B: Also wir beschäftigen uns ganz zentral mit Annotationssprachen unterschiedlichster Art: Wir verwenden TEI; wir versuchen zunehmend auch uns mit Fragen der ontologiebasierten Modellierung von Wissensdomänen zu beschäftigen. Das sind also Aufgaben, wo man sagt, da sind wir an der Fachwissenschaft, da werden auch solche Fragen zu methodologischen Fragen der jeweiligen Disziplin. Wenn ich mir überlege: Wie kann man in der Germanistik oder in der Kunstgeschichte ontologiebasiert bestimmtes Fachwissen modellieren, das sind Fragen, die können dann nur Fachwissenschaftler beantworten. Wie man diese Ontologien dann dazu benutzt, um Daten zu erschließen und einer prädikatenlogischen Abfrage sozusagen zu erschließen, das ist die Frage des Informationswissenschaftlers, ... darauf basierend.

**I22-3#I: Was haben Sie sich dann vor diesem Hintergrund von TextGrid erwartet und von dem Workshop, an dem Sie teilgenommen haben?** B: Also von TextGrid habe ich mir erwartet..., also ich bin in bestimmten Bereichen in meinen Erwartungen bestätigt worden. Ich habe also von TextGrid erwartet: ein standardisiertes Instrument, das es mir ermöglicht, solche Modellierungsfragen sozusagen mit entsprechender Usability für Fachwissenschaftler zu ermöglichen. Ich habe den Überblick, den ich bekommen habe..., da habe ich den Eindruck, das geht genau in diese Richtung. Ich meine, es ist zutiefst sinnvoll zu sagen: Nicht jeder erfindet ständig die Welt neu, sondern ein standardisiertes Tool, das noch, wie sich zeigt, sehr modulhaft und erweiterbar sich gestaltet, so dass man das wirklich als Plattform sehen kann, wo sozusagen unterschiedlichste Werkzeuge einfließen können, vom intelligenten Scannen, aber auch zum Auszeichnen von Dokumenten zur Bearbeitung, zur XML-basierten Verarbeitung. Wo ich zuvor ein bisschen enttäuscht war fürs erste, das war für mich der Aspekt, das ist angeklungen, in bestimmten Bereichen zu wenig standardbasiert sich bewegt in TextGrid. Für mich war das nicht nachvollziehbar, was z.B. was die Auszeichnung von diesem grafischen Editor betrifft, dass dann dieser grafische Editor nicht irgendwelche standardbasierten Daten raus schreibt, sondern dass das irgend so ein natives Format ist.

**I22-4#**Wo ich auch ein bisschen enttäuscht war, wo ich noch sehe, da könnte man sich einfach weiterentwickeln, ist die Strukturierungsmöglichkeit von Informationen im System, die einfach noch zu filebasiert ist. Also ich glaube, wir haben ein System, wir bewegen uns da FEDORA-basiert, das ist auf der Ebene der Datenstrukturierung wirklich als ein objektorientiertes Modell konzipiert. Also, da kann ich unterschiedliche Datenströme zu einem Objekt zusammenfassen und dann auch auf diesen Datenströmen basierend die Methoden definieren und über diese

Methoden erschließt sich sozusagen der Inhalt eines solchen Objektes. Also Objekte sind bei uns mehr als nur so ein File, es ist wirklich eine logische Einheit, in der ich z.B. ... ich hätte eine Transliteration einer Handschrift in Verbindung mit den Bildern, wo ich sagen kann: die Transliteration als TEI-Dokument, die niederauflösende Variante eines Bildes als Zip-Pak und die hochauflösende Variante des Bildes ergeben zusammen ein Objekt. Die Transliteration in Verbindung mit dem Bild und einer XSLT-Transformation wird zu einer HTML-Seite oder einem PDF-Dokument als Funktionen, die darauf operieren – dann noch die Möglichkeit, Zugriffsrechte auf die unterschiedlichen Entitäten eines solchen Objekts zu definieren. So kann man sagen, die niederauflösende Variante des Bildes im Kontext der Website ist für alle einsichtig, aber die hochauflösende Variante des Bildes kann ich nur über Authentifizierungsprozesse erschließen, weil ich möchte nicht, dass jeder meine hochauflösenden Bildvarianten lesen kann, und das Ganze ergibt eine Einheit, die ich logistisch im System gemeinsam verwalten kann. Ein Nutzer kann sagen: Erstelle mir eine Folioseite z. B. und die umfasst sozusagen diese ganze komplexe Funktionalität. Da war mir dieser Ersteindruck, da hab ich das Gefühl gehabt, das ist File-System-basiert, das war mir zu wenig, ehrlich gesagt. **I: Ja, es hat also zu sehr diese Ordnerstruktur, die man auch von anderen Textverarbeitungssystemen her kennt...? Und es müsste dann mehr in Richtung Funktionalitäten wie in einem Hypertext gehen, oder...?** B: Kann auch sein... ganz allgemein formuliert: Es müsste stark in Richtung einer Objektorientierung gehen, dass ich die Möglichkeit habe – das geht unter FEDORA – Inhaltsmodelle zu definieren, die komplexere Strukturen sind. Also man könnte sagen, mehrere Datenströme in Verbindung mit bestimmten darauf operierenden Methoden in Verbindung mit XSLT-Style-Sheets, die diese Inhalte irgendwie transformieren, kann ich in ein Objekt zusammenfassen und dann kann ich sagen, ich kann Objekte des gleichen Typs anlegen. Ich hab z.B. eine Handschriftenedition und ich habe für jede Folioseite ein solches Objekt und das kann ich dann befüllen. Ich habe auch eine Grundstruktur, die auch mit dem Anlegen des Objekts bereits existiert und auch die darauf operierenden Funktionalitäten, genau diesen Aspekt möchte ich in G hervorheben. Das wäre für mich ein Wunsch, dort, in diesem Referat, weil das wirklich aus meiner Sicht wirklich etwas ist, an dem man etwas lernen kann. **I: Das wäre ja jetzt ein systemübergreifender Aspekt, das würde ja auf alles zutreffen, auf alle Tools, auf alle Fachwissenschaften...** B: Das ist also eine Grundfrage der Datenorganisation im System für mich. Vor allem dann, wenn man in Richtung Publikationsveröffentlichung denkt. Ich glaube, das ist ja auch ein Bereich, der in TextGrid noch

