

16. Jan. 21.35.41

RES200_Die_Nationale_Forschungsdaten-Infrastruktur

Willkommen zum Forschungspodcast der Helmholtz-Gemeinschaft.

Ich bin Holger Klein.

[Musik] Resonator Diesmal geht es um die nationale Forschungsdateninfrastruktur.

Und weil ich überhaupt keine Ahnung habe, worüber wir jetzt gleich reden, mache ich es mir leicht und zitiere einfach mal von deren Webseite.

Die Mitgliedsorganisationen von NFDI, also Nationale Forschungsdateninfrastruktur, bilden die Wissenschaftslandschaft in Deutschland in der Breite ab.

Zu den Mitgliedern zählen Wissenschaftsorganisationen, Universitäten und Hochschulen, außeruniversitäre Forschungseinrichtungen, Fachgesellschaften und Vereine.

Und zwei Vertreter bzw.

Vertreterinnen, davon habe ich jetzt an der langen Leitung.

Das sind einmal Astrid Schneidewind vom Forschungszentrum Jülich, also eine Vertreterin einer Wissenschaftsorganisation der Helmholtz-Gemeinschaft, und Philipp Bechtle von der Uni Bonn als Anwender dieser Forschungsdateninfrastruktur.

Hallo ihr zwei.

Hallo.

So, und jetzt stellen wir uns mal nicht nur ganz dumm, sondern ich bin es auch.

Worüber reden wir jetzt?

Was macht ihr?

Ja, wir wollen langfristig Daten fair machen.

Und fair heißt in dem Falle auf Englisch findable, accessible, interoperable and reusable.

Das heißt, die mit ziemlich viel Steuergeld gewonnenen Forschungsdaten sollen halt nicht nur für den einzelnen Wissenschaftler oder die einzelne Gruppe heute und vielleicht morgen verfügbar sein, sondern wir wollen versuchen, diese Daten langfristig auch für andere wertvoll und verwertbar zu machen.

Und dazu muss man die Daten langfristig speichern können.

Sie müssen lesbar sein, suchbar sein.

Und wenn ich Daten gesucht habe, weil ich mich für ein bestimmtes Wissenschaftsfeld oder ein bestimmtes Material interessiere, dann will ich sie nach dem Suchen natürlich finden, auffinden.

Und dann muss ich damit natürlich irgendwas machen können.

Und dazu muss ich verstehen, was in diesen Daten drin ist.

Und das bedeutet, dass ich hinterlege nicht nur das Ergebnis einer Messung, was ich als Experimentalphysiker machen würde, sondern hinterlege, was habe ich wann, zu welcher Zeit, unter welchen Bedingungen gemacht.

Das nennen wir Metadaten.

Damit jemand, der jetzt nicht dabei war bei dem Experiment, versteht, wie diese Daten gewonnen wurden und unter welchen Bedingungen, was ich berücksichtigen muss bei der Auswertung, um die Daten zu nutzen.

Das ist jetzt aus der Physikersicht beschrieben und das Gleiche will man natürlich machen für Daten aus allen möglichen Wissenschaftsfeldern.

Die NFDI möchte weitgehend die gesamte deutsche Forschungslandschaft abdecken und dort eben die gewonnenen Forschungsdaten für die Breite der Gesellschaft, andere Wissenschaftler, öffentlich nutzbar, nachnutzbar machen.

Wenn du sagst, weitgehend abdecken, welche Wissenschaften sind denn da nicht dabei?

Die Juristen.

Entschuldigung.

Die Juristen spielen eine ganz wichtige Rolle.

Es gibt nämlich in dieser NFDI eine Sektion, dazu kommen wir vielleicht noch zu der Struktur von dieser NFDI, die hat nämlich zwei Säulen.

Da gibt es auch eine Sektion, da geht es unter anderem um legale Fragestellungen, weil natürlich Open Science, offene Wissenschaft, nicht unbedingt heißt, dass ich Daten zu jeder Zeit für jeder Mann wirklich zugänglich machen muss.

Warum?

Warum will man das denn?

Man will das doch eigentlich, oder?

Wenn ich mir vorstelle, dass wir Patientendaten über Krankheiten für die Forschung zugänglich machen, dann wäre das extrem wichtig, wie wir alle in Covid gelernt haben, aber natürlich nicht mit den persönlichen Daten der entsprechenden Patienten und auch nicht so, dass man aus diesen zur Verfügung gestellten Daten am Ende auf die Person schließen kann, einfach über irgendwelche Überlegungen oder Crosslinks oder sowas.

Und da ist es ganz wichtig, dass wir sozusagen Datenschutz in persönlichen Dingen machen, aber auch in anderen Dingen zum Beispiel haben wir, ich komme ja aus der Großforschung, ich arbeite tatsächlich am Ende nicht in Jülich, sondern an der Forschungsneutronenquelle in Garching bei München.

Und da bewerben sich Wissenschaftler für Experimente, werden ausgewählt nach wissenschaftlichen Kriterien und wenn sie diese Messzeit tatsächlich bekommen, dann haben sie einen Zeitraum, den wir Embargo nennen, wo sie alleine Zugriff auf diese Daten haben, also als Gruppe, bevor die Daten öffentlich werden und jeder andere gucken kann, was man daraus machen kann.

Ich sprengte jetzt ganz fürchterlich vermutlich, dieses Embargo, ist das dann in die Daten einkodiert?

Also hebt sich das von alleine auf dann im Idealfall?

Na nicht in dem, was die Experimentdaten in unserem Falle sind, aber das gehört zu den Metadaten im Katalog, dass ich dort hinterlege, wann läuft das Embargo ab oder haben die Daten ein Embargo?

Ja, nein.

Und wenn sie ein Embargo haben, dann sind sie zwar im Katalog, aber nicht frei zugänglich.

Aber warum haben die ein Embargo, Philipp?

Also warum sagst du nicht, hier, tolles erfunden?

Ja, also es ist natürlich so, dass man einfach sagt, das, was wir als Forscherinnen und Forscher tun, nämlich uns kreativ auf Sachen auszudenken, was wir machen wollen.

Es ist ja nicht jedes Experiment gleich, das ist jetzt wahrscheinlich in jedem Forschungsfeld so.

Es ist natürlich auch so, dass wenn man eine gute Idee hat, unter welchen exakten Bedingungen man mit den Neutronen welches besondere Material jetzt untersuchen möchte, mit welchem Detektor, man natürlich auch was davon haben will.

Und jetzt ist es ja so, dass es natürlich einem gewissen wissenschaftlichen Wettbewerb attraktiv ist, zu sagen, okay, wie finde ich jetzt für mich mein spannendstes Resultat, das ich machen kann?

Und da will man eben unter Umständen diese Motivation aufrechterhalten, zu sagen, okay, wenn ich die gute Idee hatte, wenn ich den Wettbewerb um die Strahlzeit gewonnen habe, wenn ich den Aufwand betrieben habe, dieses Experiment überhaupt selbst zu betreiben, das macht sich ja nicht von selbst im Normalfall, dann darf ich auch was davon haben.

Hast du schon mal einen Wettbewerb um Strahlzeit gewonnen?

Weil das ist ja Goldstaub, oder?

Also in dem Fall, da ist es vielleicht eher wieder eine Frage für Astrid, in meinem Bereich ist das ein bisschen anders.

Wir bauen die Experimente sozusagen als Kollaboration in meinem Bereich einfach "über Jahre selbst" und betreiben sie selbst und da ist es dann ein bisschen anders.

Da stellt sich auch die Frage mit dem Embargo ein bisschen anders.

Aber das, was Astrid beschrieben hat, ist eben in ganz vielen Wissenschaftsbereichen, also Astronomie ist ein anderes Beispiel, Strahlzeitperteleskopen, ist es eben genau so, dass es eben mehr standardisierte Aufbauten gibt und man dann mit einer bestimmten Frage, mit einem bestimmten Material, mit einer ganz bestimmten Konfiguration an diesen Aufbau geht und dann eben die Fragen des Embargos sich eben genau so stellt, wie Astrid das beschrieben hat.

Aber das ist eben was, was total unterschiedlich ist, wie das jetzt genau sich darstellt für die unterschiedlichsten Wissenschaftsbereiche, nur dass es eben solche rechtlichen Fragen oder auch prozeduralen Fragen gibt, wie Wissenschaft organisiert ist intern und wie die internen Motivationsmechanismen und Rewardmechanismen der Wissenschaft aufrechterhalten werden können.

Das gehört dann eben auch dazu, wie mit den Daten umgegangen wird.

Das heißt, solange deine Publikation nicht durchs Peer-Reviewing gegangen ist, möchtest du auch gerne, dass es deine Publikation bleibt?

Ganz genau.

Und ich glaube, wenn man es ein bisschen weiterfasst noch, dass ganz viel von dem, was wir hier machen, glaube ich, hat damit zu tun, Vertrauen herzustellen.

Und zum einen braucht man natürlich als wissenschaftliche Gruppe oder einzelne Wissenschaftlerinnen oder Wissenschaftler Vertrauen, dass wenn die Daten, die man hat, öffentlich sind, dass damit kein Mumpitz gemacht wird.

Und zu dem Mumpitz kann natürlich auch gehören, dass es jemand einfach sloppy auswertet und dass diejenigen, die sich am meisten Gedanken darüber gemacht haben, auch das erste Wort darüber haben, ist vielleicht auch schon ein gewisser Schutzmechanismus davor, dass dann Sachen, die eben sloppy jetzt im weitesten Sinne...

Rosinenpickerei und sowas.

Genau.

Also man kann ja mit Daten, also das ist vielleicht etwas anderes, worüber wir nachher noch viel reden können, Daten sind ein winziger Teil von Wissenschaft.

Und Astrid hat es eben schon angesprochen, als es um die Metadaten ging, wie Daten zustande kommen und was mit Daten gemacht wird, was eben im weitesten Sinne eben Durchführung der Experimente oder Planung, Durchführung, Datennahme, vielleicht Datenreduktion irreversibel, Datenreduktion bei der Datennahme und Datenauswertung ist.

Das sind ja alles wichtige Teile der Wissenschaft, die im Grunde genauso einen Öffnungsprozess durchlaufen können wie die Daten selbst.

Und insofern gehört das eben alles zusammen.

Und diese Fragen, wann wer was mit Daten machen kann, und vertraue ich, dass kein Mumpitz mit den Daten gemacht wird, kann ich nachvollziehen, was jeder und jede mit den Daten gemacht hat, sind eben genauso wichtige Fragen in dem Zusammenhang.

Gilt das auch dafür, dass der, wie nennt man das denn, der Erschaffer dieser Daten, also der Forscher oder die Forscherin, die diese Daten ursprünglich mal produziert hat, auch keinen Mumpitz gemacht hat, weil jemand Kompetentes jederzeit dann auch nachgucken kann, in diesen Metadaten, hast du denn überhaupt sauber gearbeitet?

Im Idealfall ja, auf jeden Fall.

Also ich glaube, nichts davon ist absolut foolproof, aber umso mehr wir uns den Anspruch setzen, an uns selbst eben Sachen klar zu dokumentieren, in Metadaten und Daten klar zu hinterlegen, umso schwieriger wird es und vielleicht auch umso motivierter ist man, eben alles so sauber wie möglich durchzuführen,

würde ich sagen.

Diese Geschichte mit der Erzeugung der Daten ist total spannend, weil bei dem Versuch, die Wissenschaftler zu bewegen, ihre Datenproduktion nachvollziehbar zu machen, und das trifft vielleicht weniger, aber auch für die Datenauswertung zu, gibt es schon eine relativ große Hemmschwelle, weil sozusagen Wissenschaft ist ja irgendwas Neues und zu sagen, das sind standardisierte Abläufe, die man irgendwie dokumentieren kann, wird nicht so ohne weiteres unbedingt sofort akzeptiert.

Aber das ist genau der spannende Punkt, weil wenn ich natürlich darüber nachdenke, wie ich meine Arbeiten durchführe und wie ich sie erklärbar mache, dann ist das ja auch ein Erkenntnisprozess bei den Experimenten selber.

Und wir reden ja auch über ganz viele Daten, die jetzt vielleicht nicht unbedingt aus Experimenten kommen.

