

I25-1#Interviewer: Inwiefern arbeiten Sie denn im Bereich E-Humanities bzw. digital unterstützt? Wie würden Sie da ihre Erfahrungen beschreiben? Befragte/r: Nahezu 100%.

Das ist der Grund, warum ich hier in der H Bibliothek angestellt worden bin, weil ich das in K auch in einer Ausbildung sozusagen genossen habe, in Anführungszeichen. Dort gibt es seit ein paar Jahren einen Studiengang „Historisch-kulturwissenschaftliche Informationsverarbeitung“, allgemeiner gesagt „Informationsverarbeitung in Geisteswissenschaften“ und ich bin in dem Teilbereich „Historisch-kulturwissenschaftliche Informationsverarbeitung“ bei T tätig gewesen. Studiert habe ich das nicht, weil man das noch nicht konnte. Als ich fertig war, ist aber dieser Studiengang eingerichtet worden und dann habe ich dort angefangen zu arbeiten und habe also mit so Handschriftendigitalisierung, Codierung von Daten, Metadaten, Schemata, also Metadatenformate ganz allgemein, TEI im Speziellen, extrem viel zu tun – und betreue hier in der H Bibliothek, wenn man also auf das Türschild oder meine E-Mail-Adresse schaut, digitale Editionen, im weitesten Sinne digitale Erschließung und das bedeutet, dass ich Kollegen, die z. B. Handschriften katalogisieren, betreue darin, das im Format TEI zu tun und dann diese Informationen in Datenbanken zu überführen nach Manuskripta Mediaevalia z.B oder auf der eigenen Homepage in Datenbanken zu publizieren; betreue die Kollegen, die digitale Editionen machen, habe auch selber ein Projekt zur digitalen Edition, bin Mitglied im Institut für D. Also ja, das ist ...mein ganzes Berufsleben ist eHumanities im Prinzip. Und dazu kommt noch, das bauen wir aber gerade erst auf, dass in G ja ein Studiengang dazu aufgebaut wird, das haben Sie vielleicht mitbekommen in anderen Interviews, und wir sind ein offizieller Partner dieses Studiengangs an der H Bibliothek. Ob das mal mit Lehrveranstaltungen einhergehen wird, muss man abwarten. Aber in verschiedenen Zusammenhängen, z. B. für das Aufbaustudienprogramm Skripto, in dem Kollegen der Umgang mit mittelalterlichen Handschriften gelehrt wird, da betreuen wir, insbesondere ich mit F in mittlerweile Göttingen zusammen, das Modul „EDV-Kenntnisse“.

I25-2#I: Also in allen Bereichen eHumanities. Welche Erwartungen hatten Sie vor diesem Hintergrund an TextGrid und den Workshop? Was hat Sie dazu bewogen, das kennenlernen zu wollen? Welche Erwartungen hatten Sie? B: Nun ja, TextGrid ist mir natürlich von Anfang an ein Begriff gewesen. Ich habe das mitverfolgt, wie es aufgebaut worden ist und hatte auch schon vorher an Veranstaltungen von TextGrid teilgenommen. Da ging es z. B. in D, von J organisiert, mal um eine Festlegung, was digitale Edition und TEI und so etwas

anbelangen sollte. Und da war ich zu diesem Treffen auch schon mal hingefahren, d. h. TextGrid und die beteiligten Personen sind mir seit Langem und sogar eigentlich auch ganz gut bekannt. Ich war mir aber nicht bewusst, wofür die Software eingesetzt wird, das TextGridLab. Ich hatte mir das zwar schon mal heruntergeladen und mir einen Account besorgt, aber da noch nicht richtig mit gearbeitet und wollte die Tatsache, dass ich sowieso bei der Tagung gewesen bin, dann dazu nutzen, um eben auch das TextGrid nochmal dort kennenzulernen. Und das war ja insofern ganz zielführend, dass wir ja dann damit gearbeitet haben, um zu sehen, was man damit tun kann. Das waren meine Erwartungen: zu sehen, was da tatsächlich programmiert, hergestellt wird und wofür man dieses Lab auch benutzen kann. **I: Hatten Sie da auch schon vorher Vorstellungen, Vermutungen, wozu man das Lab benutzen kann? Inwiefern hatten Sie da konkrete Erwartungen?** B: Naja die Schlagwörter, was dort entwickelt werden sollte, waren ja bekannt. Dass es dort also eben um XML-Editoren, digitale Editionen im weitesten Sinne gehen sollte. Vor dem Hintergrund dieses Grids gibt es eben eine Möglichkeit, dort Daten zu akkumulieren oder dezentral zu bearbeiten. Das war mir schon klar. Ich habe aber, wie schon gesagt, das noch nicht in Realität benutzt gehabt oder noch nicht so gesehen. Insofern war das die Möglichkeit, das zu tun.

