



EIN MÄRCHEN MIT WAHREM KERN
DIE FAIR-PRINZIPIEN FÜR FORSCHUNGSDATEN

A FAIRy tale

Publiziert durch: FAIR project

Text verfasst von:

Karsten Kryger Hansen ORCID 0000-0002-2407-8764

Mareike Buss ORCID 0000-0002-1459-1345

Lea Sztuk Haahr ORCID 0000-0003-2211-6138

Redaktionelle Gestaltung und Illustrationen: Paulina Halina Sieminska

Korrektorat: Lotte Stehouwer Øgaard

Deutsche Bearbeitung: Ilona Lang ORCID 0000-0002-7202-5982 (Text) und

Lena Dreher ORCID 0000-0002-4531-9432 (Satz) unter Verwendung des Original-Layouts

Lizenz: CC-BY-SA 4.0 Attribute: 'DK Fair på tværs' -<https://creativecommons.org/licenses/by-sa/4.0>

Created by AAU, CBS, DTU KU, The Danish National Archives, The Royal Library.

Supported by DeiC and Deff.



Error

Page not found!



Vielen Dank für den Input und die Unterstützung an alle am Projekt Beteiligten.
Vielen Dank an DM Forum und seine Sponsoren, die dies möglich gemacht haben.

Inhalt:

Einleitung	6
Findable #1: (Meta-)Daten werden mit global eindeutigen und persistenten Identifikatoren versehen	8
Findable #2: Daten werden mit umfangreichen Metadaten beschrieben	10
Findable #3: Metadaten enthalten klar und deutlich den Identifier der Daten, die sie beschreiben	12
Findable #4: (Meta-)Daten sind in einer durchsuchbaren Ressource registriert oder indexiert	14
Accessible #1: (Meta-)Daten sind anhand ihrer Kennung unter Verwendung eines standardisierten Kommunikationsprotokolls abrufbar	16
Accessible #1.1: Das Protokoll ist offen, frei und universell einsetzbar	18
Accessible #1.2: Das Protokoll ermöglicht bei Bedarf eine Authentifizierung und Autorisierung	20
Accessible #2: Metadaten sollten auch dann zugänglich sein, wenn die Daten nicht mehr verfügbar sind	22
Interoperable #1: (Meta-)Daten verwenden eine formale, zugängliche, gemeinsam genutzte und allgemein anwendbare Sprache zur Wissensrepräsentation	24
Interoperable #2: (Meta-)Daten verwenden Vokabulare, die den FAIR-Prinzipien folgen	26
Interoperable #3: (Meta-)Daten enthalten qualifizierte Verweise auf andere (Meta-)Daten	28
Reusable #1: (Meta-)Daten sind reichhaltig mit einer Vielzahl von genauen und relevanten Attributen beschrieben	30
Reusable #1.1: (Meta-)Daten werden mit einer klaren und zugänglichen Nutzungslizenz freigegeben	32
Reusable #1.2: (Meta-)Daten sind mit einer detaillierten Provenienz verbunden	34
Reusable #1.3: (Meta-)Daten erfüllen domänenrelevante Gemeinschaftsstandards	36
Ende	38

Einleitung



Es war einmal im schönen Königreich Datamania ein Prinz namens Fairhair. Obwohl er ebenso gutaussehend wie sanftmütig war, ließ ihn sein Vater die Liebe seines Lebens nicht selbst auswählen. Nein, es war ihm bestimmt, eine Frau aus dem benachbarten Königreich zu heiraten. Er kannte nicht einmal ein Gemälde von ihr und wusste nur, dass sie als My Fair Lady bezeichnet wurde. Bevor die Mutter von My Fair Lady bereit war, in die Heirat einzuwilligen, stellte sie Prinz Fairhair eine Aufgabe. Nur wenn er diese Aufgabe erfüllte, würde er sich der Prinzessin als ebenbürtig erweisen und sie heiraten dürfen. Die Aufgabe lautete, das Rezept für die Verwandlung von Wasser zu Gold zu finden. Eine abenteuerliche Suche, die das Sammeln einer Unmenge von Datentruhen und Hinweisen erfordern würde.

Glücklicherweise war Prinz Fairhair nicht allein. Einer der Schlossflügel beherbergte eine Reihe von Zauberinnen und Zauberern, die ihm bei der Entschlüsselung und Untersuchung der Datentruhen helfen konnten. Es war diesen zaubernden Datenassistentinnen und Datenassistenten jedoch nicht möglich, selbst auf die Jagd nach den Daten zu gehen. Um ihnen zu helfen, wurde ein große Anzahl von untergebenen Elfen darauf trainiert, nach Datentruhen zu suchen. Den Elfen wurden Bücher, Zeitschriften, Comics und sogar Gedichte zu lesen gegeben, damit sie lernten, wo sie suchen mussten. Die Suche begann und die Elfen machten sich auf die Suche nach Datentruhen im ganzen Königreich Datamania und in weit, weit entfernten Reichen.



Die FAIR-Prinzipien, von denen hier die Rede ist, wurden erstmals im Jahr 2016 veröffentlicht. Sie enthalten Richtlinien für eine gute Datenmanagementpraxis, die darauf abzielen, Daten auffindbar, zugänglich, interoperabel und nachnutzbar (Findable, Accessible, Interoperable und Reusable = FAIR) zu machen. Insgesamt umfassen die FAIR-Prinzipien fünfzehn Unterpunkte.

Obwohl sie aus den Lebenswissenschaften stammen, können die FAIR-Prinzipien in anderen Forschungsdisziplinen, einschließlich der Geistes- und Sozialwissenschaften, angewandt werden – und werden es auch bereits. Seit ihrer Veröffentlichung haben sowohl die Europäische Union als auch einzelne Förderer und Universitäten ihre Unterstützung und Zustimmung zu den FAIR-Prinzipien erklärt. Dies reicht von der Schaffung von Datenmanagement-Tools und -Infrastrukturen bis hin zur Festlegung von Richtlinien für den grundsätzlichen Umgang mit Daten. Einige Implementierungen halten sich eng an die ursprünglichen Definitionen, während andere vom Geist der FAIR-Prinzipien inspiriert sind.

Eine erste Grundvoraussetzung für das Verständnis von FAIR ist, dass sowohl Menschen als auch Maschinen mit Daten interagieren. Nur so kann ein Ökosystem entstehen, das schnell auf Veränderungen reagiert und sich automatisch an neue Erkenntnisse oder Veränderungen anpasst. Das ist auch der Grund für die Konzentration auf Datenstandards, Identifikationsmechanismen und beispielsweise die Verfügbarkeit von Daten. Zweitens gelten die FAIR-Prinzipien sowohl für Daten als auch für ihre Metadaten – d.h. für die beschreibenden Datensätze zu den eigentlichen Datensätzen. Deshalb wird in den Grundprinzipien der Begriff "(Meta-)Daten" verwendet. Drittens beziehen sich die Prinzipien nicht nur auf offene Daten. Mit Daten, die nicht für die öffentliche Verfügbarkeit bestimmt sind, können Sie ebenfalls im Sinne von FAIR arbeiten.

Die FAIR-Prinzipien stellen keinen Qualitätsstandard dar, den Sie zur Bewertung von Werkzeugen, Daten, Richtlinien usw. verwenden können. Dadurch würden die Prinzipien bald veraltet und nicht mehr in allen Forschungsdisziplinen anwendbar sein. Die Umsetzung von FAIR kann eine allmähliche und systematische Anpassung neuer Arbeitsroutinen oder ein großer Sprung sein, bei dem Sie eine Art von Struktur durch eine andere ersetzen. Die Umsetzung der Prinzipien sollte an jeden Forschungsbereich angepasst werden, d.h. jede Forschungsgemeinschaft wird die Prinzipien in ihrem jeweiligen Kontext zur Anwendung bringen.

