**Interview mit Dr. Dorothea Iglezakis vom 03.07.2023**

**Fabian Boehlke (FB): Vielleicht könnten Sie ja zu Anfang noch mal allgemein noch etwas zu FoKUS erzählen, was Sie machen und welche Angebote Sie haben.**

**Dorothea Iglezakis (DI):** Also FoKUS ist ja das Forschungsdaten-Kompetenzzentrum der Uni Stuttgart, was größer klingt als es ist. Wir sind ein Querschnittsteam aus Bibliothek und Rechenzentrum und auch noch aus Data Steward Research Engineers und Mitarbeitern in den NFDI-Konsortien an der Uni Stuttgart, die sich darum kümmern, das Forschungsdatenmanagement an der Uni Stuttgart voranzubringen und die Mitglieder der Uni Stuttgart dabei zu unterstützen, ihre Daten zu managen, zuzuteilen und zu veröffentlichen. Der Schwerpunkt liegt darauf, eine Infrastruktur aufzubauen, in der die Arbeit mit der Datierung von Forschungsdaten einfacher und automatischer wird. Damit soll ein Workflow unterstützt werden, in dem Daten dann einfach und leicht auffindbar werden und nutzbar für andere sind.

**FB: Auf der Website habe ich gesehen, dass Sie das in die Bereiche Beratung, Schulung, Dienste und Services sowie Projekte unterteilen. Teilweise sind das ja Ihre eigenen Dienste, aber Sie verlinken auch auf externe Angebote?**

**DI:** Ja genau. Also unser Hauptdienst, den wir anbieten, ist das DaRUS, unser Datenrepositorium, welches zur Veröffentlichung von Daten da ist, aber auch für interne Projekte und solche über institutionelle Grenzen hinweg. Es dient dazu, die Daten FAIR zu machen. Dafür ist DaRUS unser Hauptservice. Und dann haben wir auch so verschiedene Services, die teilweise im Aufbau sind, teilweise auch so kleine Services rund um DaRUS. Diese Services sollen die einfache Integration und den Forschungsprozess unterstützen oder kleine Aufgaben erfüllen. Ein größerer Service, der im Aufbau ist und der auf jeden Fall bald kommen wird, ist Jupyter Hub. Das ist so ein Service, mit dem man Code ausführen und damit wissenschaftliche Workflows nachvollziehen kann, quasi live in so einer Plattform. Was auch kommen wird, ist eine RDMO-Instanz zur Erstellung von Datenmanagement Plänen. Davon hatten wir bereits eine eigene, die aber wenig genutzt und von uns auch wenig gepflegt wurde. Wir haben das quasi als Service angeboten bekommen von der ULB Darmstadt und können das weiterführen. Ich glaube, das sind so die die Hauptservices, die wir anbieten. Alles andere ist ein bisschen drumherum.

**FB: Ich hatte ja schon in der E-Mail geschrieben, dass meine Fragestellung sich vor allem an die geisteswissenschaftlichen Forschungsdaten richtet. Die Universität Stuttgart besteht ja zu sehr großen Teilen aus den Bereichen Naturwissenschaften, Technik, Ingenieurwesen etc. Spiegelt sich das auch in Ihren Angeboten wider?**

**DI:** Na, sagen wir, was wir so in unseren Projekten tun und auch in unseren Kooperationen mit der Welt, das hat schon den Fokus auf die technischen Fächer und die Ingenieurwissenschaften. Einerseits deswegen, weil wir das hauptsächlich an der Uni Stuttgart haben. Andererseits auch, weil in den Bereichen bisher auch sehr wenig an Forschungsdatenmanagement-Praktiken und -kultur vorhanden ist. Da ist also viel zu tun. Was im Bereich Geisteswissenschaften tatsächlich bei uns gemacht wurde, das war im Re-

Play-DH-Projekt ein Tool, das helfen soll, Metadatierung im Alltag für die Digital Humanities zu ermöglichen. Dies geschieht auf Basis von einem Git-Client. Das, was Git im Hintergrund tut, wird so abgeschirmt von dem Nutzer, der die Prozesse im Alltag dokumentiert. Das wurde im Rahmen dieses Projektes entwickelt. Es ist allerdings jetzt nicht so, dass uns jetzt alle Leute aus den Geisteswissenschaften dafür Tür eingerannt haben.