nicht angegangen ist. Im Moment ist das ein Tool zur Bearbeitung. Ich kann mir gut vorstellen, wenn man schon damit begonnen hat, das Tool in unserem Kontext, in unserem Denken und in unserem System einzusetzen, als Tool, mit dem wir Inhalte generieren, sehr hilfreich für uns schon und ganz gut und wichtig, brauchbar.

**I22-5#I: Das ist ja dann ein Aspekt, der aus der Nutzerperspektive auch völlig neue Möglichkeiten bieten würde, nämlich den Datenzugriff verändern.** B: Also, ja genau, wir sind bei der Konzeption unseres System davon ausgegangen..., wir haben immer gesagt: Archivsysteme sind Folgendes: Das sind Datenbanken, in die Benutzer was hineinspeisen, ganz mühsam, und wo nie irgendwas rauskommt. Wir haben an vielen solchen Projekten gearbeitet und da kommt nix raus. Und da fragt sich jeder Fachwissenschaftler: Warum tue ich das? Wie gesagt, ein Archiv ist etwas, was in der Benutzung entsteht, das war unser Grundkonzept und eine Prämisse von uns, Archive sind etwas, was in der Benutzung entsteht, und was immer einen unmittelbaren Nutzen für die Befüller haben muss, d. h. aus dieser dynamischen Sicht, aus unserem Repository werden dynamisch die ganzen Projektwebseiten generiert. Das heißt, das Repository, das Archiv ist sozusagen die Datenbasis, aus der dynamisch Webseiten oder PDFs generiert werden. Also, wir haben viele Editionsprojekte, da gibt es im System kein PDF, es gibt einen TEI-modellierten Inhalt und es gibt pro Objekt Referenzen auf Stylesheets, die diesen Inhalt zur HTML-Seite machen oder zu PDF. Bei Anklick wird das PDF generiert, mit dem Vorteil gleichzeitig, wir können über die Veränderung einer Transformationsvorschrift in einem Rutsch 5000 Dokumente verändern. Wir müssen also nicht das alles wieder angreifen, da setzt man im Prinzip diese Prämisse, die eh bekannt ist von XML, die Trennung von Inhalt und Repräsentation. Wir haben einen gut modellierten Inhalt, was für uns auch den Vorteil bringt, dass, wenn wir Projekte angehen, die Fachwissenschaftler beginnen können, den Inhalt zu erstellen, ohne dass noch das Layout oder noch bestimmte andere Fragen geklärt sind. Da können die schon anfangen, Inhalte zu produzieren, weil dann kommt die Grafikerin, macht das Grafikkonzept und dann setzt bei uns jemand im Haus die entsprechenden XSLT-Stylesheets um, die dann den Inhalt sozusagen in das entsprechende Repräsentationsformat überführen.

**I22-6#I: Das würde ja auch bedeuten, dass ein Ziel ist, dass man Software anbieten müsste für Fachwissenschaftler, die kein XML können.** B: Ja unbedingt. Das ist bei uns Grundvoraussetzung. Also wir verwenden zur Inhaltserstellung..., oder haben vor TextGrid sozusagen, in bestimmten Bereichen können wir das verwenden, aber wir verwenden auch noch immer in Projekten folgenden Workflow: Wir definieren unter OpenOffice, mittlerweile sind wir

dabei da umzusteigen, bis jetzt haben wir immer Word verwendet, aber wenn wir dahin kommen, möchten wir OpenOffice verwenden. Wir definieren unter Word, also in der Textverarbeitung Stylesheets, ...Dokumentenvorlagen, die kann man nämlich gut benutzen, um sozusagen einzelne Textteile auszuzeichnen. Man kann sagen, dann... ich habe einen Autor, da markier ich in Word einen Namen und sage: Das ist ein Autornamen. Und dann generieren wir automatisiert über Visual-Basic, sozusagen XML, in dem Fall bei uns immer Standardformat mittlerweile TEI, TEI-Dokumente. D.h. der Fachwissenschaftler, der auszeichnet, braucht nur systematisch mit Word zu arbeiten, drückt dann einen Knopf und hat die XML-Repräsentation des Dokuments. Das war wirklich für mich eine Offenbarung unter OpenOffice, dass unter OpenOffice sozusagen es möglich ist... dass also dieses Open Document Format wirklich, dass das Dokument als Open Document Format unter OpenOffice repräsentiert ist, d. h. man kann XSLT, sprich standardisierte Technologien, wieder benutzen, um aus dem OpenOffice-Dokument ein TEI-Dokument zum machen. Ich kann einfach ein Filter definieren und dann schreibe ich von OpenOffice sozusagen über diesen Filter ein entsprechendes Dokument als TEI-Dokument aus. Diesen Weg möchte man diesbezüglich sicher auch gehen und das ist in vielen Projekten ganz wichtig. Also man kann nicht jedem Fachwissenschaftler in jeder Fachwissenschaft sagen..., also auf's Auge drücken: Du musst TEI, also jetzt XML kodieren. Und noch dazu, auch das ist ein Weg, den wir verfolgen, gibt es in vielen Disziplinen eingeführte, aus einer bestimmten Perspektive natürlich proprietäre Weisen zu annotieren, z. B. die Historiker haben das Leidener Klammersystem, das eingeführt ist, und da gehen wir auch da den Weg und sagen: Wir nehmen das System, das eingeführt ist, und entwickeln einfach Tools, die dieses Leidener Klammersystem dann für die Weiterverarbeitung in ein TEI-Dokument überführen. Dann können die wiederum ihre gewohnten Formen der Auszeichnung verwenden. Wenn man das XML-basiert machen würde, würde das leicht unübersichtlich. Und wir schaffen dann einfach ein Tool, das in diese Form der Notation einfach überführt, in ein XML-Dokument für die Weiterverarbeitung.