Also wenn wir die Forschungslandschaft abdecken, dann sind das nicht wirklich immer Experimente, die jetzt wirklich mit Setups und Aufbauten zu tun haben, wenn wir den Blickwinkel wieder ein bisschen erweitern.

Ich glaube, viel wichtiger oder genauso wichtig ist die Datenanalyse.

Und da gibt es inzwischen so tolle technische Möglichkeiten, die Arbeitsschritte nachvollziehbar zu machen und sozusagen auch reduzierte Daten zwischendurch, also teilverarbeitete Daten nachher abzuspeichern und das sozusagen mit einem Workflow-Protokoll zu verbinden, dass man hier diese Mumpitz-Geschichte natürlich viel besser bewältigen kann, indem man sagt, ich hinterlege, was ich mit den Daten gemacht habe.

Und der andere Wissenschaftler kann einfach sagen, ich habe aber eine viel bessere Idee, ich will eine andere Formel ausprobieren.

Und dann ändert er einfach die Zeile in der Software, in dem Programm und rechnet das halt mit einer anderen Formel und guckt dann, was kommt raus.

Und das ist natürlich das, was Wissenschaft machen soll.

Und da muss man deswegen nicht noch mal messen gehen oder ein Experiment machen oder eine neue Probe herstellen oder was auch immer.

Und das ist ja richtig cool, wenn ich mit einmal erzeugten Daten verschiedene wissenschaftliche Modelle probieren und ausprobieren kann oder vielleicht später mit neuen Erkenntnissen einfach das Ganze auf ein neues Niveau heben kann.

Dazu muss ich aber auch erst mal wissen, dass diese Daten überhaupt existieren.

Wisst ihr das voneinander nicht?

Ihr seid beide Physiker.

Wisst ihr voneinander nicht, welche Experimente ihr gemacht habt und welche Daten ihr erzeugt habt?

Das hängt, glaube ich, sehr stark daran, wie sich bestimmte wissenschaftliche Felder voneinander abgrenzen.

Das ist eben auch eines der Ziele von der NFTI, sozusagen eine übersichtliche Auffindbarkeit.

Das hatten wir eben schon, also dass es um das FAIR-Prinzip geht, wo eben findable, auffindbar und zwar eben auch nach Maschinen lesbaren Prinzipien auffindbar, eben das erste dieser Prinzipien ist.

Und da eben eine übergeordnete Kataloge auf einem Metaniveau zu schaffen, die eben das Mögliche machen, auch Sachen aus anderen Feldern, von denen man eben nicht quasi innerhalb der schon existierenden Werkzeuge des eigenen Feldes wusste, ist eben eines der Ziele.

Das heißt, dass die Sozialwissenschaftler vom UFZ auch tatsächlich mal auf die Daten der Physiker von Jülich zugreifen können?

Genau, das klingt jetzt so in der Form vielleicht ein bisschen abstrus, aber wenn man jetzt an andere Sachen denkt, also zwei Sachen, die mir immer einfallen, also zum einen, wir haben jetzt die Erlebnisse mit der Pandemie so ein bisschen hinter uns oder machen sie noch in den Resten durch.

Und da sieht man schon, dass an der Stelle ganz viele soziale und medizinische Fragen natürlich einfach zusammenfinden.

Zum Beispiel, wie beeinflusst Verhalten und wie beeinflusst Mobilität irgendwie eine Ausbreitungsdynamik?

Und das ist natürlich was, wo plötzlich ganz viele Felder zusammenkommen und ganz viele Felder, die vorher nicht so richtig viel miteinander zu tun hatten, weil ein Epidemiologe, klassischer deutscher Epidemiologe oder Epidemiologin, die irgendeine Kohortenstudie mit irgendwelchen Leuten von zwei Jahren bis 70 Jahren irgendwie seit Ewigkeiten betreut, natürlich bisher nicht mit solchen Fragestellungen, wie Google irgendwie Mobilitätsmist quasi konfrontiert war.

Da stecken jetzt schon wieder ganz viele Sachen drin, nämlich die Abhängigkeit von Internetkonzernen zum Beispiel, sieht man an der Stelle.

Also möchte man quasi eine Demokratisierung und deshalb, also ich glaube, da können wir an ganz vielen Stellen darauf zurückkommen, das hat ganz viel, also es klingt als total trockenes Thema, dieses NFDI, aber wir können ganz viele Beispiele machen, wo das eben ganz viele gesellschaftliche Implikationen hat.

Nicht nur das Vertrauen in die Wissenschaft, was wir vorher schon angesprochen hatten, sondern eben auch ganz viele andere Fragen oder andere Sachen, auch Klimawissenschaft zum Beispiel, wo ganz viele Sachen irgendwie zusammenkommen, wo es ganz viele neue Messverfahren im Laufe der Zeit jetzt gab, in den letzten 10, 20, 30 Jahren, wo eben mit Mitteln der zum Beispiel Teilchenphysik oder Hadronphysik mit Hilfe von Neutronen irgendwelche

Feuchtigkeitswerte im Boden messbar gemacht wurden oder eben neue Detektortypen auf Satelliten zur Anwendung kamen, wo man plötzlich Sachen messen musste und konnte, die man vorher nicht messen konnte.

Also finden plötzlich ganz verschiedene Sachen zusammen, die eben plötzlich offen auf ihre Daten zugreifen mussten.

Und vielleicht ist da auch, ich weiß nicht, ob das woanders in deinem Podcast schon mal auftauchte, das Copernicus-System von dem European Center for Medium-Term Weather Forecasts, ein umständlicher Name, schon was, was vielleicht schon mal aufgetaucht ist, weil das finde ich ein ganz, ganz großes positives Beispiel, die eben schon vor langer Zeit damit angefangen haben, in einem einfach zugänglichen System ein komplettes Klimamodell der Erde in einer Reanalyse, also quasi das gleiche Modell, das man für die Wettervorhersage verwenden würde, einfach nochmal an die existierenden Messdaten gefittet, nicht als Vorhersage, sondern eben als bestmögliche Approximation, dass das der gesamte Zustand der Atmosphäre der Erde ist zu einem bestimmten Zeitpunkt und den einfach öffentlich zugänglich gemacht haben.

Und da kann man ganz viele Sachen damit machen.

Und zum Beispiel jetzt wiederum Leute, die Vorhersagen dafür machen oder auch Analysen, wie man unseren Energiebedarf mit erneuerbaren Energien decken kann, die brauchen so was.

Und da hängen plötzlich ganz viele Sachen zusammen.

Und von daher finde ich dieses Copernicus-System ein ganz positives Beispiel, was eben schon funktioniert vor der NFTI.

Und was genau diese Ideen an zusammenzubringen verdeutlicht und was eben auch zeigt, was für einen unheimlichen gesellschaftlichen und wirtschaftlichen Impact dann auch so was haben kann, wenn das funktioniert.

Wenn du jetzt aber alle Daten demokratisierst, das wird natürlich auch ein

überkomplexes, unübersichtliches, endimensionales Schachspiel, das man da mit Daten dann treiben kann.

Und da brauche ich dann ja wieder irgendwas, wodurch ich mich zurecht finde.

Und das ist ja im Zweifelsfall die Verschlagwortung.

Wie viel Platz für Verschlagwortung ist denn in den Metadaten?

Weil ihr müsst ja da auch irgendwas vorgeben, dass das nicht aussieht wie auf Twitter.

Ja, das ist ein gutes Beispiel.

Also ich glaube, auch das ist etwas, wo ich sagen würde, ja, das kann ich jetzt vielleicht jedenfalls nicht sagen, wie das am Ende aussehen wird.

Aber wichtig dabei ist, dass man sagt, ich habe ein gemeinsames Verfahren, auf das ich mich einigen kann.

Weil natürlich ist es so, dass die Daten jetzt von einem Teilchenphysik-Experiment, an dem ich arbeite, eine andere Struktur haben als die Daten von Neutronen-Streuungsexperimenten von Astrid und jetzt wiederum andere Strukturen als jetzt so ein Klimamodell in Kopernikus.

Von den Sozialwissenschaftlern nur ganz zu schweigen.

Und von denen ganz zu schweigen, genau.

Und auch da gibt es spannende Sachen.

Also NFDI for Culture kann ich dir sehr empfehlen.

Das sind hochspannende Sachen auf jeden Fall.

Aber zurück, dass man sich mehr quasi auf ein gemeinsames Verfahren einigt, wie man diese Beschreibung zugänglich macht.

Also ein gemeinsames Format darin, wie ich überhaupt mal die Formatbeschreibung für die Metadaten finde.

Und von da aus hangelt man sich dann sozusagen von oben nach unten durch, dass man eben durch verschiedene Abstraktionslevel den eigentlichen Daten immer näher kommt.

Und das andere Spannende daran ist auch, dass was wir mit Daten meinen, das hat Astrid, glaube ich, eben auch schon angerissen, kann ja auch ganz unterschiedliche Abstraktionsniveaus wiederum haben.

Dann kann ich sagen, ich habe einen Detektor, also nehme ich jetzt vielleicht mal eine Kamera als einfaches Beispiel.

Dann habe ich irgendwelche Pixel und ich lese aus, wie viele Elektronen ich innerhalb von einem bestimmten Zeitraum, bei einer bestimmten Temperatur des Sensors und einer bestimmten Einstellung der Optik, das wären jetzt die Metadaten, eben aus diesem Sensor ausgelesen habe.

Und jetzt will ich vielleicht nicht irgendjemandem diese Zahlen an Elektronen in die Hand drücken, weil vielleicht kann derjenige oder diejenige damit nichts anfangen.

Und dann kann man eben schon wieder anfangen zu sagen, okay, ich weiß, ich habe einen Filter davor, der macht irgendwelche Farbkorrekturen, also der filtert die Farben.

Und jetzt rechne ich erstmal diesen Filter raus und mache aus diesem Bild überhaupt mal ein Farbbild.

Und das ist dann ja sozusagen schon eine verarbeitete Niveaus, aber auch Daten.

Und jetzt ist vielleicht jemand da drin interessiert, ein triviales Beispiel oder beliebiges Beispiel, wäre es auf dem Bild zu sehen, dann könnte ich sagen, ich lasse irgendeine Bilderkennung darüber laufen und zähle, wie viele Menschen, Hunde und Katzen ich da drauf gefunden habe.

Das sind auch wieder Daten.

Und das ist was anderes, was halt schon zeigt, dass es gibt nicht die Daten, also auch nicht in einem Feld kann es nicht die Daten geben, sondern in den Metadaten muss gesagt werden, welche Abstraktionsform der Daten habe ich da an, was hat er dazu geführt sozusagen.

Ich glaube, es wird helfen, ich mache mir gerne so ein Kochbeispiel zu dem Ganzen.

Das Ganze geht ja darum, wirklich, wie wir es glaube ich jetzt schon viel gesagt haben, die Methodik von dem Ganzen voranzubringen.

Und deshalb finde ich es immer ganz lustig, quasi mit einer Analogie anzufangen, die sozusagen alte Formen von Wissenschaft beschreibt.

Und das Beste, was mir dazu einfällt, ist immer Wolfram Siebeck.

Achso, ich weiß nicht, ob denn alle deine Hörerinnen und Hörer noch kennen.

Ja, legendärer Gastronomiekritiker, der einige Mitschnitte hinterlassen hat, die gelegentlich in Podcasts in Deutschland zu hören sind.

Auch das, umso besser.

Und jetzt ist es so, wenn man so eine Wolfram Siebeck-Kolumne im Zeitmagazin gelesen hat, dann weiß man quasi nachher, dass er irgendwo war, wo er unglaublich grandiose Sachen gekocht hat oder gegessen hat oder unglaublich schlechte Sachen vorgesetzt gekriegt hat.

Und dass es also zum Beispiel der absolute Wahnsinn war und dass man es auch mal ausprobieren soll.

Und das wird dann veröffentlicht.

Und das ist jetzt natürlich ein Horrorszenario sozusagen.

Du meinst datentechnisch ein Horrorszenario?

Genau.

Also ich weiß nicht, was überhaupt gemacht wurde.