Findable #1:

(Meta-)Daten werden mit global eindeutigen und persistenten Identifikatoren versehen



Die Elfen kehrten einer nach dem anderen in die Burg zurück, und einige von ihnen waren wirklich frustriert. Sie waren Pfaden zu Datentruhen gefolgt, die akribisch beschrieben worden waren, aber irgendwie waren die Datentruhen entfernt worden und hatten nur Löcher im Boden hinterlassen. Fimble war einer dieser Elfen, der ziemlich verwirrt über einige seltsame Codes war, die er gefunden hatte. Er konnte sie nicht entziffern und wusste daher auch nicht, wohin er gehen sollte.

„Schau“, sagte Fimble zur Datenassistentin, „ich habe diesen seltsamen Code 10.1234/abbb, und ich weiß nicht, was er bedeutet!“

„Oh, die sind wirklich sehr nützlich“, sagte die Datenassistentin. „Wir können die Codes in diesen riesigen

Büchern nachschlagen. Lass mal sehen. 10 ist das große Land des Datentals, und wir sollten in der Hausnummer 1234 nachsehen.“ Sie wies im Buch auf eine Karte. „Dorthin solltest du fliegen.“

„Sind Sie sicher, dass sie noch da ist?“, sagte Fimble, der keinen weiteren Schritt auf einer Jagd nach Datentruhen verschwenden wollte, die er nicht finden konnte.

„Auf jeden Fall. Diese Bücher sind magisch. Wenn jemand die Datentruhe an einen neuen Ort bringt, wird das Buch es merken.“

„Großartig“, sagte Fimble und startete mit einem Sprint. Bald darauf kehrte er glücklich mit einer Datentruhe zurück.

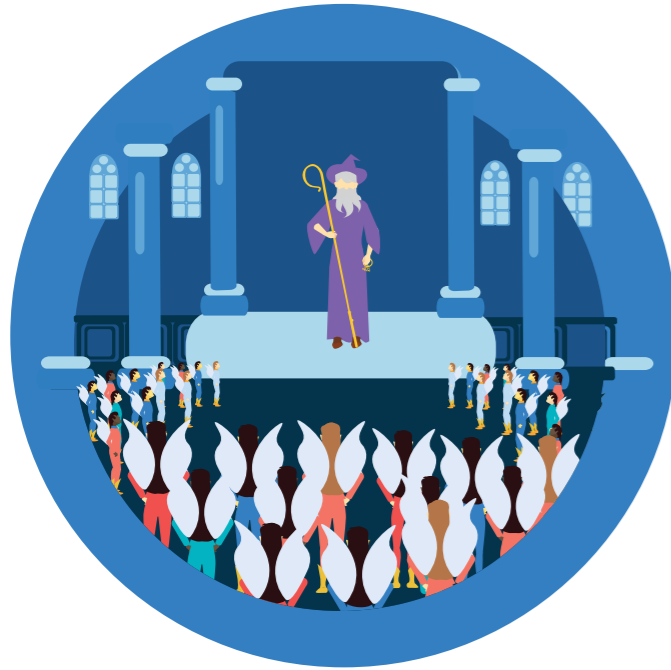
Aber nicht alles war Glückseligkeit. Ein anderer Elf kehrte nie wieder zurück. Sein einziger Hinweis war jemand namens Zhang Wei, mit dem er sprechen sollte. Soweit wir wissen, geht der Elf immer noch von Tür zu Tür und spricht mit Menschen, die diesen Namen tragen.

Ein großes Problem beim Zitieren eigener und fremder Daten mit Verweis auf einen Ort im Internet ist, dass sich der Ort stetig ändern kann. Dies kann durch die Verwendung persistenter Identifikatoren gelöst werden. Sie funktionieren wie ein großer Index oder ein Register, bei dem jedem Datensatz ein eindeutiger Schlüssel (den Identifikator) zugewiesen wird. Wenn jemand versucht, dem Identifikator zu folgen (oft als „Auflösen eines persistenten Identifikators“ bezeichnet), zeigt der Resolver auf die richtige Webadresse (URL). Wenn sich die URL ändert – d.h. wenn Daten verschoben werden – ist derjenige, der den Schlüssel erstellt hat, dafür verantwortlich, dem Resolver den neuen Standort zur Verfügung zu stellen. Dies verhindert, dass Sie in einer Internet-Sackgasse „Seite nicht gefunden“ landen. Deshalb nennt man die Identifikatoren „persistent“ (PI).

Ein DOI (Document Object Identifier) ist ein Beispiel für einen solchen Identifikator und sieht wie folgt aus: 10.1234/abba (Präfix/Suffix). Dieser kann aufgelöst werden und zeigt auf die URL hin. Die meisten Datenrepositorien geben DOIs oder andere PIs aus und pflegen sie. Zusätzlich enthalten PIs normalerweise einige grundlegende beschreibende Metadaten wie Titel und Autor.

Andere PIs werden verwendet, um Mehrdeutigkeiten zu verhindern, z.B. indem einer Person eine Nummer statt eines Namens gegeben wird. Damit wird das Problem der Unterscheidung zwischen Thomas Müller und Thomas Müller gelöst – Ja, es gibt mehr als eine Person mit diesem Namen! ORCID ist ein Beispiel für einen Dienst, bei dem einer Person ein eindeutiger Code zugewiesen wird, z.B. "0000-0002-1825-0097". der auf eine – und nur eine – Person verweist.



Findable #2:**Daten werden mit umfangreichen Metadaten beschrieben**

Als die ersten Elfen in die Burg zurückkehrten, sahen ihre Funde alles andere als vielversprechend aus. Sogleich versammelten sie sich um den leitenden Datenzauberer in der Halle der Burg. Er trat auf ein Podium, um ein paar Worte zu sagen: „Ihr befindet euch auf der wichtigsten Mission von allen. Es gibt jedoch ein paar Regeln und Ratschläge, die ihr beachten müsst. Wenn ihr euch auf die Suche nach Datentrühen begeben, müsst ihr euch an die Wege halten, die ihr bereits aus euren Studien kennt – wir Zauberinnen und Zauberer können euch in den Wäldern nicht finden, wenn ihr euren eigenen Weg geht und euch verirrt. Wenn ihr eine Truhe findet, stellt sicher, dass ihr alles über die Truhe auf dem Etikett lest, um zu sehen, ob ihr Inhalt für uns relevant ist. Lest sie sorgfältig durch, da manche Etiketten vielleicht nur nebulöse Aussagen über den Inhalt der Truhe

ermöglichen. Und schließlich dürft ihr die Truhe niemals selbst öffnen! Überlasst das uns, den Datenassistenten.“

Obwohl der leitende Datenzauberer wusste, dass die Elfen ihr Bestes tun würden, drückte er alle drei Daumen. Die Elfen waren wahrhaftig klug; dennoch bezweifelte er, dass sie alle relevanten Trühen dort draußen finden würden.

Wenn Menschen und Maschinen nach Daten suchen, sind Metadaten oft der erste Anhalts- und Kontaktpunkt, da Metadaten in der Regel in Suchmaschinen usw. indexiert sind. Daher sind es oft die Metadaten, die darüber entscheiden, ob der zugehörige Datensatz für ein bestimmtes Nutzungsszenario als relevant wahrgenommen wird oder nicht.

In den Metadaten sollten die gleichen Suchwörter vorhanden sein, die ein Mensch zum Suchen eines Datensatzes verwenden würde. Dabei handelt es sich um Metadaten über den Kontext und/oder die Voraussetzungen für den Datensatz, Qualitätsfragen usw. sowie eine Reihe disziplinspezifischer Daten, wie z.B. den Stichprobenumfang oder verwendete Materialien. Dazu gehören auch Details über den Datensatz, die für Sie vielleicht nicht wichtig sind, die aber verwendet werden könnten, um Ihre Daten außerhalb Ihrer eigenen Disziplin auffindbar zu machen. Versuchen Sie also, beim Hinzufügen von Metadaten zu Ihren Daten, über den Tellerrand hinauszudenken.