**I22-7#I: Ja, sehen sie das so, dass der Weg dahin recht unproblematisch ist oder welche Barrieren sehen sie da? Das klingt jetzt so, als müsste das nur fertig programmiert werden.**

B: Also ich sage immer oft, unsere Arbeit besteht zu 90 Prozent sozusagen in didaktischer Arbeit. Wir müssen die Fachwissenschaftler überhaupt erst dorthin führen, dass sie diese Formen des Arbeitens annehmen. Bei Geisteswissenschaftlern ist es oft noch viel schlimmer, weil sie manchmal einfach fürchterlich unsystematisch denken, d. h. man muss sie überhaupt erst

hinführen, dass sie in gewisser Weise systematisch an ihre Aufgabenstellung herangehen. Sie haben Systematik im Kopf, aber sie sind oft nicht in der Lage, sie zu realisieren. D.h. das ist in unseren Projekten immer der wichtige Schritt, dass wir ..., das ist der erste Schritt und aus unserer Sicht der wichtigste. Oft war das so, von uns Fachwissenschaftlern: Wir machen was und bitte nehmt das und macht damit irgendwas. Diesen Weg gehen wir nicht. Das heißt, am Anfang steht immer sozusagen ein von uns moderierter Prozess, in dem die Fachwissenschaftler darauf kommen müssen, was sie eigentlich von einem solchen System wollen können im weitesten Sinne. Also, es ist nicht so, dass Sie es uns sagen und wir setzen es um, sondern wir moderieren sozusagen ihre Überlegungen dabei: Was können sie überhaupt wollen von einem solchen System. Das ist ganz wichtig. Und das geht auch deswegen, weil bei uns im Haus, wir sind eigentlich alle Geisteswissenschaftler, die sich mit einem technischen Hintergrund an die Dinge heranzubewegen. Ich stelle auch oft fest, dass oft Techniker, das stelle ich auch fest, weil bei uns im Haus Mitarbeiter existieren, die auf der Technik studieren, dass Techniker einfach nicht in der Lage sind, zu erkennen, wo die effektiven Probleme, nämlich die methodologischen Probleme von Geisteswissenschaftlern liegen, weil sie eine ganz andere Sprache haben, meistens aus einem strukturierten..., mit einem sehr kategorialen Denken an die Dinge herangehen. Und in der Geisteswissenschaft stellen sich Probleme anders, nicht so sehr kategorial. Und XML insgesamt, auch das beobachte ich auch oft, wird von den Technikern oft völlig unterschätzt in der Bedeutung. Für die Geisteswissenschaftler ist es ganz ein wichtiges Instrument, da es uns in der Informationsverarbeitung den Text zurückgegeben hat, wie ich immer sage. Wir können also am Text plötzlich sein und dort arbeiten, dort Dinge anbringen. Daher sage ich, das ist für uns ein viel größerer, wichtigerer Paradigmenwechsel wie für die Informationsverarbeiter, die das oft unterschätzen. Wir haben auch bei uns in unserer Ausbildung zunehmend Studierende der TU, die Technische Universität in G, die zu uns kommen, das hat sich rumgesprochen, die uns sagen: So hören wir das nie auf der TU, so haben wir das noch nie gehört, wie ihr das so sagt. Die kommen auch zunehmend gerne in unsere Veranstaltungen. **I: Das heißt, der Prozess funktioniert so: Man muss einerseits fragen: Was wollen die Nutzer? Andererseits den Nutzern wieder sagen: Das gibt es. Und dann wieder fragen: Was wollt ihr wirklich und so weiter?** B: Ja, das ist ein iterativer Prozess. Oft wissen die gar nicht, was sie wollen können. Und wenn man ihnen zeigt, sozusagen, was Modellierung unter Umständen ermöglicht, also wenn man einen semantisierten Text vorlegt, was dann an Funktionalitäten möglich ist, dann gehen einem die Augen auf und sagen plötzlich: Ja ok, das hat ja wirklich eine Bedeutung auch für

unser wissenschaftliches Denken, weil das ermöglicht uns andere Formen des Herangehens an bestimmte Probleme. Das ist insofern immer ein wechselseitiger Prozess.