Es gibt noch ein anderes Projekt, was auch in die Richtung Digital Humanities geht. Das wurde nicht direkt von FoKUS durchgeführt, sondern von der UB, da war mein Kollege der Hauptakteur. Das ist eine Erweiterung von DaRUS, um urheberrechtlich geschütztes Material rechtssicher zur Verfügung zu stellen. Da geht es um Textcorpora, bei denen sich ja die Forschung oft gar nicht ran traut, überhaupt irgendwelches urheberrechtlich geschütztes Material zu verwenden. Und in dem Projekt ging es darum, dass zwanzig Prozent legal für die Forschung verwendet werden dürfen. Und eben diese zwanzig Prozent sollen zur Verfügung gestellt werden. Dass man sagen kann, die zwanzig Prozent kannst du bekommen, aber eben nicht mehr. Es ging also darum, ein Tool zu entwickeln, welches dann in DaRUS integriert und womit sichergestellt werden kann, dass niemand mehr als die zwanzig Prozent bekommt. Mit den zwanzig Prozent kann dann Forschung betrieben werden. Also das war ein Projekt, dass letztes Jahr abgeschlossen wurde und noch so ein bisschen im Prototypen-Status ist. Das ist auch so ein Service, der nicht unbedingt von so vielen genutzt wird, sondern eher den Status Proof of Concept hat. Ich würde sagen, das sind so ein bisschen die Services, die sich an die Geisteswissenschaften richten. Bei uns ist so, dass wir jetzt für DaRUS immer so Bereiche vergeben, normalerweise auf Institutslevel, dass Institute, aber auch größere Projekte wie Exzellenzcluster oder SFBs ihre Bereiche, die sie bekommen, auch selber verwalten können. Und da ist es jetzt so, dass wir bspw. aus der Linguistik auch Datensätze haben.

Oder aus der Geschichte der Technik, da haben wir eine große Sammlung von Kreiseldaten. Ich meine, dass ein Großteil der geisteswissenschaftlichen Daten aus dieser Sammlung kommen. Da können wir aber jetzt nicht sagen, dass wir ganz speziell ein besonderes Angebot für die Geisteswissenschaften auf DaRUS haben. Abgesehen davon, was ich vorher meinte, da was so in Projekten entstanden ist. Da kam jetzt aber auch noch kein größerer Bedarf an uns heran, als dass man gesagt hätte, man bräuchte jetzt ein extra Metadaten-Schema, um die Sachen zu erkennen und zu beschreiben, oder man bräuchte extra Schnittstellen zu anderen Bereichen. Das IMS [Institut für Maschinelle Sprachverarbeitung] ist auch im Clarin-Verbund mit drinnen und betreibt dort ein eigenes Repositorium. Also die sind auch ein bisschen woanders mit dabei.

**FB: Und kann man da ungefähr sagen, wie das bei DaRUS so ist mit dem Anteil der geisteswissenschaftlichen Forschungsdaten?**

**DI:** Also wie gesagt, je nachdem, wie man es definiert.

**FB: Ich weiß, die Definition ist ein bisschen schwammig.**

**DI:** Eine schwierige Definition, was genau geisteswissenschaftliche Daten sind. Also wie gesagt, aus der Linguistik haben wir Daten, die definitiv linguistisch sind. Und diese Gyrologs-Sammlung, ob das jetzt geisteswissenschaftliche Daten oder ob das technische Daten sind, ist ein bisschen schwierig zu sagen, weil sie so bisschen beides ist. Also je nachdem, aus