Findable #3:

Metadaten enthalten klar und deutlich den Identifier der Daten, die sie beschreiben



Einige Elfen kamen ohne Datentruhen zurück. Stattdessen brachten sie hübsche kleine Blätter mit eingravierten Texten, die den Inhalt der Truhen beschrieben. Eine der Elfen, die zurückkam, war Faruty, und sie war sehr stolz auf das goldene Blatt, das sie gefunden hatte.

„Schau“, sagte sie zu der Datenzauberin namens Wiuxa, „Das habe ich gefunden, und es beschreibt die faszinierendste Truhe, von der ich je gehört habe. Sagen Sie mir, wohin ich gehen soll, und ich werde sie holen.“

„Lass mich mal sehen“, sagte Wiuxa und drehte das Blatt immer wieder um. „Seltsam Es gibt keine Einzelheiten darüber, wo die Truhe zu finden ist. Bist du sicher, dass es neben dem Blatt keine Truhe gab?“

„Ja, sicher, sicher“, sagte Faruty. „Ich fand diese hier direkt am Rande des Zauberwaldes, und ich verspreche, dass keine einzige Truhe in Sicht war. Ich habe in alle Richtungen und Dimensionen geschaut.“

„Dann habe ich keine Ahnung, wo ich suchen soll“, sagte Wiuxa. „Mein bester Vorschlag wäre, den schrecklichen Herrn von Uguly zu fragen, ob diese Truhe irgendwo in seinem giftigen Sumpf versteckt sein könnte.“

„Ha ha ha ha, Sie sind so lustig“, sagte Faruty und rollte mit einem Auge. Aber das Lächeln verschwand bald, als sie Wiuxa ansah, auf deren Gesicht nicht das geringste Anzeichen von Ironie zu erkennen war.

„Ab mit dir“, sagte Wiuxa. „Diese Truhe brauchen wir wirklich“, und schickte Faruty auf eine Mission, die gefährlicher war als der Kampf gegen einen Syzx-Drachen.

Metadaten und Daten sind zwei getrennte Dinge und sollten als solche behandelt werden. Ein Beispiel sind Metadaten, die für große Indizes gesammelt werden, die nicht den Inhalt der Dateien enthalten oder indizieren. Genauso wie Metadaten zu Artikeln, die für die Indexierung in kommerziellen und nichtkommerziellen Suchmaschinen abgerufen werden, ohne dass diese sich den Inhalt des Artikels selbst wiedergeben. Wenn Metadaten keinen Verweis auf die von ihnen beschriebenen Dateien enthalten, ist es – insbesondere für eine Maschine – zweifelhaft, ob die von den Metadaten beschriebenen Daten jemals gefunden werden. Daher sollte in den Metadaten immer auf den zugehörigen Datensatz mittels eines persistenten Identifiers hingewiesen werden, damit die Daten eindeutig auffindbar sind.



Findable #4:

(Meta-)Daten sind in einer durchsuchbaren Ressource registriert oder indexiert



Auf ihrer Suche nach Datentruhen hatten die Elfen alle Arten von Literatur gelesen, um Hinweise zu erhalten, wo sie diese finden konnten. Doch der Elf Fusky hatte es wirklich satt, nach Datentruhen zu suchen, also setzte er sich für eine kurze Pause unter den Apfelbaum im Hof des Schlosses. Er bemerkte, wie die Elfe Fyrty mit einer großen hölzernen Baumtruhe über den Hof ging.

„Was ist das, Fyrty?“ fragte er.

„Das ist Gold“, antwortete Fyrty mit einem breiten Lächeln auf dem Gesicht. „Oder zumindest sieht es so aus, als hätte jemand vor uns versucht, Gold herzustellen.“

„Wo hast du es her, und wie hast du es gefunden?“ fragte Fusky, ungeduldig zu erfahren wo er als nächstes suchen sollte.

„Das ... Nun, ich ... habe es im Wald gefunden. Es lag einfach so auf dem Boden. Ich bin sicher, wir haben es nicht in den Büchern“, sagte Fyrty etwas verwirrt. „Das ist doch in Ordnung, oder?“

„Sicher. Es ist großartig“, sagte Fusky. „Aber das bedeutet, dass wir in jeder Ecke jedes Königreichs nach Kisten suchen müssen. Das wird Jahrzehnte dauern!“

„Ich weiß“, sagte Fyrty und versuchte, das Lächeln auf ihrem Gesicht zu bewahren. „Aber ich habe gerade Gerüchte gehört, dass Wildy, der Datenassistent, mit den Leuten im Bibliotheksturm spricht. Vielleicht haben sie eine Karte der Zauberkiste gefunden.“

„Aufregend“, sagte Fusky und vergaß dabei fast, dass er es stattdessen, nach Datentruhen zu suchen. „Wenn alle Truhen kartiert sind?“

„Vielleicht“, antwortete Fyrty, „Wer weiss ...?“

Das gesamte Internet nach (Forschungs-)Datensätzen zu durchforsten, ist weder möglich noch machbar. Und es lässt zu viel Raum für zufällige Funde, was manchmal zwar schön sein kann, aber bei der strukturierten Suche nach Datensätzen nicht wünschenswert ist. Zudem erhöht die alleinige Bereitstellung von Forschungsdaten auf Projektwebseiten etc. das Risiko, dass diese nur per Zufall gefunden werden.

Repositorien sind verwaltete Orte zur Aufbewahrung geordneter Dokumente und stellen, meistens als Webseiten, eine gängige Methode zum Aufbau strukturierter Indizes von Metadaten und Datensätzen dar, die wiederum in das Repositorium hochgeladen werden. Die Indizes folgen oft der Beschreibung der Daten einem bestimmten gemeinsamen (Metadaten)Standard. Dies ermöglicht sowohl dem Repositorium als auch anderen Suchmaschinen diese Register zu sammeln und zu indexieren, wobei sie oft zu größeren Indizes aggregiert werden, die schließlich auch übergreifend durchsucht werden können.



Repositorien gibt es in vielen Formen und Ausprägungen. Einige sind generische Repositorien, die fast jeden Datensatz aufnehmen können, während andere auf bestimmte Disziplinen oder Forschungsdatentypen ausgerichtet sind. Repositorien sind in der Regel Eigentum von Institutionen, Forschungsgemeinschaften oder privaten Unternehmen und werden von diesen betrieben. Bei der Frage, wo genau Ihre Daten hinterlegt werden sollen, geht es darum, das beste Repositorium für Ihren spezifischen Datensatz zu bestimmen und dadurch seine Auffindbarkeit und sein Potenzial zu maximieren. Dies sollte von Fall zu Fall individuell beurteilt werden.

Accessible #1:

(Meta-)Daten sind anhand ihrer Kennung unter Verwendung eines standardisierten Kommunikationsprotokolls abrufbar



Alok war der traurigste Elf im ganzen Königreich Datamania. Das dachte er jedenfalls. Ein Zauberer fand ihn beim Dryaden-Brunnen auf dem Weg zurück zum Schloss.

„Was ist los?“, fragte der Zauberer ein wenig spöttisch.

„Ich wurde in das Königreich Dovia geschickt, um alle Beschreibungen ihrer Datentruhen zu lesen und zu sehen, welche wir kaufen konnten. Leider war ich nicht in der Lage, eine einzige zu lesen. Ich bin schrecklich darin!“ jammerte Alok.