**I22-8#I: Ja, ein weiteres Problem oder sagen wir Herausforderung, was auch die Fachwissenschaftler angeht: TextGrid ist ja als Infrastruktur geplant. Welche Potentiale, aber auch welche Barrieren sehen Sie in dem Bereich?** B: Dazu kenn ich es zu wenig. **I: Es muss nicht TextGrid-spezifisch sein, sondern digitale Infrastruktur.** B: Digitale Infrastruktur? Im Prinzip gibt es immer noch sehr große Ressentiments gegenüber digitalen Strukturen, das merke ich zunehmend. Das klassische Argument des Fachwissenschaftlers: Das Buch, das ist gedruckt, das liegt vor, aber wie ist das mit digitalen Strukturen? Da wissen wir nicht, die können verschwinden. Dann kommen diese Argumente. Da kann man das nicht lesen. Was ja auch völlig Gültigkeit hat, da haben wir heute ein Paradebeispiel dieses Falls: Ein Kollege bringt eine Powerpoint-Präsentation mit, startet die auf dem Laptop – geht nicht. Da, wenn man sich auf Microsoft einlässt, wenn es im Bezug auf das Thema Nachhaltigkeit geht, ist man selbst Schuld. Hingegen, wenn man XML-basiert arbeitet, deswegen auch Informationsmodell..., wenn man sagt: Ok, ein Dokument ist sozusagen selbstreferenziell insofern als es für sich steht und alle Informationen, die notwendig sind, um dieses Dokument zu verstehen, in sich trägt. Das ist ein ganz ein wichtiger Aspekt von einem digitalen Archiv. Diese Sachen darf man nicht nur ausschließlich datenbankorientiert lösen, sondern wir sagen, zunächst ist das Textdokument, was für sich steht, das alles, was es enthält, auch in sich beschreibt, also redundant, also und dann kann ich ein solches Dokument nehmen und wenn ich drauf komme, relationale Strukturen sind gut, um das bereitzustellen, dann nehme ich das Dokument und importiere es automatisiert in diese Umgebung. Aber zunächst besteht die Datenqualität darin, einzelne atomar für sich existierende Dokumente zu haben. Was Textgrid..., Was war die Frage? **I: Welche Potentiale Sie in der digitalen Infrastruktur sehen, die ja unter anderem durch TextGrid aufgebaut werden soll bzw. welche Barrieren?** B: Es ging sozusagen um..., es gibt sozusagen noch große Ressentiments gegenüber digitalen Strukturen. Die Zukunft liegt sicher in der Hybridedition, nehme ich an, die zunehmend wahrscheinlich abgelöst wird von rein digitalen Formen. Jedenfalls erforderlich dafür ist eine Übergangsphase, in der auch dieses Vertrauen wachsen kann. Der Fachwissenschaftler in diesen digitalen Strukturen...braucht natürlich Projekte wie URN und ähnliche Ansätze .... Persistent Identifier, Link Resolver – also dieser Problemkreis. Es braucht natürlich auch dort Strukturen, die sozusagen sicherstellen, dass sozusagen digitale Ressourcen wirklich nachhaltig zitabel werden. Die gibt es, aber sie sind noch nicht eingeführt und so

akzeptiert, dass sie wirklich von Bedeutung wären. Zunehmend wird es in diese Richtung gehen, aber es ist noch ein langer Weg, noch ein langer Prozess...

**I22-9#I: Ja, Sie haben eben das Stichwort Usability erwähnt und jetzt im Zusammenhang mit Ressentiments ist das ja ein wichtiger Faktor, gerade bei weniger technikaffinen Nutzern. Was für einen Eindruck hatten Sie da von TextGrid, was die Usability angeht?** B:

Die Usability erscheint mir eigentlich zufriedenstellend. Es ist eine eclipse-basierte Umgebung und die Frage der Sprache..., für mich ist es kein Problem, wenn es Englisch ist. Wir stellen einfach fest in unserer Ausbildung, vor allem weil auch zunehmend Kollegen das wahrnehmen, das Angebot bei uns, dass wenn wir englische Lehrveranstaltungen anbieten, dass dann noch eine zusätzliche Ebene hereinkommt. Also vor allem technische Inhalte englisch angeboten macht ein Problem, auch wenn die Kollegen und Kolleginnen überhaupt kein Problem mit Englisch haben, aber es ist eine zusätzliche Ebene, die verunsichert. Insofern ist eine mehrsprachige Plattform ideal. Gleichzeitig stellen wir fest in unseren Projekten: Es ist einfach mühsam immer zu übersetzen. Und wir tendieren zunehmend, was die Website unseres Institutes betrifft, was Applikationen betrifft, die wir entwickeln, zu einer englischen Sprache. Weil ewig mehrsprachig zu denken ist so mühsam. Machen wir es doch gleich in einer Sprache. Aber es ist gleichzeitig auch problematisch und natürlich gehen dann auch gewisse kulturell-räumliche Spezifika verloren, was manchmal auch schade ist, vor allem auch in den Geisteswissenschaften. Insofern bin ich da ambivalent, was dieses Thema betrifft. Aber grundsätzlich die Usability finde ich gut. Eclipse, aber da bin ich natürlich auch insofern vorbelastet, weil Eclipse für mich täglich Brot ist, da ich die Umgebung kenne und mich natürlich auch orientieren kann. Ist vielleicht besser jemand anders zu fragen, wie der das sieht. Weil für mich ist das gewohnte Umgebung, daher habe ich keine Probleme damit.

**I22-10#I: Ja und Sie sehen als Barriere, was den Infrastrukturaufbau angeht, was Sie als Ressentiments nennen, auch, dass da eben Hemmungen bestehen, etwas einzustellen, damit ein Repository entstehen kann?** B: Ich habe auch länger mit der Frau R gesprochen. Das war für mich auch etwas überraschend, auf die Frage, welche konkreten Projekte – TextGrid läuft doch jetzt schon relativ lang – welche konkreten Projekte in TextGrid abgewickelt worden sind. Sagt Sie mir: Bis jetzt noch kein einziges. Also es gibt nur ...Testdatensätze da drinnen? Das war für mich doch überraschend nach der langen Laufzeit. Wir haben seit fünf Jahren auf unserer Plattform 18 Projekte abgewickelt, die wir hosten. Natürlich macht das uns das insofern Probleme, dass wir nicht mehr heute sagen können, ab morgen verwenden wir FEDORA 3.0, wir



verfahren noch immer mit FEDORA 2.2., weil wir einfach die Verantwortung für diesen Datenbestand haben. Aber trotzdem war für mich überraschend, dass es keine Projekte da drin gibt, keine realen. War ein bisschen überraschend ganz ehrlich.