I25-3#I: Und welchen Gesamteindruck hatten Sie jetzt nach diesem ersten Test? B: Naja, die Ziele sind hoch gesteckt, wie meistens bei den Projekten. Das mit der Realisierung ist etwas langsamer als man das so erwartet und dazu kommt vielleicht noch, dass als versiertem Benutzer es dann relativ leicht fällt, schnell an die Grenzen heranzukommen und die Probleme oder sozusagen noch offene Enden dann zu entdecken. Das ist ja während des Workshops auch geschehen, dass wir doch eine ganze Reihe von Anregungen oder Fehlermeldungen dort hinterlassen konnten, wo Dinge eben nicht so funktioniert haben, wie man das erwartet hätte vielleicht als Benutzer und wir das auch unmittelbar an die Kollegen, die zur Betreuung dieses Workshops dort dabei waren, direkt weiterleiten konnten, so dass die auch eine relativ lange Liste an offenen Punkten dann... ja sozusagen mit nach Hause gefahren sind. Ich stelle mir vor, dass das vielleicht für den ein oder anderen Forscher interessant sein könnte, dieses Lab später mal zu benutzen, wenn die Funktionalität noch weiter ausgedehnt ist. Für mich persönlich kommt es vielleicht weniger in Frage, obwohl ich das weiter mit Interesse verfolgen werde. Weil ich selber entwickle – in Anführungszeichen –, also Dienste in der H Bibliothek z. B. aufzubauen. Ob ich das TextGridLab deswegen als Ressource, als Arbeitsumgebung mal nutzen werde, mag dahingestellt sein. Aber sich damit auszukennen, dass auch gegebenenfalls Kollegen, die

vielleicht außerhalb ihrer Tätigkeit hier an der H Bibliothek damit arbeiten wollen... Das wird sicherlich interessant bleiben.

I25-4#I: Was sind denn die entscheidenden losen Enden, die Sie festgestellt haben? B: Ach, es sind teilweise aktuelle Fehler gewesen, wo das Programm nicht so reagiert hat, wie man sich das vorstellte. Wo einfach noch Programmierfehler, oder -probleme drinsteckten. Es hat allerdings auch etwas damit zu tun, dass – glaube ich – dieses TextGridLab sehr weit geneigt ist, dem Benutzer entgegenzukommen, womöglich vollständig von der Kenntnis von XML und TEI und so etwas zu entlasten, obwohl ich einfach denke, dass man das nicht kann. Es geht nicht, digitale Editionen zu erstellen, ohne wirklich in den Quelltext, in XML, hineinzuschauen. Und deshalb sind wir einfach relativ schnell auf solche Probleme gestoßen, wie, dass bestimmte Aspekte, wo XML ja gerade problematisch ist: überlappende Strukturen, was kann ich wo hineinschachteln und so etwas, dass das auf dieser Oberfläche, die dem Benutzer dort geboten wird, nicht realisierbar ist, weil man dafür wirklich in die spitzen Klammern reingucken muss. Und das sind dann Dinge, wo ich es verstehen kann, dass man dem Benutzer weit entgegenkommen möchte, weil der vielleicht nicht gewöhnt ist, mit spitzen Klammern zu arbeiten und so etwas, wo ich aber sage, der Gegenstand, eine digitale Edition zu erstellen, ist so komplex, dass das bisschen Komplexität, in spitzen Klammern zu arbeiten, man auch zumuten muss, weil das Arbeiten sonst nicht erfolgreich sein wird. Weil genau da, an diesen Stellen, z. B. überlappende Strukturen, Seitenfälle oder andere Dinge, die in der Quelle drinstecken könnten, wenn man die codieren möchte, dann muss man das im Quelltext tun. Und da nützt es nichts, irgendwie eine Draufsicht zu haben, wo dann möglichst alle spitzen Klammern verschwinden. Das halte ich da für nicht zielführend. Und dieses relativ schnell festzustellen, das war z. B. eine der Sachen, die schnell offensichtlich wurden, aus meiner Kenntnis heraus wie XML funktioniert, wo die Probleme sind, man relativ schnell bei so einem TextGridLab feststellt, dass die Probleme nicht gelöst sind, im Gegenteil sogar vielleicht Probleme verschleiert werden, die dem Benutzer bewusst sein sollten, wenn er damit arbeitet. Dass da Grenzen sind, das bleibt so ein bisschen versteckt und wurde an der Funktionalität auch nicht deutlich. Sondern man führt dieses Lab, diesen XML-Editor schnell an eine Stelle, wo man denkt, wieso reagiert er jetzt nicht so, wie er sollte, nämlich dass er entweder eine Fehlermeldung rausgibt oder das macht, was er soll. Ja, das ist also so ein prominentes Beispiel. Wie gesagt, wir haben da eine ganze Reihe von Dingen herausgefunden, z. B. auch mit befreundeten Kollegen, die genauso erfahren sind wie ich, zusammen gesessen haben und ausgelotet haben. Man kann nicht mehr alles nachvollziehen. Aber das ist so einer der

markantesten Punkte gewesen für mich, wo ich denke: So eine Arbeitsumgebung ist ja sehr schön, aber man muss es dem Benutzer nicht zu einfach machen, weil es nicht geht, vom Inhalt her.