welchem Blickwinkel man drauf guckt. Sonst muss ich schon sagen, der Schwerpunkt unserer Daten, die wir auf DaRUS haben sind einerseits Simulationsergebnisse, Experimentaldaten, in ganz unterschiedlichen Kontexten, Materialwissenschaften und chemische Daten. Wir haben einen Schwerpunkt, in dem es sehr viel um porösen Medien geht. Da gibt es dann teilweise CT-Scans von Materialien oder eben auch Simulationen. Also vieles, was wir haben, sind Experimentaldaten, Kommunikationsdaten, auch nicht wenig Forschungssoftware als Ergebnis, entweder selbst oder als Analyseskripte. Das ist jetzt schon der Schwerpunkt, was bei uns reinkommt. Teilweise auch Daten für, für Data Science, so dass man sagt, man hat Daten, die man gut für KI nutzen kann, um Modelle zu lernen.

Eine ganze Menge sind tabellarische Daten aus verschiedenen Bereichen. Obwohl ich da sagen würde, da ist jetzt nicht der Schwerpunkt der Geisteswissenschaften. Also alles Mögliche, eine große Bandbreite. Wir haben auch etwas aus der Architektur, so Terminologien und Ontologien, die in dem Bereich entwickelt werden. Da ist das auch wieder Ansichtssache, sind das jetzt Architekturdaten oder könnte man das auch als geisteswissenschaftliche Daten interpretieren? Also es wird immer alles, was wir haben mit einbezogen.

**FB: Sie haben ja schon kurz über Re-Play-DH erzählt. Wie wird das Tool so angenommen?**

**DI:** Also soweit ich sehen kann, überschaubar. Es kann sein, dass ich da nicht alles mitbekomme, was herankommt, aber zumindest über FoKUS kam da eigentlich gar nichts an Anfragen zur Nutzung der Software. Da wären eventuell meine Kollegen die besseren Ansprechpartner, die man nochmal nachfragen könnte, inwieweit Leute auf sie zukamen. Beides ist veröffentlicht auf DaRUS also sowohl der Client wie auch das Metadatenschema, was dem zugrunde liegt. Ich meine, die Downloadzahlen stehen dort, da könnten Sie dann selber gucken.

**FB: Dann gucke da noch mal nach. Es gibt ja bei Ihnen noch die beiden anderen Bereiche, Beratung und Schulung. Kann ich da als Geisteswissenschaftler genauso sagen, ich würde gerne an der Schulung teilnehmen und hätte dann Vorteile davon?**

**DI:** Selbstverständlich. Wir haben regelmäßige Forschungsdatenmanagement-Schulungen für die Zielgruppe der Doktoranden und auch für Juniorprofessuren. Faktisch sind die nicht ganz so regelmäßig wie siegedacht sind. Und wir machen hauptsächlich Schulungen auf Anfrage. Das heißt, wenn jemand zu uns kommt und sagt, wir haben hier ein Graduiertenkolleg oder wir haben hier in unserem Projekt oder in unserem Institut Seminare, wir wollen jetzt DaRUS nutzen und es wäre schön, wenn jemand vorbeikommen würde und uns erklären würde, wie das geht. Oder wir wollen mehr wissen zu den Anforderungen der Forschenden. Oder wir wollen allgemein etwas wissen zu Forschungsdatenmanagement. Oder die DFG hat uns gesagt, so geht das nicht mit unseren Anträgen, wir müssen mehr zu Forschungsdatenmanagement schreiben. Wer auch immer das ist, da kommen wir und schulen und beraten. Und das gilt für Geisteswissenschaftler natürlich genauso wie für Ingenieure. Ein großer Teil dieser Schulungen ist auch nicht spezifisch für technische Fächer. Natürlich haben wir Services, die speziell für Technik noch mal was anderes bieten. Aber das ist eigentlich ein kleiner Bruchteil unserer Schulungs- und Beratungstätigkeit. Also der größte Teil davon ist eigentlich unabhängig vom Fach und steht selbstverständlich allen Mitgliedern

der Uni Stuttgart zur Verfügung. Und ab und zu kommt auch mal jemand von außen. Das ist dann jemand, der nicht von der Uni Stuttgart kommt und der irgendwie an unsere Kontaktadresse kommt, und dem helfen wir dann auch soweit wir können. Da kam dann auch mal jemand, der Geisteswissenschaftler ist. Und dann helfen wir so gut wir können. Aber das ist natürlich nicht unsere eigentliche Aufgabe, wir sollen die Forscherinnen und Forscher unterstützen.