**I22-11#I: Ja, ein Thema, das damit zu tun hat: Kollaborationsmöglichkeiten? Welche Bedeutung messen Sie diesen zu und...?** B: Auf dem ersten Paper zu unserer Absichtserklärung, was das betrifft, war das einer der wichtigsten Punkte. Wir sind immer davon ausgegangen, dass die Zusammenarbeit von unterschiedlichsten Institutionsebenen auch möglich sein muss, also die Zusammenarbeit zwischen Bibliothek und Fachwissenschaft z.B. was bei uns in O... was bei ihnen jetzt überhaupt nicht gelingt. Die Bibliothek sagt immer: Das ist unser Geschäft, die anderen sagen immer: Die Bibliothek versteht davon gar nichts, was wir tun. Es wäre aber wichtig, dass auf der einen Seite die Fachwissenschaftler die fachbezogenen Metadaten, im Sinne des Semantisieren des Inhalts einbringen und dass die Bibliothekare sozusagen die beschreibenden, deskriptiven, die Ressource erschließenden Metadaten pflegen. Da wäre es unmittelbar wichtig, sozusagen Formen der Zusammenarbeit zu unterstützen vom Tool. Auch in Projekten ist es immer wieder notwendig. Gibt ja viele Editonsprojekte, wo mehrere Personen unter Umständen an ein und demselben Dokument arbeiten. Auch da ist es natürlich wichtig, Dinge zu unterstützen. Ich weiß nicht inwieweit bei TextGrid daran gedacht ist, sozusagen SVN-Strukturen zu integrieren in der Versionierung von Dokumenten. Mir erscheint das ganz ein wichtiger Aspekt im Hinblick eines entsprechenden Tools.

**I22-12#I: Sie verstehen unter Informationsmodellierung auch Prozesse, die etwas abbilden, wenn ich das richtig verstanden hab? Also Visualisierungstechniken?** B: Modellierung kann dann den Zweck haben, in entsprechenden Repräsentationsformaten entsprechende Visualisierungstechniken zu unterstützen. Wir haben z. B. die M-Edition, die wir in Frankfurt vorgestellt haben, da haben Sie die Möglichkeit, in der Transliteration des Textes sich mit dem Cursor zu bewegen, und kontextbezogen erscheinen sowohl Editionscommentare, sich im Text bewegend, als auch an der Cursorposition der jeweilige Ausschnitt der Zeile. Also wenn Sie in einer bestimmten Zeile sind, sehen Sie unmittelbar an der Cursorposition vergrößert in der Lupenform sozusagen die Zeile aus der Handschrift, aus dem Faksimile. Das wird möglich eben über Modellierung, weil in dem Dokument sind sowohl grafische Information drin, SVG-basiert, als auch die ganze Transliteration abgebildet in diesem TEI-Dokument.

**I22-13#I: Angenommen diese Infrastruktur würde es geben. Es gibt Modellierungen und die Grid-Struktur, die ermöglicht ja, eine riesige Datenmenge darin unterzubringen.**

**Welche Möglichkeiten würden Sie als gut ansehen, da Orientierung zu schaffen, sich also da zurecht zu finden in dieser vernetzten Wissenschaftsumgebung?** B: Das schließt unmittelbar an das Vorhergesagte an. Ich habe von Objektmodellen gesprochen. Also ich glaube, das wäre eine ganz wichtige Form. Also „Projekte“ ist ein Ordnungskriterium. Im Projekt muss es mir möglich sein, sozusagen inhaltliche Konfigurationen zu bilden, im Sinne dieser Inhaltsmodelle. Ich kann erkennen, im System, eine bestimmte Art von Datenstruktur, die ich definieren kann, einmal definieren... Jetzt möchte ich mehrere solcher Objekte anlegen. Das ist für mich ganz ein wichtiger Aspekt sozusagen dieser Organisation von Daten in einem solchen System. **I: Die müsste dann aber auch individualisierbar sein. Sie haben ja modularen Charakter erwähnt.** B: Genau, ich müsste sozusagen solche Inhaltsmodelle in einem solchen System modellieren können, auch die Inhaltsmodelle, dass ich sagen kann: Ok, ich möchte... Jetzt brauch ich eine neue Form von Dokument, jetzt kann ich sagen: Ok, diese und diese Form von Datenströmen macht diese Form von Dokument aus, diese und diese Operationen operieren auf diesen Datenströmen, so hängt das zusammen, das muss ich modellieren können und dann ja... **I: Auch als nutzender Fachwissenschaftler?** B: Ja, das ist die Frage sozusagen. Ich glaube, dass man Fachwissenschaftler wahrscheinlich immer beraten sollte, aber es muss eine Ebene geben, die im System solche Möglichkeiten zur Verfügung hat. Das macht man auch so. Das machen natürlich nicht die Fachwissenschaftler, sondern die kommen zu uns sozusagen und sagen... Wir stellen im gemeinsamen Gespräch fest: Das ist die Erfordernis in einem solchen Projekt und dann designen wir auf Basis dieser Erfordernisse das entsprechende Inhaltsmodell. Wir schauen: Reicht eines der vorhandenen Inhaltsmodelle aus? Wenn nicht erstellen wir ein neues oder erweitern ein bestehendes im Sinne der Vererbung von Eigenschaften, also wirklich im objektorientierten Sinn.