„Sei nicht so hart mit dir selbst“, antwortete der Zauberer mit süßer und sanfter Stimme und bedauerte seinen vorherigen Tonfall. „Sag mir genau, was passiert ist.“

„Jedes Mal, wenn ich nach einer Truhe griff, um zu lesen, was sich darin befand, änderten alle Buchstaben auf dem Etikett ihre Position. Ich versuchte, ihnen zu folgen, aber sie änderten sich einfach ständig. Ich sollte in der Lage sein, sie zu lesen, nicht wahr?“ murzte Alok.

„Oh, du bist auf einige alte dovianische Datenkisten gestoßen, die im wirren Dovianisch geschrieben sind. Für uns ist das völliger Unsinn – und für die auch, wie ich hinzufügen möchte. Niemand kann Dovianisch lesen, und es ist eine Schande, dass wir nicht einmal die Beschreibungen der Truhen lesen können. Das letzte Lebewesen, das sie lesen konnte, war eine sprechende Kröte, aber sie ist vor einigen Jahren gestorben.“

„Das ist so kompliziert“, seufzte Alok. „Ich sollte gehen und ein paar Tage schlafen. Meine Augen sind wirklich müde.“

„Gute Idee“, sagte der Zauberer. „Erhol dich und danach mach weiter so.“

Sobald jemand entweder Ihre Metadaten oder die Daten selbst gefunden hat, sollte diese Person – oder ihre Maschine – in der Lage sein, über standardisierte Mechanismen auf die (Meta-)Daten zuzugreifen. Dieses Prinzip besagt, dass der Zugriff über ein standardisiertes Protokoll erfolgen sollte. Meistens handelt es sich dabei um Protokolle, die wir aus dem Internet kennen – z.B. http(s) oder FTP. Dies ist in der Regel dann der Fall, wenn die Daten in einem vertrauenswürdigen Repositorium abgelegt wurden.

Es kann jedoch Fälle geben, in denen Sie zusätzliche Mechanismen wie Vertragsinformationen oder Ähnliches von einer Person benötigen, bevor diese auf Ihre Daten zugreifen darf. Dies entspricht vollkommen den FAIR-Grundsätzen, wenn Sie dies in den Metadaten eindeutig darlegen. Dies kann durch die Angabe von Kontaktmöglichkeiten geschehen, die allgemein akzeptiert und einfach zu benutzen sind. Beispiele hierfür sind Telefonnummern oder E-Mail-Adressen.



Accessible #1.1:

Das Protokoll ist offen, frei und universell einsetzbar



Die Elfe Agon kehrte in die Burg zurück und sah aus, als hätte sie gerade einen Drachen mit vier Köpfen gesehen.

„Ich brauche eine Katze aus dem Königreich Stiodor, einen Strick, einen Hering und ein Zwiebelhorn“, sagte sie.

„Entschuldigung“, antwortete die Datenassistentin. „Wozu brauchst du diese Dinge? Willst du einen Codun-Drachen bekämpfen?“

„Nein“, antwortete die Elfe. „Um Himmels willen, nein! Ich folgte einem ungewohnten Weg, um an eine bestimmte Truhe zu gelangen. Plötzlich stand ein Gryrvos-Kobold direkt vor der Truhe und befahl mir, eine stiodorische Katze mit einem Seil zu schwingen, während ich einen Hering essen und auf ein Zwiebelhorn treten sollte. Sonst dürfte ich die Brücke nicht passieren, um an die Truhe zu gelangen.“

„Hmmm“, sagte die Zauberin. „Ich glaube wirklich nicht, dass wir dafür die Zeit haben. Ich kann dir ein Seil, einen Hering und ein Zwiebelhorn besorgen, aber die Katze Muss sie aus Stiodor sein? Abgesehen davon, dass sie sehr selten sind, sind sie extrem teuer.“

„Ich weiß“, sagte Agon. „Aber ohne sie lässt mich der Kobold die Brücke nicht passieren. Keine Stiodor-Katze, kein passieren.“

„Ich glaube nicht, dass wir von Stiodor eine Katze bekommen können. Und schon gar nicht, wenn wir ihnen sagen, was wir mit ihr vorhaben“, antwortete die Zauberin etwas verärgert. „Du musst eine andere Truhe suchen, hoffentlich werden wir auf diese Truhe verzichten können.“

Ein Protokoll ist der Fachbegriff für den Standard, wie Daten über ein Netzwerk, z.B. das Internet, transportiert werden. TCP/IP ist ein Beispiel für ein offenes, freies und universell einsetzbares Protokoll, das für den größten Teil des Internets verwendet wird. Das bedeutet, dass jeder es benutzen kann, ohne Nutzungsgebühren zahlen zu müssen. Wenn Sie sich für ein Protokoll mit Nutzungsbeschränkungen entscheiden, könnten Sie anderen Personen den Zugriff auf Ihre (Meta-) Daten verwehren und damit die Nutzung der von Ihnen veröffentlichten Daten erschweren oder sogar unmöglich machen. Wenn Sie ein vertrauenswürdiges Repository für die Datenveröffentlichung verwenden, wird das Repository sicherstellen, dass dieses Prinzip eingehalten wird.



Accessible #1.2:

Das Protokoll ermöglicht bei Bedarf eine Authentifizierung und Autorisierung



Der Elf Albon war auf einer der gefährlichsten Missionen unterwegs – er sollte eine Datentruhe holen, die vom Drachen Guardo bewacht wurde. Niemand kannte das Temperament des Drachens, aber Albon hatte keine Angst. Er brachte Gold und Opfergaben zu Guardo, um Zugang zu der Datentruhe zu erhalten. Außerdem trug Albon ein vergoldetes und von Prinz Fairhair unterschriebenes Zertifikat bei sich, um zu beweisen, dass der Prinz ihn tatsächlich geschickt hatte. Zudem hatte Myrtimar, der Troll, dem die Daten gehörten, das Zertifikat mit einem Zauberspruch belegt.

„Willkommen“, schnaufte der Drache, als Albon auf ihn zukam. „Ich bewache diese Datentruhe, und nur diejenigen, die sich als würdig erweisen können, dürfen die Truhe von hier heraustragen.“

„Nun ...“, sagte Albon, „ich trage dieses unterschriebene und verzauberte Zertifikat bei mir, und ich kenne auch das geheime Passwort.“

„Gib mir das Zertifikat, und sag den Satz“, sagte Guardo und schaute dabei in fünf verschiedene Richtungen gleichzeitig.

„Hier“, sagte Albon, als er dem Drachen die Urkunde überreichte. „Und das geheime Passwort lautet: Wasser zu Gold, vom Glück überrollt.“

Der Drache drehte sich um und blickte auf die kleine Höhle hinter ihm.

„Gib mir die Geschenke. Die Truhe ist da drin, du kannst sie mitnehmen. Aber denke daran, dass nur Prinz Fairhair die Truhe öffnen kann. Wenn du es versuchst, wirst du verbrennen“, sagte Guardo mit bedrohlicher Stimme und die Tür zur Höhle öffnete sich und enthüllte eine glänzende Truhe mit wertvollen Daten.

Ein weit verbreiteter Mythos über FAIRe Daten ist, dass alle Daten offen verfügbar sein müssen, was bedeutet, dass diese von jedermann frei heruntergeladen werden können. Das ist nicht der Fall.

Dieses Prinzip besagt lediglich, dass, wenn man Daten hinter einer Art digitalen Mauer, einer Paywall oder einem einfachen Genehmigungssystem für den Zugang platziert, das System eine Art Authentifizierung und Autorisierung ermöglichen muss. Dies gilt sowohl für Menschen als auch für Maschinen.

Bei der Authentifizierung geht es darum, einem System mitzuteilen, wer Sie – oder Ihr System – sind. Autorisierung hingegen ist der Prozess, bei dem das System evaluiert, ob Ihnen der Zugriff auf eine bestimmte Ressource erlaubt ist.