**FB: Und wie ist da so der Zulauf zu den Schulungen?**

**DI:** Immer gut bei denen, die auf Anfrage stattfinden. Davon machen wir so fünf bis zehn im Jahr. Und das sind dann, ich würde sagen, zwischen fünf bis zu 30 bis 40 Personen, die dort sind. Das sind teilweise Summer Schools für irgendwelche Projekte oder Instituts-Seminare, die wöchentlich stattfinden, wo mal auch mal jemand was über Forschungsdatenmanagement erzählen soll. Bei den Kursen, die wir durchgeführt und einfach so angekündigt haben, da waren dann auch so zwischen zehn und dreißig Teilnehmern. Das machen wir einigermaßen regelmäßig. Es gibt auch noch ein Online-Webinar, bei dem es um Software geht, was ein Schwerpunkt von uns ist in den Projekten. Das war immer etwas überschaubarer, etwa zwischen fünf und zehn Personen, die immer dabei sind.

**FB: Und die Beratungen sind wiederum individuell, dass man sich da anmelden kann?**

**DI:** Genau, die Beratungen sind einfach individuell. Das läuft in der Regel so, dass jemand über unsere Support-Adresse anfragt. Ganz oft ist es so: Ich muss in meinem Forschungsprojekt einen Datenmanagementplan erstellen oder ich muss etwas zu Forschungsdatenmanagement schreiben in einem Projekt und hab keine Ahnung was ich machen soll. Das ist der größte Teil der Beratung, dass man sagt okay, so geht das und das können Sie machen. Aber das ist immer individuell und immer auf die spezifische Situation der jeweiligen Forschenden, des jeweiligen Projekts, der Gruppe angelegt. Aber was wir nicht haben, wie beispielsweise in Dresden, sind feste Termine, für die man sich dann anmelden kann, die man dann buchen kann. Das haben wir nicht. Wir machen das auf Anfrage. Jeder, der kommt, wird beraten. Aber wir haben keinen festen Termin dafür.

**FB: Bei der Beratung ist es das Gleiche, dass das fachunabhängig ist?**

**DI:** Genau, dadurch gehen wir da individuell auf das ein, was die Leute für Daten haben. Und das variiert eben stark. Wir sind natürlich nicht in jedem Fachgebiet die Experten, was fachspezifische Standards oder die Verwaltung von Daten angeht. Aber ein Großteil des Forschungsdatenmanagements ist tatsächlich auch einigermaßen unabhängig von allem

**FB: Und da würden Sie nicht sagen, wenn der Linguist ankommt, dann ist das komplexer, als wenn der Ingenieur kommt und ein paar Messdaten hat?**

**DI:** Beides ist auf seine Art komplex. Also in beiden Fällen geht es darum zu gucken, okay, um was für Daten geht es bei dem Linguisten, was ja in der Regel weniger komplex ist. Da geht es meistens nicht um so große Datenmengen, wie bei den Technikern mit ihren Terabytes an Daten. Da stellt sich auch die Frage, wo wird es gespeichert, wie wird das übertragen? Wie kann man das überhaupt händeln? Die Fragen habe ich normalerweise bei