**I22-14#I: Welche Möglichkeiten sehen Sie auch anhand von anderen Infrastrukturen, die es schon gibt, den Aufbau zu betreiben? Sie kennen sich auf dem Gebiet ja gut aus. Jetzt mal nicht TextGrid-spezifisch, wie würden Sie das beschreiben?** B: Ja, also wie gesagt: Wir verwenden in unserem Umfeld lange Zeit solche Workflows, wo wir immer davon ausgegangen sind von der Prämisse: Wir müssen sozusagen... Wir möchten das in die gewohnte Arbeitsumgebung der Fachwissenschaftler irgendwie hineinimplementieren – also Workflows, die so ausschauen, dass sie mit Word oder OpenOffice ihre Dokumente erfassen können und ihre Daten produzieren können. Natürlich, das kommt bei uns immer wieder vor, dass wir oft vor das Problem gestellt sind, Software-Archäologie zu betreiben. Es kommt jemand zu uns, ja wir haben

da eine Access-Datenbank, die ist vor zehn Jahren angelegt, niemand weiß mehr, was das ist und für was das überhaupt gut ist. Wir wissen nur: Das ist die Bildsammlung X. Dann gehen wir her, versuchen das zu analysieren im Sinne der Software-Archäologie, und generieren dann aus diesen Strukturen programmbasiert entsprechende Objekte für unser System. Das ist auch ein Weg, der da hinführt. Im Prinzip sage ich, es muss über verschiedene Quellen möglich sein, in einem solchen System auch Masseningestierungen durchzuführen. Das versuchen wir. Wir haben jetzt wieder eine Bilddatenbank der Kunstgeschichte. Alle Bilder, die für eine Lehrveranstaltung „gekonnt“ werden müssen in Anführungszeichen, haben wir aus einer Exceltabelle, in einem Masseningest sozusagen, in dieses System übernommen. Also auch das muss unterstützt werden – unterschiedlichste Zugänge, auch die proprietären. Also ich sage jedem, Access jetzt zu verwenden für solche Projekte ist Wahnsinn, aber natürlich passiert es immer wieder. Es muss eben Möglichkeiten geben, diese proprietären Formate in ein entsprechendes nachhaltiges Format zu überführen – und das ist auch ein Gutteil in unserer Arbeit immer wieder.

**I22-15#I: Ja, wenn man sich jetzt vorstellt, man muss Software-Archäologie betreiben. Dort gibt es Projekte, die machen etwas ähnliches, haben sich aber ihre eigenen Tools programmiert. Das ist ja sehr stark verteilt und sehr differenziert. Wie könnte man da sonst vorgehen? Sie haben jetzt ein Institut, das sich damit beschäftigt, aber was wäre der Weg dahin, das jetzt wirklich zu vernetzen wider alle Ressentiments?** B: Ich kann nur sagen, wie das bei uns ist. Wir hatten ja auch die Situation – sogar im Haus bei uns. Jeder Fachwissenschaftler ist gekommen und hat ein eigenes Projekt erfunden. Und der Programmierer hat sich einmal hingesetzt und hat dann einmal dieses und einmal dieses und einmal jenes verwendet, bis ich gesagt habe: Das ist ein Weg, der macht keinen Sinn, der führt in nicht wartbare Strukturen. Also im Bezug auf unsere Fakultät, kann ich jetzt sagen, sind wir zu so einer Integrationsfigur diesbezüglich geworden. Wenn jemand zu uns kommt, dann denken wir in die Struktur, die wir entwickelt haben, hinein. Es sind also keine technologischen Fragen, sondern es sind politische Fragen, die entschieden werden müssen. Es sind natürlich auch Fragen der Community. Eine Universität ist eine geschützte Werkstätte. Das ist ein Haufen Individualisten. Jeder glaubt, er muss die Welt neu erfinden. Diesen Prozess, diesen Wahnsinn muss man sozusagen durchbrechen. Das gelingt natürlich, wenn es auch institutspolitisch geht. Unser Dekan sagt z. B. schlichtweg, wenn jemand ein Projekt denkt und er denkt es nicht mit uns, dann kriegt er kein Geld. Das ist ganz simpel, dann kommen die Leute natürlich zu uns. Es braucht also die Strukturen, Strukturen, die auch adäquat sind. Die Strukturen müssen so flexibel sein, dass es

nicht zu enge Vorgaben gibt. Also ein System, das alles festschreibt, eine relationale Datenbank, wenn man es konkret formuliert, ist meiner Meinung nach ungeeignet. Das XML-Datenmodell als so weit gedachte Form ermöglicht so viel Flexibilität, dass ich unter ein und demselben Dach ganz unterschiedliche Dinge denken kann und das macht es dann auch wirklich möglich. Aber gleichzeitig braucht es dann – auch das sind in der Scientific Community noch Prozesse, die im Gang sind, und die noch lange Zeit brauchen werden – es braucht eine gewisse Form der Standardisierung, dass sich das durchsetzt und man muss den Fachwissenschaftlern auch zeigen, dass das für Sie Erleichterung bringt, dass es sie gleichzeitig nicht einschränkt, dass es aber dadurch flexibler für sie wird und dass es ihnen auch Erleichterungen bringt. Es ist einfach eine politische Frage. **I: Da müssten sich also Integrationsfiguren unterhalten und auf der Ebene müssten erst Strukturen entstehen, die dann in die Fachwissenschaften...** B: Ja, ich glaube, das ist einfach notwendig, dass in den Disziplinen das Bewusstsein entsteht, es macht keinen Sinn, wenn ihr die Welt neu erschafft, sondern wir müssen uns zusammensetzen und z. B. mal definieren, was unsere Erfordernisse an eine Annotationssprache sind, und das verwenden wir dann. Das heißt ja nicht, dass das in Stein gemeißelte Dinge sind, das kann man ja dann auch wieder sozusagen modifizieren. Aber wichtig ist, dass sie diesen Weg beschreiten und diese Form der gemeinsamen Standardisierung. Da ist man in verschiedenen Disziplinen noch weit weg davon. Es gibt Ansätze, aber es wird noch lange dauern, da bin ich optimistisch, dass das noch lang dauert.