Es gibt viele gute Gründe, Daten hinter Authentifizierungs- und Autorisierungsmechanismen zu stellen. Aber bitte denken Sie daran, dass wenn Sie manuelle Prozesse zur Evaluierung des Datenzugangs haben, das Repositorium oder Speichersystem in der Lage sein muss, Sie zwecks Genehmigung zu kontaktieren. Bei anderen Systemen können sich Benutzer oder Systeme selbst registrieren und erhalten daraufhin Zugang zu den Daten, wobei ein Logbuch darüber geführt wird, wer Zugang zu den Daten erhalten hat.



Accessible #2:

**Metadaten sollten auch dann zugänglich sein,
wenn die Daten nicht mehr verfügbar sind**



Die Elfe Ahlerta kam mit Tränen in den Augen zum Zauberer Uku zurück.

„Was ist los?“, fragte der Zauberer und studierte die weinende Elfe.

„Sie war da. Ich weiß, dass sie da war, aber ich konnte sie nicht finden“, schluchzte Ahlerta. „Was?“, fragte der Zauberer. „Was war wo?“, fragte der Zauberer, „Das verstehe ich nicht.“

„Ich hatte einen perfekten Anhaltspunkt für eine Datentruhe. Ich war so scharf darauf, sie zu finden. Aber alles, was ich fand, waren einige leere Truhen mit winzigen Goldbrocken und eine seltsam klebrige Substanz. Aber da war absolut nichts, was nach Daten aussah“, wunderte sich Ahlerta und schüttelte den Kopf. „So nah und doch so fern. Es wurde in einem Buch erwähnt ...“

„Schrecklich frustrierend!“, rief der Zauberer. „Hast du da draußen nach Hinweisen gesucht?“

„Habe ich nach Hinweisen gesucht?“, antwortete Ahlert empört. „Ich habe mich stundenlang umgesehen und sogar versucht, mit Leuten von der Ogohu-Insel zu sprechen, wo die Truhe sein sollte. Aber nichts. Nichts!“

„Gut, gut... Halt die spitzen Ohren steif und such weiter, es gilt noch viele andere Truhen zu finden“, beschwichtigte der Zauberer.

Es gibt viele gute Gründe, warum Daten aus Repositorien und ähnlichen Orten verschwinden. Sie können beispielsweise aus Kostengründen zurückgezogen werden oder wenn niemand mehr Zugang zu den Daten hat. Wenn jedoch auch die Metadaten über den Datensatz verschwinden, werden Mensch und Maschine in einem „unerfüllten“ Zustand zurückbleiben, wenn sie versuchen, die Daten abzurufen, z.B. beim Abrufen eines persistenten Identifikators oder beim Folgen eines Links zum Eintrag eines Repositoriums. Stellen Sie daher sicher, dass Metadaten mit der Aussage „ja, die Daten waren hier, aber sie sind nicht mehr verfügbar“ erhalten bleiben. Das Hinterlassen von Metadaten kann auch Informationen über den Kontext, die Autoren und die Institution, in der die Daten erstellt wurden, für diejenigen bieten, die nach weiteren Details suchen.



Interoperable #1:

(Meta-)Daten verwenden eine formale, zugängliche, gemeinsam genutzte und allgemein anwendbare Sprache zur Wissensrepräsentation



Es dauerte nicht lange, bis alle Räume im Untergeschoss des Schlosses mit riesigen Stapeln von Datentruhen aus Königreichen in der Nähe von Datamania und Reichen in der Ferne überschwemmt waren. Die Zauberinnen und Zauberer hatten Mühe, mit all den Truhen Schritt zu halten, die die Elfen zurück in die Burg brachten. Die schiere Masse der Datentruhen war jedoch nicht die einzige Quelle der Frustration. Es wurde schnell klar, dass nicht alle Truhen etwas damit zu tun hatten, Wasser in Gold zu verwandeln, obwohl sowohl „Wasser“ als auch „Gold“ auf dem Etikett standen. Die Zauberin Igly schüttelte verzweifelt den Kopf. Sie war besorgt, und sie hatten noch nicht einmal begonnen, nach Truhen mit den Aufschriften „Gold“ und „H2O“ zu suchen.

„Was ist los?“, fragte Ila, eine kleine Elfin mit einer quietschenden Stimme, die der Zauberin sanft den Kopf streichelte.

„Sieh dich um“, seufzte Igly. „Sieh dir all diese Kisten an. Es ist gut, dass du sie hierherbringst, aber viele von ihnen sind überhaupt nicht relevant. Und ich fürchte, dass einige der relevanten Kisten nicht gefunden werden.“

„Ich verstehe“, antwortete Ila, obwohl sie die Sorge der Zauberin nicht wirklich verstand. „Was ist das für ein Geräusch? Ist es... Musik?“

„Oh, dieses Geräusch“, antwortete die Zauberin fast verraten schauend. „Das ist Musik. Jemand kam mit einer Datentruhe mit der Aufschrift Gold zurück, aber es stellte sich heraus, dass es sich um Musik der Echse Helene Fairshare handelte.“

„Meeeeega“, schrie Ila so laut, dass man es bis in die Gänge des Schlosses hören konnte. „Lasst uns drei Tage und Nächte lang Spaß haben und tanzen. Ist das nicht großartig?“

„Nicht wirklich, ...“, murmelte Igly, aber Ila war schon lange weg.

Menschen und insbesondere Maschinen können sich mit der Interpretation von Daten schwertun. Die meisten Wörter sind mehrdeutig, und die Vielzahl der gesprochenen und geschriebenen Sprachen erhöht die Komplexität zusätzlich. Die Probleme reichen von der Unmöglichkeit, den Wert einer Zelle aufgrund fehlender Informationen über die verwendeten Metriken zu interpretieren, bis hin zu komplexeren Situationen, in denen Sie nach vielen verschiedenen Begriffen suchen müssen, die dasselbe Objekt beschreiben, oder über Wörter mit unterschiedlicher semantischer Bedeutung in verschiedenen Disziplinen stolpern.

Die FAIR-Prinzipien gehen auf dieses Problem ein, indem sie die Verwendung gemeinsamer Datenstandards für die Darstellung von Daten sowie die Verwendung von Vokabularen und Ontologien zur Darstellung von Werten und maschinenlesbaren Markup-Daten empfehlen. Vokabulare und Ontologien werden oft innerhalb der Forschungsgemeinschaften definiert und sind ein eindeutiger Weg, um Ihren Daten eine semantische Bedeutung zu verleihen. Ein einfaches Beispiel ist, sich bei der Klassifizierung von Blumen auf eine Blumenontologie zu verlassen.

Die so strukturierte Arbeit mit Daten kann Ihre Daten nützlicher und leichter auffindbar machen. Sie sollten sich jedoch bewusst sein, dass diese Art von Arbeit oft auch Auswirkungen auf Ihre Methodik und die von Ihnen verwendete Software hat.



Interoperable #2:

(Meta-)Daten verwenden Vokabulare, die den FAIR-Prinzipien folgen



Die Elfe Imka hatte eine Datentruhe gefunden, die die Datenassistentin Dorky gerade öffnen wollte. Imka war so neugierig, dass sie bleiben und zusehen durfte, wie Dorky die Truhe öffnete.

„Schau“, rief Imka, als sie die Truhe öffneten. „Da sind Carmix-Blasen drin. Ist das nicht erstaunlich? Diese können sicherlich helfen, Wasser in Gold zu verwandeln, nicht wahr?“

„Ja, absolut“, sagte Dorky. „Aber sieh genauer hin, die haben Markierungen. Das sind so genannte Polymixic-Markierungen. Nur sehr wenige Assistenten – wenn überhaupt – wissen, wie sie zu interpretieren sind. Sie zeigen an, wie jede der Carmix-Blasen mit den anderen verbunden ist und wie man sie in die richtige Reihenfolge bringt.“

„Polymix-Markierungen ... Ich habe noch nie von ihnen gehört“, sagte Imka und kratzte sich am Kopf.