I25-5#I: Gibt es noch weitere, ganz konkrete Funktionalitäten, die Ihnen gefehlt haben? B:

Da ich jetzt auch nicht mehr sehr viel damit rumprobiert habe, ist mir jetzt da nicht sehr viel erinnerlich. Was nicht so voll mit Funktionalität gefüllt zu sein schien, waren ja dann Fragen, dass...tatsächlich der verteilte Zugriff auf Dokumente, wie man die schützt, während die in Bearbeitung sind, sind glaube ich noch nicht ganz realisiert gewesen sind. Dann auch die Frage: Wie können eigentlich andere die Dinge nachnutzen? Da schien mir noch was offen zu sein. Es ist doch noch zu sehr in der Betaphase oder vielleicht sogar noch in der Alphaphase drin, dass man das tatsächlich für die Arbeit nutzen könnte. Und wie gesagt, es nur als Speicherort für Daten zu nehmen, dafür brauche ich es jetzt gerade nicht, weil ich sozusagen die Ressourcen der H Bibliothek und anderen Institutionen nutzen kann, um meine Dinge da zu tun, was ich tun muss. Entsprechend... Ich glaube auch, dass dieses TextGridLab und der Editor selber als eben diese Rückfallpositionen am ehesten wohl für den Forscher interessant sein werden, der eine digitale Edition erstellt, aber nicht an einer großen Universität z. B. beheimatet ist, die dann entsprechend Speicherplatz und andere Ressourcen zur Verfügung stellen kann – also vielleicht auch private Projekte, für die wird das am ehesten in Frage kommen. Und ansonsten wird TextGridLab irgendwann, wenn es erfolgreich sein möchte, sowieso auf eine Schiene kommen müssen, wo Ressourcen, wie z. B. die der H Bibliothek, eingebunden werden, egal wo sie sind, dass man sie dann unter dieser Oberfläche verlinkt, verknüpft, was auch immer damit tut. Das ist ja auch alles noch nicht irgendwie funktionabel, da wird ja noch mehr an dem Lab und der Speichermöglichkeit dort selbst gearbeitet. Das sind natürlich auch noch offene Enden.

I25-6#I: Könnten Sie auf den Punkt verteilter Datenzugriff und Nachnutzung nochmal genauer eingehen, wie stellen Sie sich das vor? Was würde da einen Mehrwert bieten? B:

Naja, ich versuche, in der H Bibliothek ein Gefühl auch sozusagen dafür einzubringen, dass solche Dinge wie die Freiheit von Daten, Open Access oder ähnliches, in der Zukunft wahrscheinlich eine wichtigere Rolle spielen werden, auch die Möglichkeit, Daten auszutauschen über OAI-Schnittstellen und Ähnliches... sind aber eigentlich auch im Bibliothekswesen im Allgemeinen schon recht weit fortgeschritten. Und wenn ich z. B. jetzt gerade dabei bin, eine Handschriftendatenbank aufzubauen, die unsere Handschriftenkatalogisate und – digitalisate und die bibliographischen Informationen zu den Handschriften der Öffentlichkeit zur Verfügung

stellt, dann wird damit verknüpft sein: eine Schnittstelle, über die man diese Daten auch beziehen kann, z. B. weil wir ein Kooperationsprojekt, ein europäisches Projekt gerade angefangen haben, bei dem diese Daten nach Europeana Regia, nein, nach europeana eingespielt werden. Europeana Regia ist das Projekt, was wir betreiben, das ist der Projekttitel. Aber in diese europäische Bibliothek, digitale Bibliothek Europeana, werden Daten ausgeliefert und dafür braucht man Schnittstellen, um Daten an diesen Schnittstellen zur Verfügung zu stellen, sie in einem standardisiertem Format, TEI mit Dublin Core, was auch immer, zur Verfügung zu stellen. Das sind meine... das ist auch mein Hauptarbeitsfeld, in dem ich mich hier umtue. Und da, an solchen Punkten, werden dann auch solche Angebote wie TextGrid ansetzen müssen, dass sie als entweder Akkumulation, als Harvester oder was auch immer auftreten und dann aber natürlich Möglichkeiten bieten, wie man die Daten anreichert und ggf. vielleicht auch wieder in Institutionen zurückspielt, so dass dann auch der Institution bewusst ist, ah, da ist mit unseren Daten was gemacht worden, das können wir für uns selber auch nutzen, auch dieses Zusammenspiel, lokale Angebote, die halt bestandsbezogen sind, und überregionale Forschungsinteressen, die irgendwo in der Welt stattfinden können, wo da etwas beigetragen wird. Das wird der Weg der Zukunft sein. Und aber darauf müssen sich auch beide Seiten einrichten, sowohl die lokalen Anbieter als auch solche Grid oder Harvester oder wie auch immer man diese Funktion benennen möchte, dass dafür eine Systematik geschaffen wird, die die Kommunikation zwischen diesen beiden Ebenen in Gang bringt und stabilisiert. Wenn man das auch in Hinsicht auf Langzeitsicherheit, wissenschaftliche Absicherung, wie zitierfähig sind Dinge und ähnliches, dann zurückbinden möchte. Weil das sind ja zur Zeit die Fragen, bei denen die Wissenschaftler am ehesten sagen: „Oh, das Digitale ist mir nicht sicher genug, das ist nicht zitierfähig, ich weiß gar nicht, was für eine Qualität das hat, wie ist das auch mit der Qualitätssicherung, Wikipedia, das kann man ja nicht zitieren“.. und solche Dinge.