Geisteswissenschaftlern eher nicht. Da geht es ja eigentlich auch immer um die Frage, in welchem Format liegen die Daten vor? Ist das ein Format, das in der Community anerkannt ist? Ist es ein offenes Format? Ist es irgendwas, was man, was andere auch nutzen können, was auch wahrscheinlich in zehn Jahren noch gut funktioniert? Die Fragestellung an sich ist unabhängig vom Fach. Und dann muss man sich halt genau die Daten angucken und schauen, wie sieht es denn aus? Wäre es möglich, das in etwas anderes konvertieren? Und das ist aber halt spezifisch, was genau für Formate vorliegen. Und da ist natürlich immer die Frage, was sind denn die Informationen, die man zusätzlich noch braucht, die man dann in den Metadaten hinterlegen muss? Und auch da, das ist in der Regel gar nicht so unterschiedlich, je nach Fach. Ich meine, überall geht es darum zu sagen, was hat man denn eigentlich betrachtet? Was war der Prozess, in dem man diese Daten bearbeitet hat und natürlich, welche Methoden da verwendet wurden und was es da für Parameter gibt oder was genau jetzt die Informationen sind, die man braucht. Braucht man eine Geo-Referenzierung oder irgendeine historische Einordnung, was dann vielleicht spezifisch für ein bestimmtes Fach ist? Aber dass man irgendwie so was braucht, das ist halt wieder unabhängig. Und natürlich ist es so, wenn jetzt jemand kommen würde, aus einem Fachgebiet, das ich noch nie hatte, was in den Geisteswissenschaften viele sein könnten, dann müsste ich auch erstmal recherchieren und gucken. Gibt es da etablierte Metadatenstandards, etablierte Daten oder Dateiformate? Allerdings würde ich dann auch immer erst mal die Forschenden fragen, weil sie das oft selbst besser wissen als ich. Aber sowas passiert dann im Dialog, wo man klärt, welche Vorgaben die haben. Was wissen die über ihren Fachbereich, was ist in der Community üblich aus der Sicht des Forschungsdatenmanagements.

**FB: Könnte es denn mit den Dateiformaten Probleme geben beim Repositorium?**

**DI:** Jedes Dateiformat kann prinzipiell irgendwann mal Probleme machen. Wir haben bei uns keine generelle Policy, dass wir sagen, nur diese Liste von Dateiformate ist erlaubt. Grundsätzlich ist bei uns alles möglich. Wir ermuntern aber spätestens bei der Publikation die Forschenden, ihre Daten in offene Formate zu konvertieren, sofern es dafür welche gibt. Jeder Datensatz kommt ja nochmal bei uns vorbei. Und natürlich gibt es aber auch Dateiformate, die wir einfach nicht kennen. Wir fragen dann in der Regel einfach nach. Das ist aber jetzt eher ein Problem bei den Ingenieuren, also dass Messdaten in einem proprietären Format vorliegen, weil das von der Maschine abhängt, mit denen diese erhoben wurden. Und da gibt es dann manchmal Möglichkeiten, die in was anderes zu konvertieren. Und das Problem hatten wir jetzt bei geisteswissenschaftlichen Daten nicht so. Natürlich gibt es so XML-Formate, und wenn das irgendwelche annotierten Daten sind, dann kann man natürlich gucken, dass man da irgendwelche Metadaten automatisch extrahieren kann. Aber auch da können die Probleme auftreten, ob man das in fünf oder zehn Jahren noch lesen kann. Und wenn das jetzt ein textbasiertes Format ist, dann würde ich sagen, sieht schon mal besser aus, dass man zur Not mit einem Texteditor die Information irgendwie extrahieren kann, als wenn es irgendein binäres Format ist. Wenn es sich um ein binäres Format handelt, mit was für einer Software kann man das öffnen? Dann versuchen wir zumindest, dass diese Information dokumentiert wird. Dass man sagt, das sind solche Daten, die kann man mit der Software öffnen. Zumindest im Jahre 2023 hat es in der Version funktioniert. Dann hat man

immer noch die Möglichkeit, das notfalls über Emulation nochmal zu öffnen. Genau, aber viel mehr machen wir im Moment nicht diesem Langzeitarchivierungsstreben.