Also ich weiß nicht, welche Rohdaten verarbeitet wurden.

Also von welchem Feld, welche Zutaten, in welchem Restaurant oder in welcher Küche, mit welchen Werkzeugen, von wem Kochrezept.

Von wem.

Ich weiß nicht, was dabei rauskam, nach welchen Kriterien abgeschmeckt wurde.

Ich weiß nicht, nach welchen Kriterien verkostet wurde.

Also das war sozusagen bei der Analyse des Ganzen.

Und das Ganze wird dann nachher eben gedruckt und ich muss das glauben.

Und jetzt will ich natürlich nicht sagen, dass irgendeine Wissenschaftlerin oder Wissenschaftler, den ich kenne, so Wissenschaft macht.

Aber das ist so ein bisschen jetzt einfach so ein Klischeebild von der Wissenschaft, von der wir weg wollen.

Das Gegenteil davon.

Na, aber am Ende geht Wissenschaft doch auch so, dass es in der Küche natürlich wesentlich, ja wie nennt man das denn, wesentlich mehr fuzzy ist.

Das ist ja klar.

Wir können jetzt durch Beispiele gehen sozusagen, durch Stufen.

Und jetzt kann man sagen, okay, wo sind wir eigentlich?

Also wenn man dieses Horrorszenario, also nichts gegen Wolfram Siebeck persönlich, aber jetzt eben angewendet auf Wissenschaft, dieses Horrorszenario, wenn man davon mal loskommt, wo sind wir eigentlich wirklich?

Dann kann man sagen, dass in vielen Bereichen wir an der Stelle sind, wo wir sagen, okay, jeder und jede, die irgendwie ihre eigene Küche betreibt sozusagen, braucht entweder ein eigenes Feld, also so ist es bei mir.

Also ich muss quasi wirklich das Feld selber bearbeiten, wo ich die Daten herkriege.

Entweder ich die Daten habe auch für mich.

Oder hat bestimmte Methoden, auf irgendwelche Daten zuzugreifen, die mehr oder weniger offen sind.

Und dann wird typischerweise das Kochrezept, also quasi wie die Daten genommen und im Experiment verarbeitet werden, vielleicht auch wie sie analysiert werden, mehr oder weniger abstrakt, relativ klar hingeschrieben, jetzt im guten Fall.

Und es wird das Ergebnis in irgendeiner Abstraktionsform, also zum Beispiel jetzt sagen wir mal als Teilzutaten, dass man sagt, okay, ich habe nachher irgendwie einen Kartoffelbrei und einen Steak, die sozusagen separat mal irgendwo

abgelegt werden und die jemand mal probieren kann sozusagen.

Das heißt, ich habe nachher die Daten, die ich da rauskomme, in irgendeiner Abstraktionsform auch veröffentlicht.

So dass man nicht nur auf den Plot gucken kann, irgendwelche Zahlen ablesen kann, sondern wirklich irgendetwas, was mit den Daten gemacht wurde, irgendein Verarbeitungsniveau wirklich nachgucken kann.

Also so ist es bei mir im Feld jetzt seit einiger Zeit.

Aber das entspricht natürlich noch lange nicht den Kriterien, die wir vorher jetzt aufgezählt haben, wo man dahin will.

Ich frage mich die ganze Zeit, wie definiert man ja im Grunde, wie so ein Datensatz aussehen muss?

Weil der hat ja unendlich viele einzelne Datenpunkte, die dann da zusammengefasst werden müssen.

Du kannst ja nicht sagen, so das ist jetzt der Datensatz und so muss der aussehen und alles, was ihr macht, muss da reinpassen.

Weil wenn morgen jemand was Neues macht, was macht der denn dann?

Also vielleicht als Beispiel, ich weiß nicht, ob du das kennst oder wie viele deiner Hörerinnen und Hörer das kennen, aber XML zum Beispiel wäre jetzt ein Beispiel davon.

Also Extended Markup Language.

Das ist quasi ein Format.

Manche kennen das daher, dass ja auch die Open Office und später glaube ich in einem etwas verhackstückten Format, ich kenne mich nämlich nicht so aus, auch

die Word Dateien, in so einer Art XML Format abgespeichert wurden.

Ein Format, das man aus mancher Anwendung kennt, wo quasi festgelegt wird, wie definiere ich überhaupt Felder, in denen Daten nachher abgelegt werden können.

Das heißt aber nicht, dass da schon drin steht, dass da nachher, um bei dem Kochrezept Beispiel zu bleiben, es darum geht, wie viel Gemüse und innerhalb des Gemüses wie viel Gurken und wie viel Zwiebeln und wie viel Obst und innerhalb des Obstes und so weiter drin steht, sondern dass da nur drin steht, wie schaffe ich so eine hierarchische Struktur.

Also wie schreibe ich es hin, dass ich jetzt Gemüse habe und dass es innerhalb des Gemüses Unterabteilungen gibt.

Und wie schreibe ich es hin, dass es ein Ergebnis geben kann, was aus dem Kochen rauskommt, also irgendein Gericht und dass das Gericht aus verschiedenen Subbestandteilen zusammengesetzt ist.

Und wie schreibe ich es hin, wie eine Prozedur aussieht, wie ich aus den Zutaten zu diesen Abschnitten komme.

Und all das ist frei definierbar.

Aber das Format, wie ich so eine Abteilung schaffe und wie ich innerhalb dieser Abteilung Unterabteilungen schaffe und wie ich hinschreibe, wie dieses Feld heißt, das ist festgelegt.

Und vielleicht gibt es so ein bisschen so einen Hinweis darauf, wie man für unterschiedlichste Datengruppen quasi komplett unterschiedliche Datenstrukturen erschaffen kann, aber trotzdem ein maschinenlesbares Format hat.

Für XML gibt es viele Werkzeuge, die das lesen können in verschiedenen Sprachen, also verschiedenen Computersprachen, sodass ich quasi mir durch

Lesen das Dokument selbst feststellen kann, welche Struktur das Dokument hat.

Und vielleicht ist das quasi so ein bisschen so ein Beispiel.

Und das übersetze ich vielleicht so ein bisschen in dieses Kochbeispiel.

Dass man sagt, in unterschiedlichsten Ländern kochen die Menschen mit den allerverrücktesten, unterschiedlichsten Sachen.

Und Insekten gibt es jetzt in der EU seit ein paar Tagen als zugelassenes Lebensmittel.

Aber wie ich es festlege, dass es zugelassene Lebensmittel gibt, das muss also quasi in einer standardisierten Weise passieren.

Aber am Ende, wenn ich das sozusagen jetzt aufnehme, das Beispiel, ist das ja aber schon wieder eine Reduktion und limitiert deinen unendlichen Raum.

Also wenn ich jetzt zum Beispiel sage, kochen...

Ich hatte gerade angefangen, mich wohlfühlen.

Ja, nein, also ich wollte den Vorteil dieser Vereinbarungen, wie schreibe ich was auf, nochmal praktisch unterstützen.

Also beim Kochen ist doch immer das Typische, eine Prise Salz.

Mein Gott, wie viel ist eine Prise Salz?

Das ist genau das, was ich zwischen die Fingerspitzen kriege.

Genau.

Und am Ende habe ich aber nur eine limitierte Anzahl von Einheiten bei diesem Kochbeispiel.

Ich habe also die Prise und vielleicht Gramm und Kilogramm.

Und dann habe ich vielleicht noch Energie, Watt und eine Zeit.

Wie lange muss ich kochen?

Oder wie lange muss ich rühren oder so was?

Aber ich habe natürlich ganz viele Dinge nicht.

Also alle Einheiten meiner Wissenschaftswelt tauchen da überhaupt nicht auf.

Und vieles andere, was im Leben auftritt.

Das heißt sozusagen allein diese Überlegung, was brauche ich dafür?

Welche Kategorien habe ich?

Und welche Einheiten habe ich dafür?

Limitiert ja schon mal den Satz an Daten und gibt sozusagen einen reduzierten, ich würde jetzt meiner Welt sagen, Raum, in dem ich mich bewege.

Und damit schaffe ich sozusagen das, was wir Communities nennen.

Also jedes Wissenschafts- oder Lebensfeld hat dann sozusagen seine eigene Vereinbarung und gibt damit eigentlich wieder einen verständlichen und auch überblickbaren Zugang zu den Menschen.

Und das ist dann auch das, was ich so sehr gerne möchte.

Ein überblickbarer und überblickbarer Zugang zu der Benutzung dieser Daten.

Deswegen finde ich das eigentlich ganz schön, das Beispiel.

Aber wenn ich innerhalb zweier Communities, ich nehme jetzt mal wieder die Soziologen und ich nehme die Physikerinnen, wenn ich innerhalb dieser zwei Communities abgegrenzte Räume habe, in denen die sich bewegen können und möchte, dass aber beide auch miteinander reden können oder ihre Daten miteinander austauschen können, brauche ich in der Mitte doch wieder ein Austauschobjekt, das sich miteinander übersetzt.

Das ist richtig.

Und dann muss man sich halt über diese Vereinbarungen unterhalten.

Und wenn es geht, auch Einheiten, also jetzt wieder in meiner Welt Einheiten verwenden, die vielleicht auch kompatibel sind.

Das mit der Prise ist vielleicht schwierig, aber Kilogramm und Stunde und Minute ist ja dann doch vielleicht was, was communityübergreifend ist.

Und dann gibt es ja diese schöne Geschichte der sogenannten Konverter.

Also kann natürlich von zwei, aus zwei Fachfeldern, am Ende ist es ja nicht nur die verschiedenen Felder, sondern vielleicht auch Wissenschaftler, die sich nicht auf diese Vereinbarung einigen können.

Dann kann ich natürlich anbieten, dass ich sage, ich mache Übersetzer zwischen diesen verschiedenen, dem, was wir Metadaten nennen, der das eine in das andere überführt und dann für beide die Daten lesbar macht.

Und das ist ein großer Teil dieser Arbeit.

Das Kochbeispiel vielleicht, man könnte sagen, wenn man ganz ursprünglich ist, dann fängt man beim Getreide an und die Bäckerin oder der Bäcker hat vielleicht eine eigene Mühle.

Natürlich gibt es das sogar für die normale Hausküche.

Und malt sich das Getreide, genauso wie manche Leute sich natürlich auch den Kaffee selber mahlen.

Das ist einfach dann was ganz Besonderes.

Man kriegt genau das raus, was man haben will, und malt sich den selber.

Aber jetzt quasi der, die normale Alltagskochköchin, hat halt nicht die Getreidemühle.

Das heißt, dann muss ich durch wissenschaftliche Arbeit, also in weiterer Sinne eben auch durch Kommunikation, einigen auf ein Format, auf das sich alle wiederum einigen können.

Und dieses Format ist halt das Mehl.

Und so geht man, wie wir eben schon gesagt hatten, in vielen Beispielen durch verschiedene Abstraktionsniveaus.

Sachen, die ich auch selber jetzt kenne in dem Bereich.

Ich habe mich auch ein bisschen mit verschiedensten Formen von Windenergieerzeugung beschäftigt, wissenschaftlich.

Ist auch eine Publikation dabei herausgekommen.

Und da geht es eben genau um solche Sachen jetzt.

Wie kriege ich jetzt irgendein Muster, wie jetzt ein bestimmtes Gerät zur Erzeugung von Windenergie jetzt wirklich aus dem Windfeld Energie entzieht?

Also eine rein physikalische oder ingenieurwissenschaftliche Frage.

Wie kriege ich die jetzt wirklich mit den Daten von so einem Klimamodell

zusammen?

Und dann tauchen alle möglichen lustigen Probleme auf.

Zum Beispiel sind die Höhenangaben jetzt in diesem Klimamodell in Drücken.

Weil der Luftdruck nimmt nach oben ab.

Das heißt, umso kleiner der Luftdruck ist, umso weiter oben bin ich.

Jetzt ändert sich der Luftdruck aber ständig.

Das heißt, ich muss noch ganz viele andere, also quasi Daten oder Metadaten, das kommt dann schon so ein bisschen auf die Perspektive an, daraus entziehen, zum Beispiel Temperaturen und Luftfeuchtigkeiten, damit ich überhaupt ausrechnen kann, an einem bestimmten Ort, zu einem bestimmten Zeitpunkt, welcher Luftdruck hat da welcher Höhe entsprochen.