I25-7#1: Ein großes Problem. Was Sie jetzt aber auch angesprochen haben, ist ja vor allem auch der Infrastrukturbereich, verstehe ich das richtig? Da sehen Sie die Hauptaufgabe von solchen Angeboten wie TextGrid? B: Ja, so verstehe ich überhaupt die Grid-Initiative, denn das ist ja, denke ich mal, gedacht als irgendwie eine Art Backbone bzw. eine vernetzte Struktur, die Ressourcen zusammenbringt oder verteilte Ressourcen aufbaut, um halt in rechnerintensiven Bereichen der Mathematik, Biologie oder ich weiß nicht was, dann eben Kapazitäten zur Verfügung zu stellen. Und solch ein textorientiertes oder eHumanities-orientiertes TextGrid müsste dieselbe Funktion für die Geisteswissenschaften haben, dass man entweder Ressourcen,

die verteilt existieren an Bibliotheken, Universitäten, Forschungsinstitutionen, Max-Planck-Institutionen und so etwas, dass man die auf irgendeine Art und Weise miteinander verknüpft und einer Forschungsöffentlichkeit zur Verfügung stellt und darauf basierend eben Tools, wie man damit umgeht, wie man annotiert, wie man selektiert, wie man Dinge tut – wie das bei TextGrid auch angedacht, aber noch nicht realisiert ist – dass dort z. B. Lemmatizer und andere sprachwissenschaftlich orientierte Tools dann angelagert werden. Das, denke ich mir, müsste die Funktion des TextGrid sein, irgendwelche Tools, Oberflächenverlinkungsmechanismen zu organisieren, damit der einzelne Forscher eben auf den Daten, die womöglich schon irgendwo anders erstellt worden sind, aufbauen kann und nicht jedes Mal das Rad von Neuem erfunden und die Transkription wieder von Neuem angefertigt werden muss, sondern eine womöglich bestehende so verändert wird, wie das die nächsten Forschungsinteressen bedürfen. **I: Ja das wäre ja jetzt in erster Linie eine Vernetzung von Inhalten, oder?** B: Ja, ich würde das nicht so getrennt sehen. Also Inhalte kann man ja sehr weit fassen. Einerseits gehören da Ressourcen wie Digitalisate dazu, die zugänglich gemacht werden können, auch über eine solche Oberfläche, Initiativen wie der DFG-Viewer oder Ähnliches, die halt eine Oberfläche anbieten für viele Ressourcen; oder auch das, was mit dem ZVDD mal angefangen hat, eben ein Recherche-Einstieg für alle Digitalisate, die in Deutschland von der DFG finanziert worden sind, das steckt auch noch in den Kinderschuhen. Aber darauf wird es ja hinauslaufen müssen, denn die H Bibliothek ist nicht die einzige Institution, die digitalisiert, aber ein Digitalisat ist ja heutzutage eigentlich die unterste Ebene eines komplexen digitalen Objektes – nenne ich es mal – digitale Edition. Da sind die Bilder ganz unten und dann hat man da Metainformationen, Metaebenen da drauf, die dann bis zum Volltext und Annotationsformen und vielleicht sogar noch irgendwelchen SVG-Anwendungen, dass im Bild lokalisiert wird, wo ist der Text eigentlich oder welche Phänomene finden hier statt, Vernetzungen mit kunsthistorisch-orientierten Datenbanken, die dann zu einer Suche zu Bildinhalten führen könnten oder Ähnlichem mehr. Also da würde ich den Informationsbegriff oder Ressourcenbegriff ganz weit setzen. Das können Bilder sein, das können Texte sein, das können Katalogeinträge sein, das können OPAC-Informationen sein, das können Wörterbücher sein. Da ist ja in der heutigen Zeit eigentlich keine Grenze gesetzt..., das kann eine Google-Maps-Anwendung sein. Das alles ist ja heutzutage mehr und mehr XML-basiert – ein sehr weiter Begriff für die Ressourcen, die da vernetzt werden können.

I25-8#I: Und welche Bedeutung messen Sie der Verknüpfung von Arbeitsprozessen zu, also der kooperative Aspekt? Sie haben ja jetzt als Beispiel genannt: Zusammenarbeit in einem

europäischen Projekt, das wäre ja dann ein weiterer Aspekt. B: Ja, dieses europäische Projekt ist natürlich in gewissen Teilen nur zu einem kleinen Teil wirklich kooperativ, weil das ein verteiltes Digitalisierungsprojekt ist, wo verschiedene Bibliotheken Handschriften digitalisieren und die über Europeana halt zur Verfügung stellen wollen. Und natürlich tauscht man sich dann aus über Datenformate und so etwas, aber es wird eigentlich von einer Person ein Aspekt bearbeitet. Aber kollaborativ könnte es dort werden, wo wir uns z. B. wie in der Überlieferung der neutestamentlichen Schriften, wie das in Münster z. B. betrieben wird, wo wir also Gegenstände haben, Texte haben, die sehr breit überliefert sind, wo aber der eine Forscher vielleicht sagt: Ich kann im Rahmen meiner Dissertation eben nur drei oder fünf Handschriften transkribieren und die anderen zwanzig Handschriften lasse ich mal als offene Enden da liegen und wenn sich noch jemand da dran machen möchte, weil in einer Bibliothek z. B. ein Praktikant eine Handschrift transkribiert oder was auch immer, dann müsste man diese Dinge in eine offene Edition einspielen können, um dann daraus eventuell neue Erkenntnisse für die Überlieferungsgeschichte eines Textes zu gewinnen. Ob das immer der Fall sein wird, ob auch die politischen Strukturen so etwas mal befördern, dass derjenige, der die digitale Edition angefangen hat, dann auch wirklich diese offenen Enden liegen lässt und man zu einem System kommt, wo z. B. in einem TextGridLab diese teilweise, sage ich mal, offene digitale Edition dann liegt und weitere Quellen hinzugebracht werden können und daraufhin dann dynamisch sich das Ganze erweitert, das wird man mal sehen. Das ist ja vor allem auch eine forschungspolitische Veranstaltung...wie finanziert man so etwas, wie wird so etwas auch mit Credits oder was auch immer belegt, um halt seine Punkte für den akademischen Werdegang dann irgendwie zu sammeln, seine Leistung da zu erbringen. Das sind natürlich Fragen, die nochmal über das Technische weit hinausgehen. Da wird sich die Gesellschaft, die Forschung irgendwie mal zu verhalten müssen. Ich könnte mir aber vorstellen, dass das eine Möglichkeit wird, wo man so etwas ganz gut machen kann, denn wir zumindest als Bibliothek sind da sehr dran interessiert, für unsere Handschriften z. B. die Transkription bei uns auch vorzuhalten und wenn irgendjemand anders hingehen sollte und sagt: Okay, eine Handschrift oder ein Text von dem Text, der mich interessiert, liegt halt in W, also lege ich meine Transkription zu der W Handschrift in W ab und die aus M wird in M abgelegt und die restlichen Handschriften, die vielleicht keinen Ort haben oder bzw. die im Ausland sind, die lege ich auf meinen eigenen Server in der Universität ab, für so etwas wären wir als Bibliothek sehr offen und das ist auch unser Weg, den wir gerade