„Ich kenne sie nur aus einer Erzählung, und ich habe tatsächlich nicht an ihre Existenz geglaubt“, sagte Dorky düster. „Einmal arbeitete eine Zauberin namens Yrky in der Burg, und sie kannte anscheinend einen Zauberer, der diese polymixischen Markierungen sortieren konnte“, sagte Dorky und kratzte ihren Kopf, „Aber sie starb bereits im Alter von 523 Jahren.“

„Aber sie muss etwas hinterlassen haben, das uns helfen kann“, rief Imka mit leidenschaftlicher Stimme. Doch der Gesichtsausdruck von Dorky ließ sie verstummen.

„Es tut mir leid, kleine Freundin“, murmelte Dorky entschuldigend. „Ohne das richtige Verständnis für die Logik der polymixischen Markierungen können wir nicht wirklich etwas tun. Aber sie schmecken köstlich“, sagte sie, als sie hineinbiss. Die Blase sprudelte ein wenig, als sie zu kauen begann.

Ein Vokabular ist nur dann gut, wenn es zugänglich ist und die richtige Interpretation der Daten ermöglicht. Dieses Prinzip macht deutlich, wie wichtig die Verwendung von Vokabularen ist, die in der Fach-Community verbreitet und gut dokumentiert sind und auf die mit Hilfe persistenter Identifikatoren verwiesen werden kann. Gewöhnlich finden Sie diese Art von Vokabular, Taxonomie usw. innerhalb Ihrer Forschungsdisziplin oder vielleicht in anderen Disziplinen, wo diese entwickelt wurden. Die Evaluierung eines Vokabulars beinhaltet oft die Suche nach seinem Urheber bzw. seiner Urheberin und die Überprüfung, ob es noch gepflegt und aktualisiert wird. Dabei kann es sich um komplexe Vokabulare handeln oder um einfache Auszeichnungszeichen wie ISO-Norm-Strings zur Darstellung von Ländern in einem Datensatz. Deutschland ist z.B. DEU in ISO 3166-1 alpha-3. Im Hinblick auf die Interoperabilität ist dies weitaus besser, als 'Deutschland', 'DE', 'Germany', 'German' usw. in Ihre (Meta-)Daten zu schreiben.



Interoperable #3:

(Meta-)Daten enthalten qualifizierte Verweise auf andere (Meta-)Daten



Der Elf Inandu kehrte ganz aufgeregt in die Burg zurück. Er hatte eine Datentruhe mit einem korrekten Aufkleber mit der richtigen Stelle gefunden, an der er später die Truhe fand. Diese Truhe enthielt einen Zauberspruch, mit dem man Wasser in Gold verwandeln konnte.

Der leitende Zauberer Ikloton öffnete die Datentruhe und begann, den Zauberspruch zu lesen.

„Das ist gut“, freute er sich. „Wirklich, wirklich gut. Inandu, ich glaube, dies ist das Ende unserer Suche.“

„Soll ich den Prinzen herbeirufen?“ fragte der Elf und errötete vor Stolz, dass er diese Truhe ins Schloss gebracht hatte.

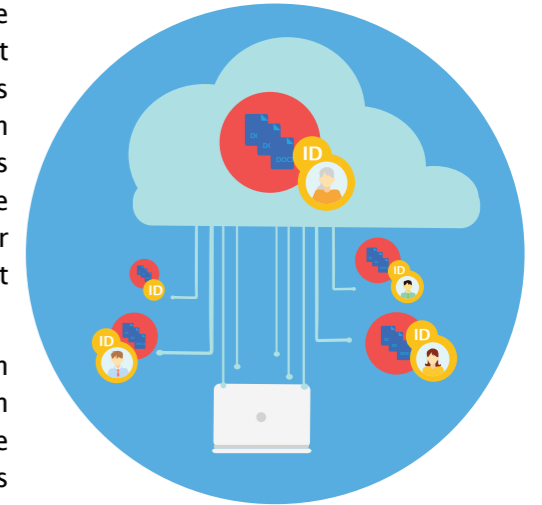
„Ja ... Aber warte!“, grunzte der höchste Zauberer, wurde zuerst weiß und dann rot vor Wut: „In Schritt 4 des Zaubers beziehen sich die Autoren des Zaubers auf ein Verfahren, mit dem sie die Schlüsselzutat dieses Zaubers beschworen haben – violette Echsen mit grünen Punkten. Wie wird dieses Verfahren genannt? Und wie kann es gefunden werden?“ Doch die Schreibweise gab keinen Hinweis darauf, ebenso wenig wie die Datenkiste oder ihr Etikett. Obwohl die Datenkiste also den richtigen Zauberspruch enthielt, war sie für die Datenzauberinnen und Datenzauberer im Schloss nutzlos.

„Es ... tut mir ... so ... leid“, murmelte der Elf mit Tränen in den Augen. Ihm war nach tagelangem Weinen zumute, was er auch tatsächlich konnte.

„Mir auch“, sagte Iklota. „So nah, und doch sind wir im Dunkeln.“

Es ist wichtig, dass Sie in der Lage sind, die Verbindungen zwischen Ihrem Datensatz und den mit ihm verbundenen Datensätzen nachzuvollziehen. Dies kann durch die Verknüpfung mit anderen Datensätzen geschehen, die nicht in Ihrer Arbeit enthalten sind. Es kann auch durch Verbindungen erfolgen, die zeigen, wie Ihr Datensatz von einer früheren Version abgeleitet oder z.B. auf der Grundlage einiger Rohdaten verarbeitet wurde.

Wie auch immer, es ist wichtig, diese Verbindungen durch Verweise zwischen den Datensätzen aufrechtzuerhalten. Wenn sich Ihr Datensatz auf die Daten anderer – oder auf Ihre eigenen – stützt, ist dies auch eine geeignete Methode, um sicherzustellen, dass die Personen, die die Daten, auf denen Ihre Daten basieren, erstellt haben, angemessen gewürdigt werden.



Reusable #1:

(Meta-)Daten sind reichhaltig mit einer Vielzahl von genauen und relevanten Attributen beschrieben



Die Elfen brachten wie verrückt Datentruhen, sodass sowohl die Elfen als auch die Zauberinnen und Zauberer in einem fast wahnsinnigen Zustand waren. Die Zauberinnen und Zauberer hatten Mühe, den Inhalt der Truhen herauszufinden. Einer der Elfen, der mit einer Truhe zurückkam, war Roscoe. Stolz wie die meisten Elfen brachte er einem Zauberer eine schön glänzende Truhe, für die er auch bezahlt hatte.

„Hmm“, murmelte der Datenassistent. „Was ist denn das? Hoffentlich besser als die letzte?“

„Nun...“, sagte Roscoe. „Es war schwer für mich herauszufinden, ob diese Truhe wichtig ist. Ich weiß, dass wir nur diejenigen

bringen sollten, denen es gelungen ist, Wasser in Gold zu verwandeln. Aber hier steht nicht, wofür der Inhalt verwendet wurde.“

„Oje“, sagte der Zauberer etwas verzweifelt. „Ich muss es untersuchen, um herauszufinden, ob es relevant ist. Aber das wird Zeit brauchen, und ich habe keine Zeit.“

„Tut mir leid“, entschuldigte sich Roscoe mit Blick nach unten.