Und all die Sachen muss ich dann nachher verheiraten mit eben diesem physikalischen Modell von dem Gerät zur Erzeugung von Windenergie, um eben zu sagen, wenn in der und der und der Höhe, an der und der Stelle, so und so ein Wind ist, also der Wind sich auch so und so dreht, wenn ich von oben nach unten gehe, oder so und so viel stärker wird, wenn ich von unten nach oben gehe, wie viel Energie kann ich dann erzeugen?

Und das sind eben einfach Sachen, wo ich dann nachher wirklich natürlich reden muss oder selber darüber nachdenken muss.

Wichtig ist aber, dass ich verstehe, was da dargelegt ist.

Also jetzt in dem Beispiel, das Astrid vorher sagte wieder, da muss jetzt irgendwie stehen, ist der Luftdruck in Hektopascal oder in bar angegeben.

Wenn ich das nicht weiß und durcheinander bringe, dann bin ich halt im Weltall, aber da ist ja gar nichts.

Das klingt nach einem aberwitzigen Aufwand, sich sowas auszudenken.

Wie weit, Astrid, seid ihr mit all dem?

Also irgendwann werdet ihr ja wahrscheinlich auch mal sagen, jetzt ist es fertig, jetzt rollen wir es aus, wie so ein Produkt oder so.

Wie lange arbeitet ihr da schon dran?

Wann werdet ihr damit fertig sein?

Wann wird das benutzbar sein?

Oder ist das so eine Work in Progress?

Also das ist nicht so einfach zu beantworten.

Das Projekt selber, also unser Projekt, was DAFNE heißt, also DAFNE for NFTI, Daten aus Neutronen- und Photon-Experimenten, was vielleicht ein bisschen speziell ist, weil sich die Wissenschaftler über die Sonde, mit der sie messen, treffen und nicht aus ihrer Wissenschaftsrichtung.

Das heißt, wir versorgen am Ende Wissenschaftler mit Experimenten und Daten aus wirklich der ganzen oder einer großen Breite der Forschungslandschaft.

Von Biologie, Pharmazie, Archäologie, Chemie, Physik, Maschinenbau, die Turbinen sind so das eine der Beispiele oder eben auch In-situ-Prozesse oder Katalyse-Prozesse, bei denen man während der Katalyse nachgucken möchte, was denn da passiert.

Das heißt, wir versorgen sehr viele verschiedene Kulturen auch in der wissenschaftlichen Arbeit, also wie man miteinander redet, wie man vorgeht und wie man auch mit Daten arbeitet.

Und das hat die Sache am Anfang ein bisschen aufwendig gemacht und sozusagen den Zeitraum, sich zu finden und miteinander zu sprechen, ein bisschen länger gemacht, als wir vielleicht erwartet haben.

Und da komme ich auf Philipp zurück.

Das Miteinander sprechen und sich einigen und sich zu einigen, was bedeuten auch bestimmte Begriffe, ist ein großer Teil dieses Aufwandes.

Aber tatsächlich sind wir jetzt in der Phase, wo wir die ersten Produkte tatsächlich zum Testen bekommen, wo wir Konzepte haben für bestimmte Einrichtungen, wie denn diese Datenaufnahme, die Katalogisierung der Daten, welche Begriffe müssen denn zum Suchen und zum Finden aufgenommen werden für die verschiedenen wissenschaftlichen Fachrechnungen und die verschiedenen Techniken, mit denen wir arbeiten.

Das ist jetzt sozusagen konzeptionell, zumindest in unserem Beispiel und einem Beispiel schon fertig.

Und wir sind tatsächlich auch schon dabei, dieses System zu testen und hoffen, dass wir das für die Neutronenquelle in Garching im nächsten Sommer den Nutzern auch anbieten können.

Aber fertig ist natürlich eine Frage.

Ich weiß nicht, ob ich mich jetzt zu sehr aus dem Fenster lehne, aber das ist natürlich am Ende nie fertig.

Fertig ist, wenn alle mitmachen.

Dann hat man viel erreicht.

Ja, dieses englische Wort "awareness" ist ein ganz großes Thema.

Wissenschaftler sind ja auch sehr verschieden und oft auch eigen.

Ob man auch wirklich alle ins Boot bekommt, das wäre schon wirklich viel, wenn man die meisten ins Boot bekommt und wenn man vor allem in der jungen Generation ein Verständnis dafür entwickelt, warum das gut und nötig ist und warum es am Ende auch Effizienz schafft.

Ich hatte die Diskussion tatsächlich gestern und vorgestern in einer Klausur von einem großen wissenschaftlichen Verbundprojekt, wo genau diese Frage diskutiert wurde, der Aufwand, den wir da treiben müssen.

Und ist es denn wirklich sinnvoll, das zu machen?

Und das bindet doch jetzt unsere schöne wertvolle Arbeitszeit, die wir sowieso nicht haben mit den vielen befristeten Stellen, wo dann in endlicher Zeit eben unbedingt Ergebnisse kommen müssen.

Aber was am Ende natürlich passiert, ist, dass ich mit diesen Vereinbarungen, auch mit einem elektronischen, sogenannten elektronischen Notebook, was also direkt die Daten aus den Instrumenten zieht vielleicht, die ich gar nicht einzeln aufschreiben muss, wo die Metadaten durch diese Vereinbarung schon vorgegeben sind, natürlich für viele Neuankömmlinge die Dokumentation und die Datenerfassung auch vereinfacht über die Standards und damit schneller mache und viel schneller in die Datenanalyse bringen kann, anstelle mich damit zu befassen, was habe ich denn da eigentlich gemessen?

Das heißt, das eine ist, dass man es richtig macht und auch sozusagen immer zielorientiert guckt, wo ist denn der Gewinn?

Mache ich das jetzt wirklich nur, weil es eine politische Forderung ist?

Oder wie kann ich das so gestalten, dass ich tatsächlich auch für die wissenschaftliche Arbeit einen Gewinn oder eine neue Qualität erzeuge?

Das andere ist aber, es kostet auch Ressourcen.

Es kostet viele Ressourcen an Zeit und Diskussion, gerade in der Anfangsphase.

Und es erfordert Techniken, die am Ende einmal aufgebaut, aber dann gewartet werden müssen.

Wer irgendwie mit Programmen und Software zu tun hat, und das ist jeder Bürger, der irgendein Telefon benutzt, weiß, dass dann irgendein Update kommt und es sieht alles anders aus.

Und die Tatsache, dass das passiert, heißt auch, dass Menschen daran arbeiten, nicht nur, weil das nächste Bild vielleicht schöner ist oder das nächste Outfit von irgendeiner Oberfläche, sondern Technik verändert sich und man muss Dinge an sich verändern, die Technik anpassen.

Das heißt, wir binden damit auch Ressourcen und werden auch neue Ressourcen brauchen, insbesondere dafür, dass diese Lösungen nachhaltig sind.

Wenn wir das Ganze nur betreiben als Projekt, was das einmal initiiert, aber nicht darüber nachdenken, wie verstetigen wir den fairen Prozess, dann würden wir in Probleme laufen.

Aber soweit ich verstehe, ist diese Verstetigung auf jeden Fall in der Diskussion in der NFDI.

Und ich verstehe auch, dass man sich in der Politik darüber Gedanken macht und dass das auch verstanden ist, das Problem.

[Siebert] Verstehe ich das richtig?

Es gibt jetzt nicht so diese eine übergeordnete Instanz, die im Grunde festlegt, wie diese Daten strukturiert sein müssen, also so etwas wie NF-DIN, sondern jedes Institut macht das irgendwie, aber ihr redet die ganze Zeit miteinander darüber, wie man es idealerweise machen würde, sodass dann am Ende gleichsam eine ideale Lösung emergiert.

[Giesing] Viele ideale Lösungen.

[Siebert] Viele ideale Lösungen.

[Hübner] Ja.

Also wie gesagt, man kann wirklich durch verschiedene Abstraktionsniveaus gehen.

Das ist schon ein bisschen das politisch vorgegebene Ziel, quasi einheitlichen Zugang dazu zu haben.

Aber es wird sich wahrscheinlich oder meiner Meinung nach sicherlich natürlich sehr schnell diversifizieren.

Wichtig ist eben, wie wir vorher schon ganz oft gesagt hatten, dass man nachvollziehen kann an jedem Schritt, wenn ich von einer Metaebene zur darunterliegenden Ebene gehe, was wird mir jetzt neu präsentiert, was ist das Format, was sind die Metadaten, was sind die Daten, unter welchen Bedingungen wurde was gemacht, sowohl Datennahme, Experimentdurchführung, Analyse und so weiter.

Und so ist es natürlich so, dass es überhaupt nicht möglich ist.

Keiner von uns kann jetzt für Medizin oder sonst irgendwas oder Kultur irgendwas festschreiben.

Und das muss sich von unten entwickeln.

[Siebert] Und das sitzt halt so die ganze Zeit in meinem Kopf, als gäbe es eben eine Entität, die sagt, so ist es, das ist der große Wurf.

Aber das ist gar nicht der Fall.

[Hübner] Genau deshalb, das ist wirklich das sehr Gute daran, die NFTI ist eine

übergeordnete Struktur, aber die NFT ist eben aufgebaut aus Konsortien.

Und diese Konsortien, die sollten sich eben ganz bewusst an ihren eigenen wissenschaftlichen Herausforderungen orientieren und aus sinnvoll gruppierten wissenschaftlichen Zusammenschlüssen entstehen.

Und das hat es ja vorher gesagt, Photonen-Neutron-Forschung ist eben sinnvoll, einfach weil die Experimente, also erstens, tolle Wissenschaft, deshalb ist es sinnvoll, zweitens, weil die Experimente einer einheitlichen Methodik folgen, mit einheitlichen oder ähnlichen Instrumenten durchgeführt werden.

Aber die wissenschaftlichen Fragen, die damit bearbeitet werden, die Felder, aus denen die einzelnen Wissenschaftlerinnen und Wissenschaftler kommen, ist hochgradig divers.

Und jetzt kann man eben andersrum hergehen und sagen, okay, ich orientiere mich jetzt weniger an den Instrumenten, sondern an Fragestellungen.

Da gibt es jetzt, glaube ich, je nach wissenschaftlichem Feld unterschiedliche Herangehensweisen.

Und ich glaube, da ist die NFT als solche eben auch sehr divers.

Aber es gibt, soweit ich mich erinnere, im Moment 26 genehmigte Konsortien.

Und die verteilen sich eben wirklich über alle möglichen Bereiche.

Also es ist wirklich eben Sachen, also aufgeteilt ist es nach Engineering Sciences, nach Humanities and Social Sciences, nach Life Sciences und nach Natural Sciences.

Und da sieht man ja schon so ein bisschen, es versucht wirklich, also ich kann dir jetzt nicht auswendig sagen, was fehlt, aber man sieht jetzt schon an der Übersicht, dass das ist nur die grobe Einteilung.

Diese 26 Konsortien, die teilen sich dann irgendwie ein und decken wirklich von Lebensmittel, Agrar, Klimawissenschaften, alles Mögliche.

Wie gesagt, NFT, I4Culture, finde ich ein sehr spannendes Beispiel.

Und wenn jetzt irgendwer, also ein Institut, eine Forscherinnengruppe sich denkt, Moment mal, wir wollen auch, dann können die sich bei der NFDI melden und sagen, wir würden gerne mitmachen.

Also das sind jetzt verschiedene Ebenen sozusagen.

Also die Nutzung all dieser Instrumente soll natürlich offen sein.

Es geht ja die ganze Zeit um offene Wissenschaft.

Ja, ja, aber wenn ich mir jetzt mein eigenes XML ausdenke, kann damit auch niemand eine Webseite darstellen.

Genau, natürlich, aber das ist jetzt wieder eine Frage von Kommunikation natürlich.

Aber das Funding, also die Förderung von bestimmten Projekten, die geht über die Konsortien und die Konsortien haben sich einmal bei der Beantragung gebildet.