**I22-16#I: Ein anderer Ansatzpunkt wäre ja, das Thema in der Lehre zu platzieren, tun Sie ja. Welche Bedeutung sehen Sie darin? Wie kann man das gut umsetzen, was ist da ihr Ansatz?** B: Ja, das ist bei uns mittlerweile ein zentrales Instrument, d. h. diese Projekte, die wir da umsetzen, die sind auch nur deswegen möglich, weil wir so ein Gesamtkonzept haben: Projekt, Strukturprojekt, einzelne fachwissenschaftliche Projekte in Verbindung mit Lehre. Und dort, wo diese Integration funktioniert, funktioniert es auch gut. Das heißt, im Regelfall gehen wir die Projekte so an, dass die Projektmitarbeiter eines Projektes zu uns kommen, die Lehrveranstalter, und dort finden wir dann eine gemeinsame Sprache für das, was sie dann später erarbeiten. Wenn das so funktioniert, funktioniert es gut. Natürlich kommen auch zunehmend Studierende. Wir haben also vor drei Jahren begonnen. Da saßen immer da in der „Einführung in die Datenmodellierung und Textmodellierung“ drei bis vier Studierende. Wir sind heute soweit, dass wir voll sind, also zwanzig. Wir bieten diese Veranstaltung so als Basisveranstaltung jedes Semester an. Also man sieht zunehmend Interesse, es wird auch wahrgenommen zunehmend und

langsam, ganz langsam wird..., fließt das dann natürlich etwas zurück. **I: Ja, also wird auch Nachwuchsgewinnung betrieben in dem Bereich?** B: Natürlich, natürlich.

**I22-17#I: Inwiefern sehen Sie Möglichkeiten, durch digitale unterstützte Forschung neue wissenschaftliche Erkenntnisse und neue Forschungsformen zu gewinnen?** B: Ich glaube, das auszuloten ist auch noch Aufgabe der Disziplinen. Ich sehe es unbedingt. In verschiedensten Bereichen müsste man da wahrscheinlich zu weit ausholen. Also in diesen Projekten, die wir haben, passiert es immer wieder. Das geht vom Sprachkorpus der slawischen Sprachen, in dem es möglich ist, nicht nur noch textbasiert zu suchen, sondern auch nach grammatikalischen Strukturen, weil dieses Korpus entsprechend ausgezeichnet ist. Ich kann dort also eine bestimmte Satzkonfiguration als Suchbegriff eingeben und finde dann Realisierungen. Da passieren unmittelbar Diplomarbeiten, da passiert Forschung, da passiert eine Verbindung von Informationstechnologie und Wissenschaft und da zeigt sich auch, dass diese Formen neue Formen der Erkenntnis möglich werden lassen. Das auszuloten ist Aufgabe der Zukunft würde ich meinen. **I: Kennen Sie da ein konkretes Beispiel aus der germanistischen Editionsphilologie. Ist Ihnen so was schon begegnet?**

B: Ich muss gerade einen Artikel schreiben und habe keine Zeit dazu, diesen Artikel zu schreiben für die Editio. Ich glaube, dass dieses Projekt, das wir in Frankfurt vorgestellt haben, genau in diese Richtung geht. Wir haben dort den Text auf Style-Glyphen-Ebene hinunter modelliert, d. h. Sie können in diesem Datenbestand nach Verzierungsformen, Diakritika-Formen, nach verschiedenen..., nach Superskripten, Subskripten und ähnlichen Dingen suchen und ich glaube, sowohl in der didaktischen Aufbereitung solcher Inhalte als auch in dem wissenschaftlichen Herangehen, dass das entsprechende Möglichkeiten bietet. Wenn ich z. B. eine Handschrift habe, in der die verschiedenen Fassungsvarianten modelliert sind, dann kann ich auf Knopfdruck eine Synopsendarstellung generieren, dann werden das unmittelbar Instrumente, die man in der Forschung einsetzen kann. Voraussetzung ist, dass man sich mit diesen Themen in der jeweiligen Fachcommunity auseinandersetzt. Dann kann man das Ding ja auch stärker standardisieren und sagen: Ok, TEI ist ja ein sehr weiter Rahmen, den ich zunächst einmal auch für mich und meinen Anwendungsfall auch einschränken muss. Wenn das aber mal disziplinemäßig einmal passiert ist, wenn also die Germanisten in einem bestimmten Bereich sagen: Wir verwenden die TEI so und so und schreiben das fest, kann man auf dieser Basis auch standardisierte Abläufe definieren. Repräsentationsformate von Text können dann nicht nur Repräsentationsformate im Sinne PDF, HTML, sondern können ja auch Analyseansichten auf den Text sein, die sie produzieren und

wenn man das standardisiert, kann man Fachwissenschaftlern Werkzeuge in die Hand geben, mit denen sie dann wirklich auch wissenschaftlich arbeiten können.