beschreiten, zu sagen: Wir haben die Originale, also wollen wir auch alle Formen von Daten zu diesen Originalen bei uns aufbewahren.

I25-9#I: Gesetzt den Fall, das funktioniert jetzt so, dieses Modellbeispiel, das sie genannt haben. Wie findet man denn so etwas wieder? Nehmen wir mal an, dass dann eine dritte Partei da ist, die sich dann genau für so was interessiert, müsste ja im Grunde eine Struktur oder eine Zugriffsmöglichkeit sein, weil das weiß ja sonst jemand möglicherweise nicht, wo was abgelegt ist? Wie stellen Sie sich das vor? B: Naja, solange das objektbezogen ist... Dass es also bspw. wie hier eine Handschrift gibt, die in unseren Bestand gehört und wir die womöglich auch digitalisiert haben, dann gibt es ja schon einen Ort im Internet, in unserer digitalen Bibliothek, wo wir die Information, die wir zu dieser Handschrift haben, versammeln und so ist jedenfalls auch unser Gedanke, die Information zu strukturieren, dass nämlich alles, was zu einer Handschrift gehört, mit dieser Handschrift gemeinsam z. B. in einer Ordnerstruktur abgelegt wird, sodass also die Handschrift weiß, ich habe eine bestimmte Signatur, und in diesem Ordner, der diese Signatur trägt, da liegt dann z. B. eine Handschriftenbeschreibung oder auch viele Handschriftenbeschreibungen, da liegen dann Strukturinformationen, da liegen Transkriptionen, da liegen andere Informationsbausteine. So stellen wir uns das jedenfalls als Infrastruktureinrichtung vor, die Objekte hat, die mal Gegenstand der Forschung sein könnten. Und wir zumindest organisieren unsere Dinge so, dass wir das zumindest sehr nah an der Signatur unserer Objekte anlagern, d. h. unsere Einträge in der digitalen Bibliothek sind nach den Signaturen wie die Bücher als Regalobjekte organisiert. Alles was wir dazu haben, wird im Prinzip mit dazu gelegt, sodass das einfach auch in unserer Präsentation klar sein wird, was für andere Ressourcen wir zu einem Objekt tatsächlich haben. Dass das gegebenenfalls natürlich in einer Universität, die ungebunden von den Objekten arbeitet oder ohne die Objekte selber zu haben, sondern in einem Forschungsprojekt,...weiß ich nicht an der Universität F ist glaube ich ein Averroes-Portal, über einen mittelalterlichen Philosophen, wo die Überlieferungszeugen dokumentiert sind. Dass es also eben eine Datenbank gibt, mit den Handschriftenzeugen und dort auch teilweise einzelne Bilder liegen oder Handschriftenbeschreibungen, so etwas. Das wird man natürlich sehen müssen, wie man solche Ressourcen dann auch mit zusammenbringt. Da ist es mir bei allen Veranstaltungen, die ich so mache, auch bei Kursen, die ich gebe und so etwas, immer ganz wichtig, zu sagen, wenn jemand in dem digitalen Bereich arbeitet, sollte er sich halt schnell mit dem Gedanken vertraut machen, dass er auch ein Ressourcenanbieter sein wird. So bald man irgendwie einen digitalen Text irgendwo im Netz zur Verfügung stellt, ist man