„Wir haben hier drin keinen Platz für weitere irrelevante Truhen“, rief der Zauberer aus. „Du musst dich auf die konzentrieren, von denen wir wissen, dass sie Wasser transformieren. Und zahl nicht für etwas, bei dem wir uns nicht sicher sind.“

„Was ...“, sagte Roscoe zweifelnd. „Truhen ... zurück ... lassen ...?“

„Ja“, sagte der Zauberer mit fester Stimme. „Ich weiß, dass es schwer ist, aber wir müssen sicherstellen, dass wir nur relevante Truhen einbringen.“

„In Ordnung“, sagte Roscoe. „Ich denke, ich werde auf dem Weg nach draußen eine Angelrute mitnehmen und sehen, ob ich einen See mit vertrauenswürdigen Daten finden kann.“

Die Kennzeichnung Ihrer Daten mit relevanten Attributen – meist in Form von Metadaten – hilft nicht nur beim Auffinden Ihrer Daten. Sie hilft auch Menschen und Maschinen, den Kontext Ihrer Daten zu verstehen. Dies kann in Form von Zweck- und Verarbeitungserklärungen, verwendeten Geräten, Softwareversionen usw. geschehen. Stellen Sie sich vor, Sie finden Ihre eigenen Daten. Denken Sie nun an die Kontextinformationen, die Ihnen bei der Feststellung helfen würden, ob die Daten für Ihre spezifischen Bedürfnisse relevant sind – und ob Sie in der Lage wären zu verstehen, wie die Daten entstanden sind. Seien Sie beim Hinzufügen von Attributen zu Ihren Daten großzügig. Was für Sie vielleicht nicht relevant ist, könnte der wichtigste Inhalt zur Filterung und Abfrage von Daten für andere Personen – oder Maschinen – sein.

Apropos Maschinen, Ihre beste Wahl wäre es, z.B. kontrollierte Vokabulare, persistente Identifikatoren oder Ähnliches zu verwenden, um die kontextuelle Beschreibung eindeutig zu machen. Häufig bieten Repositorien, die auf bestimmte Disziplinen, Gemeinschaften oder Datentypen ausgerichtet sind, sowohl für die Zuweisung, Pflege als auch für die Abfrage mit domänenspezifischen Metadaten eine optimale Unterstützung an.



Reusable #1.1:

(Meta-)Daten werden in einer klaren und zugänglichen Nutzungslizenz freigegeben



Die von den Elfen mitgebrachten Datentruhen gab es in allen möglichen Größen, Formen und Verpackungen. Einige waren sogar in Ketten und mit einem Vorhängeschloss verschlossen. Glücklicherweise konnten die Elfen für sie einen Schlüssel kaufen. Andere waren hübsch eingewickelt mit gewellten Bändern und kleinen Schildern, auf denen Dinge wie „Für dich“ standen, und sogar mit kleinen Zeichnungen versehen, um den Elfen dort zu helfen, wo sie die lokale Sprache nicht lesen konnten. Ein Elf namens Ruby kam mit einer zitternden Datenkiste zurück.

„Was stimmt mit dieser nicht?“ fragte Ruby einen Zauberer. „Sie ist eingewickelt wie die anderen, aber es ist kein Etikett daran. Habe ich etwas falsch gemacht?“

„Oh, lass mich sehen“, sagte der Zauberer. „Sie sieht genauso aus wie eine, die wir brauchen, aber wir müssen sie in den Graben werfen. Wir wissen nicht, ob sie verhext ist. Und wir können sie nicht benutzen, ohne zu wissen, ob wir sie benutzen dürfen oder nicht.“

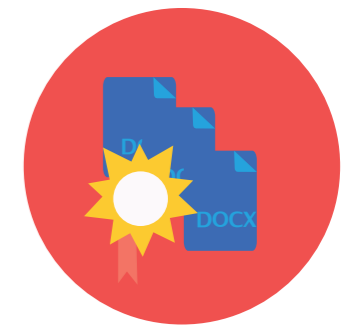
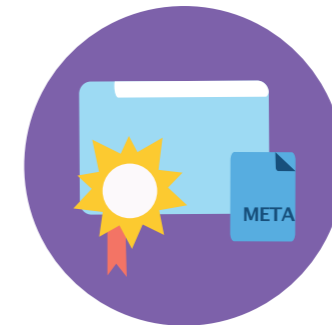
„Was!“, rief Ruby ungläubig aus. „Können Sie nicht etwas tun? Wenn es das ist, was wir brauchen, warum können wir sie dann nicht behalten?“

„Ich weiß, ich weiß“, sagte der Zauberer. „Obwohl sie wie ein Geschenk aussieht, können wir sie nicht behalten, weil sie nicht mit einem Geschenkanhänger versehen ist. Aber das ist nicht deine Schuld, Ruby“, sagte der Zauberer und tätschelte Ruby den Kopf.

Die Lizenzierung von Daten und Metadaten ist ein wichtiger Aspekt der FAIR-Prinzipien; sowohl wenn Sie einige oder alle Rechte behalten, als auch wenn Sie Ihre Daten als vollständig offene Daten freigeben wollen. Im FAIR-Kontext ist eine Lizenz eine standardisierte maschinenlesbare Erklärung, die dem/den Endbenutzer*innen genau angibt, wie und unter welchen Bedingungen er oder sie die Daten verwenden darf.

Sie können viele verschiedene Lizenzen auf Daten anwenden. Eine der gebräuchlichsten ist die Creative-Commons-Lizenz, bei der Sie explizit angeben, wie und ob Sie bei der Wiederverwendung Ihrer Daten zitiert werden sollen, zusammen mit möglichen Optionen für die gemeinsame Nutzung abgeleiteter Werke usw. Typischerweise wählen Sie eine Lizenz, wenn Sie Daten in einem Repository ablegen. Sie müssen dabei nicht auf alle Ihre Daten dieselbe Lizenz anwenden. Das Schlimmste, was Sie tun können, ist überhaupt keine Lizenz zu vergeben. Denn dann sind Ihre Daten, auch wenn Sie sie technisch gesehen öffentlich zugänglich machen, so durch die automatisch eintretende Form des Urheberrechts geschützt, dass sie nicht von anderen Personen weiterverwendet werden können.

Wenn Sie die Daten anderer Personen verwenden, sollten Sie immer die Bedingungen prüfen, wie und wann Sie die Daten verwenden dürfen. Dies kann sich auf die Art und Weise auswirken, wie Sie mit den Daten anderer Personen arbeiten dürfen.



Reusable #1.2:

(Meta-)Daten sind mit einer detaillierten Provenienz verbunden



Der Elf Rherek eilte wie auf einem Zauberwind nach Hause zum Schloss.

„Ich hab's, ich hab's, ich hab's“, rief er fast außer Atem.

„Ich habe was...?“, sagte Jimko, der Datenzauberer. „Atme tief ein, entspann dich und sag es mir.“

„Ich ... habe ... das richtige Rezept, um Wasser in Gold zu verwandeln“, flüsterte er. „Wir brauchen einen Riesen mit drei Köpfen“, flüsterte er. „Und diese Truhe zeigt, wie man einen solchen Riesen herbeizaubert.“

„Wunderbar“, sagte Jimko. „Sehen wir uns sofort die Truhe an.“

Sie untersuchten die Truhe. Der Inhalt sah aus wie Spielkarten. Sie waren jedoch alle durcheinander.

„Mh ...“, sagte Jimko. „Das sieht perfekt aus. Es stimmt mit dem überein, was wir für die richtigen Schritte halten. Es ist jedoch alles durcheinander, sodass wir keine Ahnung haben, in welcher Reihenfolge wir die Schritte unternehmen sollen. Und es wird noch schlimmer ...“

„Was?“, fragte Rherek. „Können wir es nicht einfach in verschiedenen Sequenzen versuchen? Irgendwann muss es uns doch gelingen.“

„Oder wir jagen das Schloss in die Luft. Wenn du mit dem Riesen etwas falsch machst, wird er böser als ein Buffingor-Werwolf an einem schlechten Tag!“

„Aber, aber, aber, aber ... können wir nicht diejenigen fragen, die die Karten erstellt haben?“ fragte Rherek vorsichtig.