Also es ist jetzt nicht so, dass jetzt jeder Einzelne dahin gehen kann und sagen kann, ich möchte jetzt auch Geld.

Aber der Prozess ist bewusst über viele Jahre angelegt.

Also es sollten 30 Konsortien insgesamt gefördert werden.

Soweit ich weiß, also ausdrücklich vielleicht aktuelle Zahlen, aber ich glaube 26 sind momentan gefördert, die sich über drei verschiedene Förderrunden erstreckt haben, also im Zeitraum von drei Jahren sozusagen ausgewählt wurden

Stück für Stück.

Und das sind im Grunde noch Plätze frei.

Und ich glaube, man hat das bewusst auch so angelegt, dass man sagt, das ist ein dynamischer Prozess, der von unten in irgendeiner Form mitgetrieben wird.

Und dann bilden sich eben im Laufe der Zeit Sachen heraus, die alle brauchen.

Da gibt es jetzt auch so ein Basis-Konsortium, Based for NFDI, wo eben bestimmte Sachen, die jeder braucht, gebündelt werden.

Und das sind zum Beispiel Sachen, an denen viele arbeiten oder die viele nutzen, eben Authentifizierungs- und Autorisierungsmechanismen.

Was wir viel machen in meinem Bereich ist eben föderiertes Computing.

Das heißt, wir haben nicht eine Computing-Infrastruktur, wo jetzt alle sich einloggen, um irgendwie ihre Daten zu analysieren oder darauf zuzugreifen, sondern überall auf der Welt gibt es Forschungszentren, die eben an was angeschlossen sind, was sich in unserem Fall Worldwide LHC Computing Grid nennt.

Und da sieht man schon wieder, dass die nationale Forschungszeitinfrastruktur und die eigentliche Art und Weise, wie Wissenschaft organisiert ist, sich so ein bisschen beißt.

Das war also für mich persönlich, muss ich sagen, auch am Anfang konzeptionell eines meiner größten Probleme, mir überhaupt vorzustellen, was wir da machen können.

Wir sind ja eingebunden in Strukturen, die viel größer sind, aber es macht natürlich trotzdem Sinn.

Natürlich hätte man genau das Gleiche auch auf europäischem Niveau machen

können.

Es macht aber natürlich Sinn, einfach irgendwo mal anzufangen.

Und jetzt ist es halt so, dass diese Initiative in Deutschland zustande kam.

Da beschwert man sich ja auch nicht und versucht es trotzdem so zu machen, dass es mit der Art und Weise, wie man real als Community in einem internationalen Umfeld arbeitet, eben auch verträglich ist.

Und um wieder darauf zurückzukommen, also diese Autorisierung und Authentifizierungsinfrastruktur, die brauchen wir natürlich auch in so einem Modell, das wir hier schon haben, um einfach zu sagen, bin ich überhaupt autorisiert, in Karlsruhe, am CERN oder sonst irgendwo in irgendeinem Forschungszentrum irgendwas laufen zu lassen oder ist mein Job autorisiert?

Also meine Analyseaufgabe, die ich als selbst laufender Compute-Prozess irgendwo abschicke, ist, der autorisiert sich selbst, irgendeinen Ort irgendwo auf der Welt zu suchen, den ich im Zweifel gar nicht weiß, wo der ist, und am Ende mir irgendwie das Ergebnis zurückzuliefern.

Und all diese Sachen sind eben für, die brauchen fast alle, und so macht es eben Sinn, die auch zu bündeln.

Aber das ist eben was, was erst in der dritten Runde kam, einfach um auch zu gucken, wie baut sich das auf?

Was brauchen die meisten Leute?

Was kann man sinnvoll bündeln?

Und was ist eben so, dass eben dann doch, NFTI for Culture, um in dem Beispiel zu bleiben, dann doch eben wirklich aus fundamentalen Gründen ganz andere Formate braucht, als irgendjemand sonst.

[Siebert] Vielleicht kommen wir doch nochmal auf die Struktur der NFTI zurück, weil du hast in der Anmoderation ja gesagt, dass es ganz viele Mitglieder gibt, und Institutionen, Universitäten, bin ich jetzt überfragt, was noch?

[Hier ist eine Wissenschaftsorganisation, Universitäten, Hochschulen, außeruniversitäre Forschungseinrichtungen, Fachgesellschaften und Vereine.]

[Siebert] Genau, und davon irgendwie über 200.

Tatsächlich ist der NFTI-Verein genau das, wo diese ganzen Organisationen Mitglied sind und auch bestimmte Entscheidungen gemeinsam treffen.

Und dieser Verein hat auch sozusagen ein Direktorat, was eine Art Dachorganisation ist, und die Arbeit in der zweiten Säule, die parallel dazu läuft, und das sind nämlich die Konsortien, koordiniert und bündelt und unterstützt.

Das heißt, NFTI hat zwei Säulen, den Verein mit den Mitgliedern institutioneller Art, auf der einen Seite, und auf der anderen Seite in den Konsortien, die gefördert sind über ein DFG-Verfahren, die sich in diesen drei Runden beworben haben und am Ende auch gefördert werden.

Und jetzt gibt es diese Schnittstelle zwischen den beiden Säulen, die nämlich über die sogenannten Sektionen erfolgt.

Und da gibt es eben eine Sektion, die beschäftigt sich mit den Metadaten und guckt jetzt, wie in den verschiedenen Konsortien Metadaten gehandelt werden und versucht, dort das Gespräch zu moderieren oder auch technische Lösungen zu vergleichen, anzubieten, aufzusammeln, weiterzuentwickeln, die dann gute Lösungen protegieren, Leuten helfen, die vielleicht technisch noch Probleme haben.

Und das Gleiche eben genauso über die federierten Infrastrukturen oder über die ethischen und sozialen und legalen Probleme oder auch eine Sektion Industrie.

Und jetzt fehlt mir die fünfte, es sind nämlich fünf, da müssten wir vielleicht

nachgucken.

Ja, steht dann in den Show Notes, können alle nachschauen.

Genau, steht auf der Webseite auch.

Das heißt, es gibt eine Verflechtung zwischen diesen beiden Säulen, aber im Augenblick wird im Wesentlichen diese praktische Arbeit in den geförderten Konsortien getätigt, wo tatsächlich die Wissenschaftler sich engagieren, die eben praktisch sich einbringen wollen und auch Beispiele zeigen wollen, Vorreiter sein wollen, um den anderen zu zeigen, wie geht es und wo gewinnen wir.

Es war vor wenigen Wochen die erste große Konferenz, wo sich alle diese Konsortien getroffen haben.

Und es war extrem überraschend, wie spannend diese Diskussionen waren, wie viele Prozesse ähnlich verlaufen, an wie vielen Stellen es ähnliche Probleme und Lösungen und Methodiken gibt, wo man eben extrem viel voneinander lernen kann, weil die Fachrichtungen so unglaublich verschieden sind.

Und ich finde das total spannend, abzuwarten und zu sehen, was sich da vielleicht noch entwickelt.

An wie vielen Stellen wir Synergien finden, mit denen vielleicht niemand gerechnet hätte.

Und das alleine ist schon ein Riesengewinn.

Wenn einem vor der Komplexität der Kopf platzt manchmal, und dann auf der anderen Seite zu sehen, wie viele Dinge am Ende doch systematisch ähnlich oder gleich sind, dann macht einem das total Hoffnung.

Und es weitet den Blick von den Beteiligten so ungemein, wenn man aus den anderen Fachrichtungen sieht, wie sie mit ihren Daten umgehen, welche Probleme sie haben und wo es dann doch manchmal wieder runterkocht auf

dieselben Prinzipien.

Das ist auch unglaublich bereichernd.

Das klingt, als wäre es so etwas Ähnliches wie ein Grassroots-Projekt, als würde tatsächlich die Arbeit, die Ideenfindung vor Ort auf deinem Computer, in deinem Büro, in deinem Institut stattfinden.

Und wenn du dann was Gutes gefunden hast, würdest du im Grunde die Hierarchie nach oben melden.

Habe ich das richtig verstanden?

Ja, den Begriff melden würde ich jetzt vielleicht nicht benutzen.

Ja gut, ich bin halt, ich sitze halt hier, ich habe halt nichts gelernt.

Ja, also ja, aber da ist natürlich was dran.

Also diese Idee, die mancher vielleicht auch von uns hatte, man macht da Vorschläge und nimmt den Besten und das machen dann alle, hat sich sozusagen relativ schnell als nicht machbar erwiesen und als zu optimistisch.

Aber tatsächlich ist es so, dass an ganz vielen Stellen gearbeitet wird, auch schon in den Konsortien und dann Erfahrungen ausgetauscht werden und in den Diskussionen halt gute Lösungen protegiert oder manchmal auch bessere Lösungen übernommen werden.

Und den Austausch gibt es nicht nur in den Konsortien, sondern auch zwischen den Konsortien.

Und ich weiß nicht, ob ich den Begriff Grassroot jetzt nehmen würde, weil wir sind ja schon innerhalb der DFG-Förderstrukturen mit all ihren administrativen und sonstigen Komplexitäten.

Aber von dem Teil der wissenschaftlichen Arbeit ist da ganz viel dran.

Aber es klingt jetzt halt nicht so, als würde weder innerhalb eines Konsortiums noch zwischen den Konsortien und irgendeiner übergeordneten Instanz irgendetwas vorschreiben, sondern als würde es entstehen aus sich heraus.

Würde ich schon unterstützen.

Ja, also bestimmte Zielsetzungen sind natürlich vorgegeben, allein schon durch die Förderausschreibung.

Aber wie das jetzt gemacht wird, ist eben was, was im Laufe der Zeit entsteht.

Ich glaube, das ist auch die einzige Möglichkeit, das zu machen.

Und es gibt ja auch andere Sachen, die auch langsam wachsen.

Also wenn man im europäischen Umfeld denkt, die European Open Science Cloud, EOSC, ist ja auch ein großes Projekt, was ganz ähnliche Zielsetzungen in manchen Bereichen verfolgt.

Aber auch da muss noch vieles wachsen.

Und auch da kann später mit den Sachen, die wir jetzt hier ausprobieren, vieles zusammenwachsen.

Und von daher ist es eben so, dass Forschung natürlich immer was ist, was sich entwickeln muss und was man nicht planen kann.

Das ist schon das eine triviale Beispiel, dass ich eine Erfindung nicht vorhersehen kann.

Also so wie ich bei Civilization in so einem Entwicklungsbaum nachgucken kann.

Wenn ich sagen will, ich möchte irgendwann mal irgendwie das Auto entwickeln,

dann muss ich vorher mal das Rad entwickeln.

Das ist natürlich nicht wie Forschung funktioniert, weil wenn ich noch nicht mal das Rad erfunden habe, habe ich ja keine Ahnung, was das Auto sein soll.

Und so ähnlich ist es hier halt auch.

So etwas funktioniert überhaupt nicht.

Wobei ein spannender Teil dieser Arbeit natürlich genau der ist.

Philipp hat gerade gesagt, wie Wissenschaft funktioniert.

Aber natürlich gibt es auch ganz viel Arbeit, die gemacht werden muss.

Und was man vielleicht jetzt mit Programmieren oder IT-Produkte schaffen, umschreiben könnte.

Und da gibt es natürlich schon die Idee, dass man eigentlich ein Work Package hat oder einen Milestone Plan.

Und wie heißt das?

Tickets, die man abarbeitet.

Und wo man natürlich schon sagt, da will ich hin.

Und die Mittel habe ich zurzeit zur Verfügung.

Und das arbeite ich jetzt Schritt für Schritt ab.

Und das sind schon auch zwei Welten, die ein bisschen unterschiedlich sich bewegen und unterschiedliche Vorstellungen davon haben, wie Dinge gestaltet werden oder sich entwickeln oder entwickelt werden können.

Und auch das ist eine der Herausforderungen, diese zwei Welten zusammenzubringen und gemeinsam gute Lösungen zu schaffen, auch mit den sehr unterschiedlichen Vorgehensweisen und Kulturen und Erwartungen manchmal.

Nein, das würde ich voll unterstützen.

Es hat auch viel damit zu tun, wie die Geschichte und die Diversität auch innerhalb der einzelnen Forschungskommunities schon ist.