potentiell eine Ressource, die von außen verlinkt werden könnte. Und deswegen sind solche Dinge wie eindeutige Identifier, dass zum Beispiel dieses signaturenbezogene oder, dass man nachweist, an welchem Ort liegt etwas, was ist das wirklich, der Gegenstand? Das ist mir relativ wichtig. Da kämpfe ich leider immer wieder auch mit einigen Anbietern immer wieder drum, weil gerne mal die informatisch einfachste Variante eingesetzt wird. Man setzt eine Datenbank auf und ein Datenbankeintrag wird dann mit ID=5743 bezeichnet, was das aber ist, ob das ein Handschriftenkatalog ist oder ein Digitalisat oder eine Edition oder was auch immer, das geht ja aus dieser Adresse nicht hervor. Und diese Ressource selber sprechend zu machen, ist mir persönlich ein ganz großes Anliegen, trifft nur nicht immer auf die offenen Ohren, wie man sich das wünschen würde. Denn wenn das so ist, dann kann man ja auch schön die Ressourcen von außen nutzen. Dann kann ich ja sagen: Ich weiß, in der H Bibliothek liegt eine Handschrift, von der ich weiß, dass da mein Text drin ist, und jetzt gehe ich einfach mal in die digitale Bibliothek und schaue mal nach, ob unter dieser Signatur denn auch ein Digitalisat vorhanden ist, das kann man nämlich bei uns auf diese Weise machen. Das ist eine ganz selbsterklärende, algorithmisch gebildete URL, die den Zugang zu der Handschrift, der digitalisierten, ermöglicht. Dann kann ich einfach als jemand, der diese Ressource nutzen möchte, hingehen und per Algorithmus diese Informationen der digitalen Bibliothek abfragen und gucken: Ist denn da etwas zu meiner Handschrift? Ja oder Nein? Wenn ja, dann kriegt man das relativ schnell raus. Und solche Identifikationen, die werden mal nötig sein, um den Überblick darüber zu behalten, was gibt es denn eigentlich alles? Denn, wenn ich weiß, es gibt verschiedene Projektportale, die mit Handschriften etwas zu tun haben, dann kann ich auf die Idee kommen und sagen: Ok, ich weiß zwar, dass in vielleicht W nichts zu dieser Handschrift vorhanden ist, aber ich kenne doch noch diese und jene Projektportale. Dann gucke ich doch noch da, ob da nicht vielleicht noch was ist. Es gibt ja viele Ressourcen für Handschriften, Manuskripta Mediaevalia oder das Handschriftenarchiv der Berlin-Brandenburgerischen Akademie der Wissenschaften oder der Handschriften-Zensus in Marburg oder und und und... Es gibt ganz viele Möglichkeiten, wo man nachschauen könnte: Sind da eigentlich nicht noch Informationen zu der Ressource? Und das wird man mehr und mehr automatisieren müssen, wenn man den Überblick behält. Und da kommt es aber auch drauf an, dass diejenigen, die die Informationen anbieten, die entsprechenden Möglichkeiten schaffen und nicht ihre Beiträge hinter ID 3567 verstecken sozusagen, sondern die Sache sprechend machen.

I25-10#I: Also sehen Sie da eine Möglichkeit drin, dass über eine digitale Infrastruktur wie TextGrid oder eine ähnliche Struktur der Überblick geschaffen werden kann? Oder wie müsste das organisiert sein?

B: Man würde sich das vielleicht wünschen, aber meine bisherigen Erfahrungen – und da sei mal ZVDD genannt – sagen eigentlich eher, dass die Projektforscher nochmal auf neue Dinge setzen. Man macht das eben, wie man sich das so vorstellt, und schaut nicht zurzeit drauf: Wie kann das, was ich hier mache, eigentlich nachgenutzt werden? Und bis dieses Denken sich durchgesetzt hat, würde es eventuell noch drauf ankommen, auf irgendwelche Dokumentationen oder auf solche Dokumentationsportale zu setzen, die einfach Informationen sammeln und mehr oder weniger zufällig davon Wind bekommen, dass eine neue Ressource existiert, die in ihr bestehendes System von Ressourcen einbaut und dann Metainformationen über diese Projekte oder Informationsangebote dann erstellt, sodass man dann so, wie das früher bei Yahoo oder so mit Inhaltsverzeichnis von Webangeboten organisiert worden ist, oder letztendlich über Suchmaschinen das gefunden werden muss, dass man seine Ressource dann so gut benennt oder die Inhalte so strukturiert, dass man den Text, den man da anbietet, dann doch findet. **I: Also wieder eine extra Portallösung, die sich genau mit diesem Problem befasst, müsste das dann sein?**

B: Naja Portallösung? Ob das jetzt eine Suchmaschine ist, die die Dinge auffindet oder ob das eine Oberfläche wie TextGrid ist, die einfach die Mitspieler in dem Bereich eHumanities z. B. kennt und weiß, an diesen oder jenen Institutionen, Universitäten, Bibliotheken, Archiven wird digital gearbeitet und ich kenne dann also die Ressourcen, die da zur Verfügung stehen, weil vielleicht OAI-Schnittstellen oder so etwas existieren, die dann abgefragt werden und ein Informationsportal aufgebaut wird. Das wäre sicherlich eine weitere Möglichkeit neben der Suche, so. Ob das so sein muss? Wird wahrscheinlich nichts daran vorbeiführen. Die Frage wird eher sein: Wie zugänglich, also wie gut sind Informationsressourcen identifizierbar, wenn sie verteilt auf irgendwelchen Servern liegen? Kann ich die wirklich nachnutzen, kann ich in den Quellcode reinsehen oder bekomme ich nur HTML ausgeliefert? Wie kann ich das nachnutzen? Das wird natürlich dann von der Angebotsseite ausgehen. Und wenn ich jetzt von der besten aller Welten ausgehe und mir denke: Oh, solche Angebote, egal ob das jetzt Katalogisate, Volltexte oder was auch immer sind, stehen unter Open Access, unter einer OAI-Schnittstelle zur Verfügung, dann ist es eigentlich auch egal, ob es eins, fünf oder zehn Portale gibt, die das anbieten. Was frei in diesem Sinne zur Verfügung steht, kann ja doch eh von allen abgeholt werden. Dann ist es eigentlich auch egal, an welcher Stelle das kumuliert. Ob das auch ein TextGrid ist oder sich noch weitere Mitspieler finden, die da ein Interesse daran entwickeln