„Wir können... Aber die Truhe enthält keine Spur davon, wer die Karten erstellt hat. Das Finden der Person wäre pures Glück. Aber versuch, zurückzugehen und nach Spuren von Beweisen dafür zu suchen, wer die Truhe hergestellt hat“, sagte Jimko verzweifelt.

Rherek lief davon. Aber weder er noch Jimko waren überhaupt davon überzeugt, dass weitere Nachforschungen es möglich machen würden, die Karten zu verstehen.

Der Wert von Daten hängt zum großen Teil davon ab, ob eine Maschine oder ein Mensch die Herkunft der Daten beurteilen kann. Dabei entscheidet sich, ob man die Daten in einem neuen Kontext wiederverwenden kann. Dazu gehört auch die Möglichkeit herauszufinden, wie und von wem die Daten erstellt wurden und z.B. mit welchen Materialien. Wurden die Daten bereits vorverarbeitet, oder handelt es sich um Rohdaten? Wenn sie verarbeitet wurden, wie war dann der Workflow? Und so weiter, und so weiter. Dies ist ziemlich ähnlich zu dem Methodenteil eines Artikels und Sie können in Ihrem Datensatz auf eine Dokumentation im Methodenteil verweisen. Denken Sie jedoch daran, dass dies für eine Maschine möglicherweise nicht lesbar ist.

Denken Sie des Weiteren daran, die Herkunft Ihrer Person anzugeben (z.B. mit einer ORCID-ID), sowie eine Angabe zur gewünschten Zitation einzufügen, wenn Ihr Datensatz anderweitig wiederverwendet wird. Der einfachste Weg ist zu versuchen, sich selbst als Wiederverwender der eigenen Daten zu betrachten. Doch bevor Sie dies tun, müssen Sie Ihren Kopf von allem Wissen bezüglich des Datensatzes befreien. Welche Details würden Sie benötigen, um einen bestimmten Datensatz zu bewerten und ihm zu vertrauen? Wenn dies schwer vorstellbar ist, versuchen Sie, die Datensätze anderer Leute zu finden, und prüfen Sie, ob sie Ihrer Meinung nach genügend Provenienz haben.



Reusable #1.3:

(Meta-)Daten erfüllen domänenrelevante Gemeinschaftsstandards



Rebzuss war die letzte Elfe, der mit einer Datenkiste zurückkehrte. Sie strahlte vor Erwartung und Stolz, denn sie hatte eine Datentruhe gefunden, die einen Zauberspruch zur Umwandlung von Wasser in Gold enthielt.

Der Datenzauberer Fixeor Datahin sah sich das Etikett der Truhe an:

„Gut gemacht, Rebzuss. Das ist genau das, was wir suchen.“ Er öffnete die Truhe und begann die Stirn zu runzeln:

„Hmm ... Ich kann sehen, dass dies der richtige Zauberspruch ist, aber ich kann ihn nicht ganz verstehen. Es sieht so aus, als sei der Zauber von einem Daten-

assistenten geschrieben worden, der im Datenlabor der Oxwart-Universität ausgebildet wurde. Ich erkenne die eigentümliche Verwendung logischer Symbole und die Oxwart-Weise der Datenverarbeitung in den Verfahren 1 und 5. Es wird Jahre dauern, bis wir diesen Zauberspruch in Scruby übersetzen können.“

Rebzuss sah traurig aus – das war nicht die Reaktion, die sie sich erhofft hatte. Plötzlich leuchtete ihr Gesicht auf und sie sagte: „Zauberer Fixeor, warum rufen wir nicht die Zauberin Lux Datastorm herbei? Bevor sie im Schloss in Datamania arbeitete, studierte sie im Datenlabor in Oxwart. Es könnte ihr leichtfallen, den Zauber für uns zu übersetzen.“

Fixeor rief Lux Datastorm sofort herbei und zeigte ihr die Datentruhe und ihren Inhalt. Sie lachte: „Ich kann verstehen, warum Sie Schwierigkeiten haben, den Zauber zu verstehen. Dies wird genau so beschrieben, wie sie in Oxwart zaubern. Geben Sie mir einen Nachmittag und ich werde ihn in Scruby und unsere Datamanish-Verfahren übersetzen.“

Lux nahm die Datenkiste und verschwand in ihrer Kammer.

Die Arbeit mit Datensätzen aus einer Vielzahl von Quellen ist viel einfacher, wenn sich alle auf einen bestimmten Standard zur Organisation und Beschreibung der Daten einigen. Aus diesem Grund haben viele Disziplinen Metadaten-Standards für die Beschreibung von Daten geschaffen und Listen mit empfohlenen Dateiformaten etc. erstellt. Wenn man sich an diese Standards hält, werden die eigenen Daten in ein Ökosystem von Daten überführt, das für andere zur Wiederverwendung geeignet ist. Daher sollten Sie immer versuchen, innerhalb Ihrer Gemeinschaft nach Standards Ausschau zu halten um sich an diese zu halten. Aber natürlich kann nicht alles standardisiert werden und viele Forschungsdisziplinen beschreiten Neuland, wo es derzeit noch keine Standards gibt – dort können Sie sich allgemeineren Standards zuwenden oder neue erfinden.

Beachten Sie, dass ein Standard in diesem Sinne keine Qualitätsmessung ist, welche ein hohes oder niedriges Qualitätsniveau der (Meta-)Daten anzeigt. Die Qualität sollte immer von den Personen beurteilt werden, die die Daten in ihren spezifischen Kontexten wiederverwenden.



Ende



Am folgenden Tag kam Lux Datastorm mit dem Rezept zurück, wie man Wasser in Gold verwandeln kann. Alle Elfen, Datenzauberer und Datenzauberinnen waren anwesend, als sie zu dem Schluss kamen, dass das Rezept tatsächlich funktionieren würde. Alle applaudierten, als das Rezept ordnungsgemäß in einer Glastruhe aufbewahrt und Prinz Fairhair vorgeführt wurde.

Er nahm sofort sein weißes Pferd und ritt in das benachbarte Königreich. Drei Tage später kehrte er mit My Fair Lady zurück. Sie wurden in einer großen Zeremonie in der Bibliothek getraut und anschließend wurde tagelang in den Archiven gefeiert – einem der beliebtesten Orte in Datamania.

Natürlich wurde die Suche nach den FAIR-Prinzipien dokumentiert. Auf diese Weise musste niemand mehr die gleichen Schwierigkeiten durchmachen, und sie konnten alle glücklich bis ans Ende ihrer Tage leben.

About

„A FAIRy tale“ wurde im Rahmen des dänischen Projekts 'FAIR på tværs' / 'FAIR accross' geschrieben.

Konzeptentwicklung und Text:

Karsten Kryger Hansen, Universität Aalborg

Mareike Buss, CBS

Lea Sztuk Haahr, Rigsarkivet

Redaktionelle Gestaltung und Illustrationen: Paulina Halina Sieminska

Korrektur: Lotte Stehouwer Øgaard

Lizenz: CC-BY-SA 4.0-Attribut: DK Messe på tværs

Deutsche Bearbeitung:

Ilona Lang, Universität Konstanz

Lena Dreher, Universität Konstanz

Das FAIR-Projekt wird in Zusammenarbeit mit den folgenden Institutionen durchgeführt: DeIC, DEFF, DTU, CBS, AAU, KU, Rigsarkivet, Det Kgl Bibliotek

Die Originalversion in Englisch zum Ausdrucken:

Karsten Kryger Hansen, Mareike Buss, & Lea Sztuk Haahr. (2018). A FAIRy tale (p. 40). Zenodo. <http://doi.org/10.5281/zenodo.2248200>