**I22-18#I: Ja, da wäre ja dann auch ein interdisziplinärer Anknüpfungspunkt gegeben.** B: Unbedingt! Das sage ich immer... haben wir auch immer schon gesagt. Wenn z. B. der Soziologe seine Interviews TEI-basiert modelliert und seine Interpretationskategorien dahinein verarbeitet, dann produziert er..., würde er theoretisch, wenn man das jetzt interdisziplinär denkt, würde er für den Fachwissenschaftler, für den Sprachwissenschaftler bereits ein Textkorpus generieren, den er nur nehmen braucht und mit seinen Formen der Auszeichnung, Part-of-Speech-Tags oder ähnliches, versehen muss. Und er kann den unmittelbar sozusagen weiterverwenden, dann greift das ineinander und ermöglicht sozusagen in diesem Sinne Synergien.

**I22-19#I: Zum Abschluss eine zusammenfassende Frage? Worin sehen Sie den größten Fortschritt, das größte Potential in TextGrid und ähnlichen Systemen? Und wo sehen Sie noch das größte Manko, die größte Barriere?** B: Ich glaube, der größte Fortschritt besteht für mich darin, wenn es gelingt, sozusagen Workflows und Arbeitsweisen neben der Form der Modellierung noch Workflows und Arbeitsweisen in den Fachdisziplinen zu standardisieren. Ich glaube, dass das ein großer Mehrwert werden kann. Weil auch immer dieses Neuerfinden... Wenn man wirklich Werkzeuge von der einen Fachdomäne in die andere austauschen könnte, weil etwas, was der Sprachwissenschaftler zur Auszeichnung verwendet, kann genauso für den Soziologen oder die Soziologin interessant sein, und ich glaube, es macht wenig Sinn, die Welt ständig neu zu erschaffen. Wenn hier über TextGrid so eine Communitybildung und Standardisierung, was das betrifft, gelingt, erscheint mir das sehr viel. Als ich das Projekt das erste Mal gesehen habe, habe ich sofort gesagt, das macht unmittelbar und unbedingt Sinn, weil es eben neben ..., eben zur Standardisierung der Herangehensweise an bestimmte Probleme in der Scientific Community führt, was nicht heißt, dass alle das Gleiche tun. Es ist ein Rahmen, der sehr flexibel ist und sehr viel ermöglicht, gleichzeitig braucht nicht jeder die Welt neu erschaffen.

**I22-20#I: Das wäre also das größte Potential. Und was sehen Sie jetzt noch als größtes Manko oder als größtes Problem, als größte Barriere in dem Bereich? Die TextGrid-Idee haben Sie jetzt gesagt, halten Sie für ganz wichtig, aber wie ist das....?** B: Seien Sie sicher, dass, wenn wir das manchmal beobachten, dass für Fachwissenschaftler schon die Vorstellung, dass die Daten nicht am eigenen PC, sondern im Netz liegen, nämlich auf einem Server, den man zumindest noch lokalisieren kann, das ist schon oft ein Problem. Weil, es muss unbedingt an meinem PC liegen, sonst könnte es ja abhanden kommen. Wenn ich jetzt noch dran denke, dass

eine Fachwissenschaftlerin sagen muss, das liegt irgendwo in der Welt sozusagen – es ist ein Horrorszenario. Also ich glaube, dass die Akzeptanz diesbezüglich, ja dass das sicher Probleme macht. Und für mich ist das ja... Ich weiß nicht, ob ich da hineininterpretiere, aber der Ausdruck, dass es noch kein Projekt gibt, das real umgesetzt ist, der Ausdruck von dieser Problematik... so sicher diese Cloud-Strukturen sind. Das ist immer noch ein Akzeptanzproblem und das muss man glaube ich auch ernst nehmen, man muss Wege finden... Es gibt Möglichkeiten des Exports. Man muss Möglichkeiten finden, sozusagen die Daten auch herauszunehmen, so dass der Fachwissenschaftler sagt: Ok, ich kann das auf eine CD brennen, das ist ganz wichtig für die Zufriedenheit, für die kognitive Zufriedenheit im weitesten Sinne, dass das möglich ist. **I: Das ist ja dann auch ein Problem der Publikation, also wie kann man das sichtbar machen, wie ist das zitierbar, verortbar.** B: Genau. **I: Müsste in einem dann gelöst werden, oder?** B: Ja, ich glaube, diese Geschichte mit Persistent Identifiers und Möglichkeiten, das persistent zu zitieren, das ist ein ganz ein wichtiger Aspekt, das war bei unserem Projekt immer ein wichtiger Aspekt. Also es muss möglich sein, digitale Ressourcen nach dem Prinzip der Print-Publikationen zu zitieren, unabhängig abstrahiert von der zugrundeliegenden Serverstruktur oder ähnlichen Dingen. Es muss möglich sein, in einer Print-Publikation anzugeben: Da ist es. Das ist auch wichtig, nicht nur projektbezogen – das gibt es ja bei vielen Projekten – sondern es müssen verschiedene Identitäten/Entitäten zu zitieren sein, es muss eine Folioseite, eine Münze, die beschrieben ist in einer Münzdatenbank, einen Abklatsch, den muss ich effektiv zitieren können. Das ist ein wichtiger Aspekt, so dass ich wirklich in einer Verschriftlichung dieser Datenbank sozusagen einfach auf diese Ressourcen, Entitäten, in irgendeiner Form referenzieren kann. **I: Das geht ja in die Versionierung hinein, dass man sehen kann, wer hat was getan. Das ist noch ein weites Feld.** B: Ja genau. Ja, das ist noch ein weites Feld, noch viel zu tun. **I: Gut, herzlichen Dank.**