Je nachdem, wie üblich es in der Community schon war, dass man eben auch langfristige Strukturen wirklich betreiben musste, dass es nicht eben Aufgabe von einer Doktorandin oder einer Postdoc ist, irgendwie mal eben irgendwas zu erforschen innerhalb von drei Jahren und das dann aufzuschreiben und dann irgendwo anders hinzugehen, sondern eben langfristig große Infrastruktur zu betreiben, die eben auch mit einem großen Computeraufwand verbunden ist.

Umso mehr ist man sowieso gezwungen, diese zwei Welten immer gleichzeitig zu denken.

Ich denke immer noch, wie gesagt, wo man hin will, kann man sich abstrakt vorstellen.

Wie man das löst, muss man erforschen.

Und wie man es dann erhält und wirklich als Produkt umsetzt, das hat dann wiederum viel mit dem System zu tun, was eben einfach Erhalt und Erschaffen von Softwareinfrastruktur ist, von dem Astrid gerade redete.

Ich weiß nicht, vielleicht ist in dem Zusammenhang auch interessant, nochmal auf mein liebes Kochrezeptbeispiel zurückzukommen.

Du hast Hunger, ne?

Es klingt jetzt so, als wäre es ein wahnsinns Aufwand.

Ich habe Hunger, sowieso, immer.

Aber es klingt jetzt so, als wäre es ein wahnsinns Aufwand und eine wahnsinns Ablenkung.

Aber wenn man sich dann vorstellt, wie kann ich jetzt vielleicht abstrahieren, wo wir hinwollten.

Wir hatten ja vorher das Beispiel, das abstrakte Beispiel, wo kommen wir her, wo wollen wir nie wieder zurück, auch wenn es tolle Texte sind.

Und dann das Beispiel, was ist es jetzt vielleicht gerade.

Dann kann man sich natürlich fragen, wie würde das jetzt aussehen, dieses Kochen, wenn wir das hätten, was wir uns da vorstellen.

Mein persönliches Beispiel dafür, da gibt es wahrscheinlich noch ganz unterschiedliche Vorstellungen, ist jetzt, dass man eben sagt, wo wir im Moment sind, ich habe ein sauber aufgeschriebenes Kochrezept mit einer sauber ausgeführten Durchführungsanleitung.

Wenn ich Glück habe, steht nicht eine Prise, sondern zwei Gramm Salz da.

Und ich kann in den Bauernladen gehen, ich kann in den Supermarkt gehen und ich kann irgendwie noch zur Tanke gehen, wenn ich noch was vergessen habe und mir alle Sachen zusammenstellen.

Und ich kann dann eben im beliebten Online-Laden mir alle Küchengeräte, die ich brauche, zusammenstellen oder dann halt noch den Extrakühlschrank irgendwo bei sonst irgendeinem Elektrofachmarkt kaufen.

Kann mir das alles selber aufbauen mit Schraubenzieher und allem Drum und Dran.

Kann das Kochrezept durchführen und kann dann das Ergebnis, wie gesagt, veröffentlichen, zur Verkostung freigeben und das ganze Prozedere veröffentlichen.

Das ist schon mal super.

Aber jetzt können wir sagen, okay, haben wir da wirklich viel gewonnen.

Und ich denke, wenn man sich jetzt vorstellt, wir führen das Ganze noch viel weiter und sagen, wir wollen wohin, wo wir den gesamten wissenschaftlichen Prozess als so eine faire Struktur sehen.

Dann würde man ja sagen, dass ich nachher in solchen Registries, wo ich eben die wissenschaftlichen Produkte ablege, nicht einfach nur das Kochrezept und die Anleitung finde, was für Sachen ich brauche und was für Sachen ich mir selbst beschaffen muss, sondern wo ich klicken kann und dann eben quasi bildlich alle notwendigen Bestandteile der Küche inklusive gefülltem Kühlschrank einfach direkt sich sofort bei mir zu Hause einfinden.

Ein Shop-Link meinst du?

Also auf der Computer Infrastruktur.

Hier kaufen.

Also ein Shop-Link genau mit instantaner Lieferung, mit allen Küchengeräten genauso eingestellt wie in der Durchführungsanleitung sowieso schon festgelegt, sodass es im Prinzip ein Einklick, also das ist natürlich jetzt alles ein bisschen idealisiert, aber ich denke einfach mal zu zeigen, wo kann das hinführen?

Warum kann das so eine große Freisetzung von Kreativität und auch Zeit all dieser Wissenschaftlerinnen und Wissenschaftler bedeuten, die das auch wirklich machen müssen?

Die meisten sind es ja Doktorandinnen, Doktoranden, Postdocs auf

Zeitverträgen, die also nicht mehr hergehen müssen und sich fünf verschiedene oder zehn oder hundert verschiedene Programme irgendwo besorgen, runterladen, installieren, müssen die die Daten irgendwo runterladen, die versuchen müssen, dieses Metadatenschema zu verstehen, sondern die eben quasi wirklich eine komplette Maschinerie geliefert kriegen, also nicht nur faire Daten, sondern faire Wissenschaft.

Und dann man auf die Weise dann quasi einfach alle Küchenkomponenten hat, die man braucht, auf den Knopf drücken kann und dann, Astrid hat es ganz am Anfang mal gesagt, also im Workflow, der Workflow in diesen Küchenkomponenten so angelegt ist, dass man wirklich in genau vordefinierter Weise, im Idealfall automatisch, durch den ganzen Kochprozess geht, den geht es furchtbar un kreativ, ich komme gleich dazu, und am Ende ein nachprüfbares Ergebnis kriegt.

Jetzt warum ist das so toll?

Es befreit eigentlich Kreativität, weil jetzt habe ich unglaublich viel Zeit.

Ich kann mir überlegen, was ich jetzt mit meinem Steak und meinem Kartoffelbrei mache.

Ich kann wirklich überlegen, was von dem, was da passiert ist, hat zu dem geführt, was ich da als Steak und Kartoffelbrei gekriegt habe, kann anfangen, damit rumzuspielen, kann versuchen, neue Komponenten mir entweder selbst zu schaffen, hinzuzufügen, zu ändern und so weiter, oder mir vielleicht noch ein...

Nimm den Stampfer raus, tu einen Mixer rein.

Genau, jetzt könnte ich zum Beispiel jetzt einfach mir ein Kartoffel-Gratin-Rezept gleichzeitig noch runterladen, oder ich könnte mir auch ein Rezept zum Herstellen von Mehl, wenn ich jetzt auch noch irgendwie Nudeln dabei haben möchte, irgendwie besorgen, und dann könnte dann die Getreidemühle auch gleich noch mitkommen.

Und dann könnte ich das Ganze nachher als neues Paket, wenn ich sage, nee, also jetzt die Sache mit den Fettuccine hat aber viel besser geschmeckt als es mit dem Kartoffelbrei, das veröffentliche ich jetzt als neues Paket und biete es wieder genauso an für alle anderen zum instantan Ausprobieren sozusagen.

Das würde eine unfassbare Menge an Zeit, an wirklich freier geistiger Kapazität für echte Wissenschaft statt für quasi die Mechanismen, das aufrechterhaltende Mechanismen, die man braucht, bedeuten.

Und es gibt ganz viele Beispiele, wo das von unheimlicher Nutzbarkeit wäre.

Jetzt wieder auf ein wirklich aktuelles Beispiel, an dem wir auch arbeiten.

Ich hatte vorher gesagt, über die NFTI, also über Punch for NFTI, bin ich auch ein bisschen in die Epidemiologie gerutscht.

Zusammen mit Viola Priesemann machen wir Projekte.

Das ist natürlich etwas, was super spannend ist, wenn man aus einem anderen Bereich kommt, wenn man denkt, ich habe mich bisher mit dem Higgs-Poisson beschäftigt, aber jetzt mache ich endlich mal etwas gesellschaftlich Nützliches, auch wenn ich Grundlagenforschung für unfassbar kulturell wichtig halte.

Da hat man dann ja immer solche Sachen, das wurde während der Pandemie auch immer öffentlich diskutiert, in welcher Form liegen eigentlich die Daten vor?

Man hat irgendwelche Tabellen und in den Tabellen sind irgendwelche Artefakte drin.

Die bekannten Artefakte da zum Beispiel, mit denen man sich dann wirklich wissenschaftlich beschäftigen muss, sind zum Beispiel diese Wochenvariation, wo man einfach gesagt hat, Montag, Dienstag, Mittwoch sind die Zahlen immer viel höher, ich glaube, Dienstag, Mittwoch, Donnerstag sind die Zahlen immer viel höher, weil es nachmeldende Fälle vom Wochenende erst am Dienstag so richtig losgeht.

Und all das muss man da rauskriegen.

Und wenn jetzt jeder anfängt, von diesen Tabellen ausgehend Epidemiologie machen zu wollen und das vielleicht sogar mit hoher zeitlicher Auflösung machen möchte, um wirklich viel aus den Daten rauszukriegen, kann ich gleich nachher noch ein lustiges Beispiel erzählen.

Dann muss jede und jeder für sich anfangen, sich Mechanismen zu schaffen und selber zu schreiben, wie man diese Wochenvariationen heraus kalibriert.

Und die Arbeit muss nur eine oder ein oder einmal machen, richtig, und klar dokumentieren und die Mechanismen zur Verfügung stellen.

Und mit einem Klick müsste sich jeder die ganze Prozedur und das Ergebnis runterladen können.

Und das würde so unfassbar viel Zeit für echte Wissenschaft, die man damit gewinnen kann, bringen.

Und das ist eben was, was man in ganz vielen dieser Fälle, in allen möglichen Feldern sozusagen haben kann.

Und deshalb würde das Ganze sich wirklich unheimlich lohnen, auch wenn es erst mal unfassbar trocken klingt.

Aber es ist wirklich ein ganz, ganz wichtiges Metathema dafür, wie Wissenschaft eigentlich gemacht wird, wie nachprüfbar sie ist, wie effizient sie ist und wie sehr sie nachher dann auch vielleicht nach außen hin als was wirklich Vertrauenswürdiges klingen kann.

Weil ich glaube schon, dass wenn man jetzt wieder in dem Medizinbereich bleibt, also ich will mir das jetzt nicht zu eigen machen, aber es ist ja schon so ein bisschen eine triviale Sache gewesen, die sich immer durch vielleicht Twitter oder Ähnliches gezogen hat.

Ich weiß es nicht, dass halt bei medizinischen Studien immer so lange P-Hacking betrieben wird, bis halt irgendwas rauskommt, weil halt Doktorand A... P-Hacking?

Ach so, ja.

Also je nachdem, was für statistische Verfahren man verwendet, hat man ja den sogenannten P-Value, also quasi eine Probability dafür, dass wenn es keinen Effekt gäbe in Wirklichkeit, die Daten genauso signalähnlich oder noch signalähnlicher aussehen, als sie aussehen.

Wenn ich also eine Aussage dafür machen möchte, ich habe einen Effekt gefunden, dann muss ich beweisen statistisch, dass dieser Effekt nicht, statistisch nicht verträglich mit der Annahme ist, dass es den Effekt nicht gibt.

Und der P-Value, in verschiedensten Feldern verwendet man unterschiedliche Werte.

Das naive Beispiel ist eben $p < 0,05$, was eben heißt, die Wahrscheinlichkeit, dass der beobachtete Effekt so auftritt, wie ich ihn beobachtet habe, oder noch stärker, ist kleiner als 5%, wenn es wirklich so wäre, dass es den Effekt nicht gäbe.

Und dann würde man eben sagen, okay, dann habe ich was gefunden.

Das ist jetzt ein bisschen hoch, dieser Wert von 5%, weil, kann man sich jetzt schon vorstellen, wenn ganz viele Wissenschaftlerinnen und Wissenschaftler die unterschiedlichsten Studien machen und man 5% als Wahrscheinlichkeit ansetzt, dass man eine zufällige Entdeckung macht, die es eigentlich gar nicht gibt, dann kann man sich vorstellen, dass rein theoretisch, wenn man 1000 Paper schreibt, man 50 Veröffentlichungen hat von Entdeckungen, die es gar nicht gibt.