**FB: Könnten Sie noch kurz etwas zu ReSUS erzählen, das hat ja auch mit Software zu tun.**

**DI:** ReSUS ist ein Projekt, in dem eine Plattform entstehen soll, welche die Forschungssoftware als Forschungsergebnis besser verfügbar machen soll. Sie soll diese Hürde nehmen, den Code zum Laufen zu bringen. Das wird keine neue Plattform sein, sondern in DaRUS integriert, dass man Software dort veröffentlichen und es in einer ganz bestimmten Form veröffentlichen kann. In diesem Fall in Form von so einem Container, aber der auf einem spezifischen Standard beruht, in diesem Fall TOSCA. Wir haben auch ein anderes Projekt, das jetzt gerade abgeschlossen ist, das eher auf der Basis von Docker Containern beruht und dann versucht, die Software in Form von einer Web App zur Verfügung zu stellen. ReSUS geht da ein bisschen anders an die Sache ran. Da ist die Idee, die Software so zur Verfügung zu stellen, dass man sie in einer Cloud-Umgebung automatisch installieren kann. Also TOSCA ist so ein Standardwerk wie auch offener Standard, in dem man alles zusammenpackt, was man braucht. Und auch eine Beschreibung, wie man es installieren, deployen kann. Und dass man dann so etwas wie eine OpenStack-Umgebung automatisch installiert bekommt und die Software läuft und man sie ausprobieren kann, ohne große Hürden zu nehmen. Das ist der eine Teil von ReSUS. Das ist noch ein bisschen work in progress. Also da sind wir noch dran, das umzusetzen. Der andere Teil ist Hilfe, Software und die richtige Lizenz für die Software auszuwählen. Das ist so ein Problem, weil die meisten Forscher sich damit gar nicht auskennen oder auskennen wollen und sich auch nicht damit beschäftigen. Da müssen wir Unterstützung bieten. Und da entwickeln wir auch gerade ein Tool, so einen Licence-Checker, der einem hilft, die richtige Lizenz zu finden für die Software. Ja, das ist so ein bisschen der andere Teil von ReSUS.

**FB: Auch sehr interessant. Was mich noch interessiert, als ich mich durch Ihre Angebote durchgeklickt hatte, bin ich dann auch immer wieder bei den Technischen Informations- und Kommunikationsdiensten gelandet, insbesondere, wenn es um Speicherung und Archivierung geht. Inwieweit sind die jetzt auch mit dem Forschungsdatenmanagement betraut?**

**DI:** Also wir sind ja ein Querschnittsteam aus Bibliothek und eben den Technischen Informations- und Kommunikationsdiensten (TIK). Das ist unser Rechenzentrum. Und manche von uns im FoKUS-Team sind angesiedelt an der an der Bibliothek und andere sind am Rechenzentrum angesiedelt. Wir sind aber jeweils sehr eng vernetzt mit den Institutionen. Über die Bibliotheksseite arbeiten wir sehr eng zusammen mit dem Publikationsteam, das sich auch um die Open Access Publikationen kümmert und die bei uns bei dem Publikationsprozess über die Datensätze drüber gucken. Und auch auf TIK-Seite sind wir da eng verbunden. Einerseits mit den Storage Diensten, in denen es ja tatsächlich um die Speicherung von Daten geht, weil die ja irgendwo hinmüssen, insbesondere diese riesigen Terabyte von Daten. Enge Zusammenarbeit gibt es auch mit den Kollegen, die sich um die Services kümmern und uns virtuelle Maschinen zur Verfügung stellen auf denen wir dann unsere Services aufbauen können. Genau mit denen arbeiten wir sehr eng zusammen. Und wir arbeiten auch eng zusammen mit denen, die sich um Identity Management kümmern,

weil wir das auch dauernd brauchen. Also wir sind ein Querschnittsteam mit Fühlern in viele Richtungen, insbesondere zu UB und TIK, aber auch in die Forschung und Projekte.

**FB: Also man könnte sagen, das Forschungsdatenmanagement liegt auch so ein bisschen übergreifend bei unterschiedlichen Institutionen?**

**DI:** Genau.

**FB: Vielen Dank für das Interview.**