und auch Nachnutzungsmöglichkeiten, Tools oder was auch immer dann aufsetzen. Das wird sich, wenn das interessant wird, sicherlich auch kommerzielle Mitspieler finden, wenn da genügend textuelle Ressourcen zur Verfügung stehen, dass da irgendwelche Wörterbuchverlage oder ich weiß nicht was die Rechercheanbieter für Familiengeschichte, oder... gibt es ja viele Möglichkeiten...

I25-11#I: Inwiefern sehen Sie denn Möglichkeiten durch eHumanities auch neue wissenschaftliche Erkenntnisse zu erlangen? B: Naja, die Möglichkeit schwebt mir natürlich zentral vor, weil das ist ja einer der Aspekte, wieso ich das überhaupt betreibe. Also die Frage ist ja nicht nur: Digitalisieren um des Digitalisierens Willen, damit man den Computer jetzt benutzt und rumspielen kann. Sondern die Frage ist natürlich dann: Wie werden die Ressourcen, die wir aufbauen, später mal genutzt? Das heißt, wenn wir jetzt digitalisieren, z. B. Handschriften, dann geschieht das einerseits unter einem Aspekt, die Dinge zur Verfügung zu stellen, damit halt der Zugriff auf diese Objekte leichter wird, dass eben auch weit entfernte Forscher sich da in den Beständen umsehen können. Aber der entscheidende Mehrwert wird erst dann entstehen, wenn die Bilder tatsächlich ausgewertet werden, auf eine maschinelle Art und Weise. Viele haben zwar ein Problem damit, wenn unser herkömmliches historisches Arbeiten damit auf den Kopf gestellt wird. Also man merkt auch hier im Hause immer wieder Vorbehalte dagegen. Aber es sind ja schlicht neue Fragestellungen, die man damit beantworten kann. Die sind aber immer noch irgendwie im Entstehen, hat man das Gefühl. Obwohl ja eHumanities im engeren Sinne ja schon seit 20, 30, 40 Jahren existieren, wird immer noch nicht so gearbeitet, dass neue Fragestellungen drauf aufgebaut werden. Dass man auf die Bilder halt eine OCR-Software laufen lässt oder statistische Kennzahlen über die Verteilung von Texten auf einer Seite, dann hingeht oder die Erkennung von verschiedenen Händen in Handschriften versucht oder eine statistische Auswertung dann tatsächlich macht für bestimmte Aspekte, wenn man jetzt Volltexte hernimmt. Es gibt natürlich Projekte, die so etwas machen, wie ein Camena in Mannheim oder dann ein Perseusprojekt, was eben auf griechische Ressourcen aufbaut, schon sehr weit sprachwissenschaftliche Fragestellungen auch aufsetzt. Aber trotzdem hat man immer noch das Gefühl, eigentlich ist das Arbeiten doch sehr traditionell. Die Ressourcen werden vor allen Dingen so genutzt, dass man über das Elektronische einen leichteren Zugang dazu hat und dann werden am Ende die Texte ganz normal gelesen und interpretiert, auf eine qualitative Art und Weise dann ausgewertet, wie man das als Historiker, Linguist oder wie auch immer gewöhnt ist.

I25-12#I: Würde Ihnen da ein Beispiel einfallen, aus ihrer Arbeit jetzt entnommen, das ein Angriffspunkt wäre für eine neue Art der Fragestellung, für ein neues wissenschaftliches Vorgehen? B: Ja, ein paar Beispiele habe ich ja eigentlich vorhin genannt. **I: Identifizierung von Händen?** B: Genau, wenn ich mir das jetzt auf mittelalterliche Handschriften vorstelle. Oder auch bei Drucken haben wir auch im Hause ein Projekt, wo wir mit einem Software-Anbieter, der OCR-Software entwickelt, zusammenarbeiten, unter der Maßgabe, dass diese Software auf einen Font trainiert werden kann und es damit umgekehrt auch möglich sein könnte, unterschiedliche Fonts dann herauszufinden. Also nicht dann nur zu sagen, ich möchte jetzt den Text erkannt haben mit einer möglichst hohen Genauigkeit, dass also weniger Fehler drin sind und diesen Text als elektronischen Text nutzen zu können, sondern auch durchaus Aufschlüsse über den Font selber zu gewinnen, um dann vielleicht mal zu einer Stilgeschichte der Drucklettern im frühen Buchdruck zu kommen. Oder wenn ich das auf Handschrift übertrage, um zu erkennen, ob eine Hand sich geändert hat, ob also sozusagen kleine Änderungen stattfinden und die dann auch inhaltlich zu interpretieren und zu sagen, da hat offensichtlich ein Handwechsel stattgefunden oder aber es ist wahrscheinlich dieselbe Hand, die dann später weitergeschrieben hat. Es lassen sich dann schon methodische Fragen finden, die mit technischen Hilfsmitteln gelöst werden können, wenn man einfach z. B. dieses Tool... Ich versuche den Text möglichst gut zu erkennen, und dann muss es eben auf einen Font trainiert werden, auf eine Schriftart oder irgendwas, dass man das auch umdrehen kann und sagen kann, wenn du die besonders gut erkannt hast, dann wirst du ja auch die Unterscheidung zu einer anderen besonders gut sehen können. Denn dann reichen ja vielleicht schon kleine Nuancen aus, um das zu erkennen. Ja das wäre so etwas, z. B.