Und das ist eben so ein bisschen, da steckt jetzt eigentlich noch viel mehr dahinter.

Das ist jetzt nur so ein bisschen, da müssten wir jetzt nochmal...

Machen wir nochmal eine eigene Sendung drüber.

Also ein bisschen trivial.

Ich wollte nur sagen, diese triviale Ansicht, naja, der größte Teil von diesen Medizinstudien, was ich mir nicht zu eigen mache, sondern was jetzt nur einfach was ist, was man wirklich von vielen Leuten hört, von diesen Studien ist halt eh einfach nur Quatsch, weil irgendjemand mal was veröffentlichen musste, um halt irgendwie die Doktorarbeit fertig zu machen.

Und das ist ja was, was unheimlich schädlich für das Ansehen von Wissenschaft ist, weil eigentlich möchte man ja, dass Wissenschaft ein unheimlich verlässlicher, gesellschaftlicher und politischer Partner dafür ist, worauf man sich verlassen kann, wenn man Problemlösungsvorschläge macht für irgendwelche gesellschaftlichen Herausforderungen.

Und bei der Klimawissenschaft ist das natürlich noch viel offensichtlicher, weil wir ja auch Sachen, die unfassbar grandios durch beliebig viele Studien auf unterschiedlichste Weise belegt sind, trotzdem die ganze Zeit kritisiert wird.

Und deshalb ist es für mich halt quasi so auf der Meta-Ebene so ein unheimlich wichtiges Thema, dass wir sagen, nee, also diese Meta-Wissenschaft, die wir hier betreiben in diesen Konsortien und die in der NFDI betrieben wird, die ist unheimlich wichtig dafür, dass der Wissenschaftler als Institution vertraut wird, weil wir Prozesse offen und nachprüfbar machen und es eben keine versteckte Sache mehr ist, wie irgendwas zustande gekommen ist und wie irgendwas verwendbar ist.

Und das hoffentlich viel dazu beiträgt, dass man Sachen, die von unglaublicher Tragweite für unsere Zukunft sind, wie jetzt eben die Frage, wie gehen wir mit der Klimakrise um und welche wirklichen Lösungen stehen uns zur Verfügung, damit umzugehen, damit eben auch wirklich effizient und auf wissenschaftlicher

Basis zu streiten und nicht auf Ablehnung von Wissenschaft.

Ich hatte ja zwischendurch schon Angst, dass uns alle Copyköche und alle Wissenschaftler lünnen, aber du hast jetzt doch nachgewiesen, dass diese Schaffung von Freiräumen für Kreativität tatsächlich da ist und das halte ich auch wirklich für ganz wichtig und würde ich auch völlig unterstützen.

Wir reden da auch immer von Kulturwechselln an verschiedensten Stellen, die Wandel, die auch in der Wissenschaft stattfinden, genauso wie in der Öffentlichkeit, dass wir zum Beispiel Bildern glauben, ohne dass wir nachprüfen können, was sie dann eigentlich aussagen, wäre auch ein schönes Beispiel dafür, dass diese Nachvollziehbarkeit von Analysen und schlicht und einfach auch von der Datendarstellung dem Vertrauen in die Wissenschaft unglaublich viel helfen würde.

Also solche, ich weiß nicht, was heißt Colorplots, also solche Bilder, die zweidimensional sind und in dritter Dimension halt den Datenwert, die Intensität, die Zahl von irgendwas darstellt, die sind ja zum Beispiel beliebig manipulierbar und können auch sozusagen Eindrücke vermitteln, ohne dass da irgendwas wirklich an den Daten gefälscht ist.

Und wenn ich solche Dinge nachvollziehbar mache zum einen und die Möglichkeit gebe, selber mit den Farben zu spielen und zu gucken, was steckt da drin und was kommt raus und was ist wirklich da, dann erhöht das das Vertrauen in die wissenschaftliche Arbeit enorm.

Und auf der anderen Seite die Freiräume, die Kreativität, ich denke, wir werden auch wieder dahin kommen, wirklich nachzugucken, was braucht es denn eigentlich an Darstellung?

Wofür muss ich dann die Messung machen?

Sind wirklich immer die bunten Bilder und die schönen Bilder es wert, um um Experimentierzeit zu investieren?

Oder wäre es nicht viel wichtiger, mit intelligenten Methoden die richtigen Parameter für irgendwelche Kräfte, die wirken, für irgendwelche chemischen Prozesse rauszufinden und am Ende die Darstellung vielleicht in einer anderen Art und Weise zu machen, die vertrauenswürdig oder vielleicht vertrauenswürdiger ist anstelle von bildlichen Darstellungen, die vielleicht doch manchmal Eindrücke vermitteln, aber nicht wirklich wissen.

Was sagen denn eigentlich eure ausländischen Kolleginnen und Kollegen dazu, wenn ihr davon erzählt, Astrid?

Teilweise waren die einfach nur begeistert, dass Deutschland anfängt und haben sich das für sich auch gewünscht.

Was?

Wir waren die Ersten?

Soweit ich weiß, waren wir zumindest in Europa die Ersten.

Das hätte ich jetzt nicht gerechnet.

Okay.

Inzwischen haben sich natürlich die verschiedenen Länder und Communities das auch angeguckt und starten ihre eigenen Lösungen.

Aber tatsächlich gab es am Anfang doch sowas wie Bewunderung, dass man dieses große Projekt auch in dieser Größe so angeht, aber eben auch den Wunsch, Ähnliches vielleicht selber zu machen.

Das ist der eine Punkt.

Und der andere ist natürlich, dass in diesen Kooperationen, von denen Philipp sprach, die es über Jahre gibt und in der Großforschung, wo wir beide sind, eben selbstverständlich, man die Investitionen nicht nur für ein einzelnes Land hat und

dass man da extrem kooperiert und damit auch schon lange Standards hat.

Wir in unserem Fall haben die Kollegen eingebunden.

Schon in dem Proposal praktisch sind die Kollegen von den europäischen Partnern involviert worden.

Das heißt jetzt nicht, dass jetzt jeder alles das macht, was wir vorschlagen.

Aber alle wichtigen Entscheidungen werden auch in einer Art von Absprache oder zumindest Information getätigt.

Also entweder man spricht sich ab oder man versucht, sich anpassbar zu machen.

Das Konvertieren, von dem ich vorhin sprach, dass man sich also auf Standards einigt, die dann entsprechend von verschiedenen Systemen lesbar sind oder übertragbar sind.

Und das funktioniert eigentlich sehr gut.

Und diese Zusammenarbeit ist dadurch in keiner Weise beschädigt worden, sondern vielleicht auch sogar enger geworden.

Und das ist sicher bei euch auch so, Philipp?

Nee, auf jeden Fall.

Also man hängt einfach an vielen Sachen, die man natürlich als Teile von großen internationalen Kollaborationen mit beeinflussen, aber nicht alleine bestimmen oder umwerfen kann.

Also zum Beispiel jetzt, um auf das Thema von ganz von Anfang zurückzukommen, in welcher Form sind unsere Daten jetzt wirklich, auf welchem Abstraktionsniveau und wann für alle frei zugänglich.

Das ist natürlich was, was die Kollaboration als Ganze entschließen muss.

Aber wie dann damit umgegangen wird, da können wir natürlich auch einfach sagen, okay, wir haben da Ideen, die dann einbringbar sind.

Und wie wir es, glaube ich, jetzt die ganze Zeit schon gesagt hatten, nicht vieles von dem, was in den Konsortien läuft, sind Entwicklungen, wo jetzt vielleicht eh noch nicht hundertprozentig klar ist, wie das verstetigt werden kann.

Aber die Entwicklung zu haben, zu gucken, wie kann das überhaupt sowieso internationalen wissenschaftlichen Prozess, wie sich unsere föderierte Dateninfrastruktur, die wir eh haben, weiterentwickelt und wie das dann eben auch mit anderen Wissenschaftsfeldern verbunden werden kann, das kann von guten Ideen beeinflusst werden.

Und so wirkt es dann eben, auch wenn vielleicht erst mal noch gar nicht klar ist, in welcher Form dann irgendein Endprodukt davon irgendwann mal, also hatten wir sowieso mal, Ende gibt es eh nicht, aber ein überhaupt mal weitläufig verwendetes Produkt dann nachher überhaupt mal entstehen wird.

Aber wichtig ist eben wirklich, diesen Prozess in Gang zu setzen und den eben auch in der gesamten Community im Gang zu halten.

Aber es ist natürlich auch so, dass wir Randbedingungen gesetzt bekommen.

Die ESRF, also der europäische Synchrotron in Grenoble, hat sofort gesagt, wir haben jetzt investiert über zwei Jahre, auch richtig viel Geld und Manpower, wir ändern das jetzt nicht, weil ihr wenn ihr was anderes macht.

Ja, also das ist auch da, dass man dann Restriktionen hat in dieser europäischen oder internationalen Zusammenarbeit, die man dann bei den entsprechenden Arbeiten und Produkten entsprechend einbindet.

Ja, aber ihr dürft ja flexibel genug sein, das dann trotzdem ordentlich verdaten

zu können, in eurem Sinne, oder?

Genau, das ist das, was ich sage.

Wir machen dann die Produkte, also die Datenformate lesbar von den verschiedenen Systemen.

Da, wo eben schon welche vorhanden sind, wird das einbezogen.

Man kann ja Daten konvertieren, wenn man sich auf Metadaten festgelegt hat.

Und der Input und der Output kann ja abgestimmt sein, oder die Ausgabe und die Eingabeformate können ja abgestimmt sein und dann entsprechend umgewandelt werden für die entsprechenden Kataloge.

Das ist ja technisch möglich, wenn man eben weiß, was dort hinterlegt ist.

Das ist bei Industrieprodukten ganz anders.

Wir haben natürlich in der Physik an vielen Laboren Industriegeräte, wo es sozusagen das Firmengeheimnis oder das Firmenwissen ist, wie Daten erzeugt werden, in welcher Form die abgelegt und verarbeitet werden.

Und dann wird die entsprechende Software mitverkauft, die dann entsprechend die Bilder produziert oder die Datenbearbeitung vornimmt.

Das ist in unseren europäischen Kollaborationen überhaupt nicht der Fall.

Das ist alles offen und man kann damit arbeiten und sich abstimmen.

Also auch da, wo sowas jetzt quasi ungeschickter Weise der Fall ist, weil manchmal lässt sich es nicht vermeiden, also in meinem Bereich gibt es das auch nicht, wir müssen es dann selber entwickeln.

Aber wo es das gibt, ist natürlich in gewisser Weise das Gerät, seine

Dokumentation und auch sein unbekannter interner Zustand auch Teil der Dokumentation.

Selbst wenn es eine Blackbox ist.

Das ist natürlich ungeschickt, aber auch das muss man realistisch vorsehen, dass sowas passieren kann.

Naja, aber vielleicht kann ja dann in einem über oder über, übernächsten Schritt es vielleicht irgendwelche Vorgaben geben, dass Hersteller nur noch Geräte liefern dürfen, wenn sie sauber dokumentiert sind.

Kann man sich ja auch vorstellen.

Ja, ich fände sowas sehr gut, muss ich sagen.

Also es ist immer problematisch.

Das kann auch bei Software natürlich passieren, nicht nur bei Geräten.

Und ja, ich würde jetzt wieder sagen, da habe ich jetzt ein bisschen Glück.

Also die Teilchenphysik arbeitet eben mit so wenigen großen Experimenten typischerweise, dass wir wirklich jeden Sensor im Wesentlichen als Community, also nicht jeder Wissenschaftler oder Wissenschaftlerin, aber als Community jeden Sensor selber bauen und die gesamte Auslese- und Analyseinfrastruktur.

Aber auch da, also zum einen kann man es ganz konkret sehen, welche Analysewerkzeuge können wir als wissenschaftliche Community.

Und das hängt alles mit dem zusammen, was wir vorher sagten.

Es geht eigentlich, es fängt an bei den Daten, faire Daten, aber es geht eben eigentlich nicht nur um die Daten, sondern es geht um faire Wissenschaft.

Und die ganzen Analysewerkzeuge gehören dazu.

Und wenn ich jetzt quasi sage, naja, bleiben wir jetzt erstmal bei näher liegenden Sachen.

Bestimmte Algorithmen, um bestimmte Sachen zu machen, mögen in der Wissenschaft entwickelt worden sein.

Aber wenn ich jetzt eine Analyse mache, dann werde ich in ganz vielen Feldern, wenn es eine komplexe statistische Analyse ist, quasi Werkzeuge verwende, die in dem TensorFlow Package drin sind.

TensorFlow ist ein Paket, was von Google ursprünglich mal auf den Markt gebracht wurde, was glücklicherweise offen ist.

Also es ist keine Blackbox, jeder kann nachgucken, aber es ist trotzdem eben was, was eben ganz zentral von Großkonzernen mit verwaltet wird und was vielleicht inzwischen genügend Community hat, dass es auch weiter existieren würde, aber muss ja nicht.

Und da ist man eben schon relativ nah an einem Beispiel dran, wo man eigentlich mit einer Blackbox arbeitet.

Und das ist eigentlich schade, weil jetzt, um in dem Beispiel zu bleiben, wenn ich jetzt ganz viele Analysen, die sehr viele statistische Parameter mit berücksichtigen, verwenden, heutzutage eben sogenannte Hamiltonian-Monte Carlo-Modelle oder ganz allgemein von Politiker Markov, Chain-Monte Carlo, was ein bisschen allgemeinerer Begriff ist.

Und damit kann man typischerweise mit bayerischer Statistik, verwendet man dann, mit unglaublich großen Räumen umgehen.

Also sagen würde, ich habe jetzt nicht nur, wenn ich mich mit Raum bewege, drei Dimensionen, ich kann nach oben, unten, links, rechts, vorne, hinten, sondern ich habe an jedem Punkt eine Farbe, an jedem Punkt eine Temperatur, dann habe ich

schon fünf Dimensionen, ich habe an jedem Punkt noch eine Richtung und eine Stärke von einem elektrischen Feld, dann habe ich noch mal vier Dimensionen mehr und so weiter.

Man kann sich also beliebig jetzt quasi räumlich vorstellen, man kann sich einen beliebig dimensionalen Raum vorstellen und in so einem Raum muss ich jetzt quasi die optimale Lösung von irgendeinem statistischen Problem finden, also quasi eine Approximation dessen, was wahrscheinlich wirklich mal der Fall war.

Und ganz viele dieser Verfahren, also alle diese Verfahren kommen aus der Wissenschaft und viele dieser Verfahren sogar aus der Physik.

Also das, was jetzt zum Beispiel bei diesen epidemiologischen Analysen wiederum viel verwendet wird, auch bei denen, die wir machen, ist eben Hamiltonian-Monte Carlo, was eben aus der Lattice QCD, also einem Teilgebiet der Teilchen- und Hadronphysik kommt.

Und das Ganze wird dann verpackt quasi in irgendeinem Werkzeug in TensorFlow, was eben von Google sozusagen mal veröffentlicht wurde.

Und das ist eigentlich schade, weil ganz viele Sachen dann quasi schon mal hinter so einer einfachen Benutzeroberfläche verschwinden und gar nicht mehr direkt mit dokumentiert werden.

Und es wäre eigentlich besser, dafür öffentliche Sachen zu haben.

Und wenn man das Ganze jetzt weiterdenkt und das jetzt mehr, was kann ich jetzt nicht für die NFTI sprechen, aber jetzt also einfach so für die Zukunft.

Man redet ja viel davon sozusagen, wie allgemeine AI oder diese Large Language Models, wie die unsere Zukunft beeinflussen werden, ob die eine direkte Gefahr sind.

Glaube ich persönlich erstmal nicht.

Viele Leute reden ja von so einem Extinction Risk.

Aber dass diese Sachen ein unheimliches Missbrauchspotenzial haben, glaube ich, wissen wir alle.

Und dass selbst wenn sie nicht missbraucht werden, unsere Kommunikation und unseren Umgang und auch dessen, was wir eigentlich glauben.

Also wiederum so ein bisschen wie das, was Astrid vorher sagte mit diesen Farbplots.

Also ich gucke auf den Plot, kriege einen allgemeinen Eindruck und nehme den erstmal so hin.

Und so ist es ja dann auch, dass wenn ich jetzt so ein Large Language Model einfach mal als Assistent für alles verwende, wenn ich für all meine Alltagsfragen quasi einfach "Jet GPT-4" oder "Bing" oder wie heißen diese Dinge alle, frage und dann einfach glaube, dann entwickelt sich da in einer Black Box, die von niemandem nachvollziehbar ist, quasi ein Automatismus, weil genau die Interaktion mit diesen Modellen gibt bisher wieder Inputdaten in das weitere Training dieser Modelle.

Und das Ganze geben wir quasi im Moment komplett aus der Hand.

Und das ist jetzt nichts, was die NFTI direkt angehen kann, aber was ich sehr wichtig finde, einfach als politisch denkender Mensch oder Wissenschaftler eben auf dem Gebiet des Umgangs mit Daten, dass so ein Potenzial hat, unsere Kommunikation, unsere Art und Weise, quasi die Layer zwischen uns und der Welt wahrzunehmen.

Also ganz wenig dessen, wie wir den aktuellen Zustand der Welt wahrnehmen, kommt daher, dass wir irgendwie nach draußen gehen und ein Blatt am Baum anfassen.

Sondern ganz viel kommt daher, wie wir eben mit interaktiven Medien mit der

Welt kommunizieren.

Wir haben ja immer einen Vermittler dazwischen.

Und idealerweise können wir gucken, wie der Vermittler die Informationen verarbeitet hat.

Also idealerweise ist es ein Journalist, wo wir gucken können, wie funktioniert eigentlich dessen Handwerk.

Und dann nehmen wir das, was er als Rohdaten hat, werfen das in dieses Handwerk und gucken, ob das stimmt.

Genau, das ist ein wunderbares Beispiel.

Genau dieser Prozess, der wird ja quasi in einem Prozess, der quasi für liberale Demokratien sehr schmerzhaft zu sein scheint, hat man gerade den Eindruck, der wird ja dem Journalismus sehr weit aus der Hand genommen im Moment.

Also in irgendeiner Weise community-dominierendem Prozess.

Und der wird eben aus der Hand genommen durch Sachen, die in ihrer Datenbasis, wie Algorithmen auf Daten der Nutzer funktionieren, für uns und insbesondere in Europa überhaupt nicht nachvollziehbar ist.

Der ist in ChatGPT4 nicht nachvollziehbar und der ist aber auch vorher auf Facebook und Twitter und was weiß ich nicht direkt nachvollziehbar.

Und dass uns diese Sachen aus der Hand genommen sind, glaube ich, jetzt einfach noch mal als Bürger gesprochen sozusagen, aber mit dem Hintergrund der NFDI und der Datenverarbeitung.

Das ist was, was wir eigentlich zum Schutz unserer Demokratie wieder in den Griff bekommen sollten.

In dem Sinne, dass wir offene, nachvollziehbare Strukturen haben, wie unsere eigenen Daten in nützliche AI-Produkte verarbeitet werden.

Weil das so einen persönlichen Assistenten zu haben, also man jetzt quasi den persönlichen Ballot hätte, der oder Butler oder was auch immer, oder Butlerin, der einem quasi immer alle Sachen erzählt, die man gerade braucht, wäre ja super nützlich.

Hat ein unfassbares, indirektes Missbrauchspotenzial.

Und deshalb Produkte zu haben, die das quasi nachvollziehbar machen, was dabei gemacht wird, wäre, glaube ich, was, was für unsere Demokratie sich als unglaublichen Schutz herausstellen könnte.

Da gibt es, soweit ich weiß, im Moment überhaupt keinen Ansatz.

Aber für mich wäre das eben was, was ich einfach logisch aus dem Gedanken der NFDI als Weiterentwicklung.

Vieles von dem, was da aufgebaut wird, wäre eine notwendige Grundlage dafür.

Und ich glaube einfach, wenn man sich die Zeitstrukturen anguckt, müssten wir mit radikaler Geschwindigkeit sofort in die Richtung weitergehen, die ich gerade angerissen habe.

Aber das ist ein politisches Thema und kein direkter, die Physik oder andere einzelne wissenschaftliche Themen betreffendes Thema.

Aber ich glaube, da sieht man, wie gesellschaftlich wichtig es ist, wie wir mit Daten, wie wir mit Wissenschaft, wie wir mit Prozessen auf Daten umgehen.

Und genau darum dreht sich dann eben doch die NFDI.

Astrid, hat Philipp da gerade ein weiteres Konsortium erfunden?

Das kann ich jetzt nicht versprechen, weil die Struktur ja eigentlich abgeschlossen ist.

Aber ich würde das tatsächlich als Punkt aufnehmen.

Und ich behalte solche Dinge immer gerne im Hinterkopf und hole die dann bei Gelegenheit raus.

Also vielleicht das eine ist, dass man tatsächlich diese Diskussion, genauso wie Philipp das jetzt getan hat, noch viel stärker führen sollte, als wir das bisher tun in allen Dingen, die kulturellen Wandel bisher schon berücksichtigen oder den eben auf die Tagesordnung bringen.

Und das andere ist, warum nicht?

Vielleicht sollten wir eine Sektion aufmachen.

Das ist kostenneutral, ist das politisch richtige Wort dafür.

Das heißt, es kostet die DFG und das BMBF kein Geld, aber tatsächlich über eine Sektion nachzudenken, die diese Aspekte mit aufgreift.

Jedes Konsortium hat irgendwelche Leute, die sich mit Dissemination Outreach, PR und ich weiß nicht was beschäftigen.

Das wäre sogar eine coole Idee.

Lass uns drüber schlafen.

Ich glaube, es ist wirklich auch eine Frage der Größenordnung.

Ich glaube, jeder oder viele von den Konsortien beschäftigt sich natürlich mit Werkzeugen, die Machine Learning machen.

Ich hatte eben schon diese Optimizer wie Hamiltonian Monte Carlo

angesprochen, die man eben natürlich auch in dem Kontext braucht.

Natürlich machen wir für bestimmte konkrete Fragestellungen Machine Learning.

Aber so ein Large Language Model wie jetzt eben JetGPT4, das ist eben im Moment der Wissenschaft, insbesondere glaube ich der europäischen Wissenschaft, aber auch der amerikanischen Wissenschaft.

Jetzt gießt doch nicht Wasser in den Wein.

Nein, nein, nicht Wasser in den Wein, sondern umso mehr.

Ich sage, im Moment ist es der Wissenschaft komplett aus der Hand genommen.

Und eben drum wäre es wichtig, in so einer Sektion, wie Astrid das sagte, oder in einem anderen Rahmen eben darüber zu reden, wie man das auf einem relevanten Rahmen angehen kann.

Und überhaupt weiter hinzukommen, das habe ich jetzt voll unterstützt, nicht Wasser in den Wein gießen, sondern den Wein verköstigen, damit es nicht so ein Problem bleibt, nicht eben in der Richtung weiterzugehen.

Da gibt es eben auch natürlich viele andere Leute außerhalb unserer Konsortien, die da natürlich viele gute Ideen haben.

Und wir sind ja noch verheiratet in einer anderen Initiative, die sich ERUM Data nennt und uns sozusagen in der BMBF-Förderung verheiratet.

Und da geht es natürlich eigentlich um die gleichen und ähnliche Themen.

Und auch da kann man darüber nachdenken, ob man das Thema zumindest jetzt auf der, nicht auf der Tool-Entwicklungsseite, aber auf der Seite der Diskussion über die Themen und der Ansprache von jungen Wissenschaftlern oder das Thema sozusagen in der Breite zu diskutieren, in einer Korrelation zu demokratischen Prozessen und unserer Gesellschaft als Ganzer.

Auch dort könnte man kostenneutral dieses Thema mit berücksichtigen.

Also wir haben uns dort jetzt mit Sustainability beschäftigt, mit Nachhaltigkeit.

Auf der gleichen Dimension wären die Dinge, die du da jetzt angesprochen hast.

Astrid Schneidewind und Philipp Bechtele, vielen Dank.

Danke.

[Musik] [Abspann]