I25-13#I: Bei der Arbeit mit digitalen Editionen habe ich als Thema auch immer wieder genannt bekommen, dass man auch zusammen an einem Projekt arbeiten kann, also an einer Datei, an einem Text. Inwiefern ist das bei ihrer Arbeit der Fall und welche Problematik ergibt sich da? B: Das ist bei den Projekten, die ich jetzt hier in der H Bibliothek betreue, eher nicht der Fall. Könnte aber schon auch sein, dass man bspw. Ressourcen hat, wir versuchen das gerade auch aufzubauen neben vielen anderen Dingen, dass wir alle personengebundene Information, die in Projekten entsteht, sei es in Editonsvorhaben oder in der Aufarbeitung irgendwelcher Quellen, wie z. B. ein Projekt, das wir haben, die Hofapotheke, untersucht wird auf das Personal hin und auf die Arzneimittel, die dort verabreicht worden sind und so etwas. Und wenn man jetzt natürlich aber eine so zentrale Ressource hat, wo personengebundene Information drin steckt, und wir dann Projekte haben, wo der eine als eine

Person bearbeitet wird, als der Vorbesitzer eines Buches, das wir haben, der andere aber als eine Person, die eine bestimmte Medizin einnimmt, der dritte als jemand, der vielleicht in einem Bildungs-, oder politischen Zusammenhang eine Rolle gespielt hat, dann könnte es auch natürlich sein, dass die auf dasselbe Dokument zugreifen wollen. Dann denke ich aber schlicht, dass es in der Programmierung mittlerweile Tools gibt wie Versionierung und Zugangskontrolle, also Sperren von Dokumenten und so etwas, die also in Office-Paketen oder in Betriebssystemumgebungen eigentlich mittlerweile ja Gang und Gäbe sind und dann eben die entsprechenden Ressourcen aber in Repositorien und so etwas angewendet werden müssen und diese entsprechenden Inhalte dann enthalten. Das kann sein, dass man damit zu tun haben wird, aber eigentlich stehen die Lösungsmöglichkeiten auch schon zur Verfügung. Man muss sie natürlich nur immer wieder implementieren und sagen, wenn ich ein Repository mit Texten aufbauen möchte, dann sollte das halt eine solche Funktionalität enthalten und dass der gleichzeitige Zugriff auf Dokumente eben irgendwie erkannt oder gesperrt wird, oder eben die Versionskonflikte hinterher irgendwie aufgelöst werden. Also Versionierungssysteme sind ja insbesondere in der Programmierung, wo eben auch gerne verteilt gearbeitet wird, ja Gang und Gäbe. Da müssen sich vielleicht die Geisteswissenschaftler vielleicht ein bisschen dran gewöhnen, dass so etwas sinnvoll ist und nötig.

I25-14#I: Gut, abschließend würde ich noch gerne wissen, was ist der größte mögliche Fortschritt, das größte Potential, das Sie in TextGrid sehen? Und was das größte Manko oder das größte Hindernis? B: Naja, wenn es TextGrid wirklich schaffen sollte, solche Ressourcen, Tools z. B. zur Verfügung zu stellen, die ein adäquates Bearbeiten von Dokumenten ermöglichen, insbesondere dann, wenn man eben nicht auf andere Ressourcen zurückgreifen kann, dass man also keine Serverkapazität hat, die man benutzen kann und ähnliches, und wenn es TextGrid auch gelingen sollte z. B. auch Informationsarten zu erkennen oder so zu klassifizieren, dass damit ein Arbeiten möglich sein wird; und wenn es TextGrid auch gelingt, als Angebot eben ein Bewusstsein bei den Forschern dafür zu erzeugen, dass man auf bestimmte Dinge Rücksicht nehmen sollte, wie z. B. das eindeutige Identifizieren von Entitäten oder von Strukturen oder von den Objekten, die man da bearbeitet – dann könnte ich mir schon vorstellen, dass das eine Oberfläche werden kann oder sozusagen eine Rückfallposition, die der Wissenschaft dient. Aber bis dahin sind halt eine ganze Menge technischer Probleme zu lösen. Man hat das ja bei diesem Workshop gesehen, dass schon relativ lange ja TextGrid als Projekt existiert, also drei, vier, fünf Jahre vielleicht schon – ich weiß es jetzt nicht ganz genau, aber –

...und das, was bis jetzt da ist, eben die Probleme widerspiegelt, dass eben noch viele offene Enden rumliegen und in der Komplexität und ja auch immer wieder in der Eigenwilligkeit der Forscher – in Ausführungszeichen – sich darlegt. Und da liegt natürlich auch das Problem drin. Aber das ist nicht TextGrid-spezifisch, das ist menschlich irgendwie... Chance und Gefahr zu gleich. **I: Gut vielen Dank.**