

Big Data, Algorithmen und AI:

DATENGETRIEBENE FORSCHUNG IN DER WISSENSCHAFT 2030

Mit welchen neuen Herausforderungen sind Hochschulen und außerhochschulische Forschungseinrichtungen in Österreich in Hinblick auf datengetriebene Forschung konfrontiert?

Studie im Auftrag des Rats für Forschung und Technologieentwicklung
erarbeitet von winnovation – Open Innovation Forschung und Beratung



Impressum

Diese Studie wurde im Auftrag des Rates für Forschung und Technologieentwicklung (RFTE) verfasst.

Juli 2020



winnovation consulting gmbh

Karl-Schweighofer-Gasse 12/6
1070 Wien
office@winnovation.at

VerfasserInnen: Dr. Gertraud Leimüller, Stefan Benke, PhD, Lena Müller-Kress, MA (winnovation) in enger Zusammenarbeit mit Dr. Anton Graschopf (RFTE)

Deckblatt: *Photo by guilherme-stecanella on Unsplash*

Inhaltsverzeichnis

Impressum	4
Inhaltsverzeichnis	5
Executive Summary	9
Glossar: Wichtige Begriffe dieser Studie	15
Abkürzungsverzeichnis	18
1. Einleitung	20
1.1..Definition und Bedeutung datengetriebener Wissenschaft	21
1.2..Zunehmende politische Bedeutung von datengetriebener Wissenschaft international und national	22
1.3..Zielsetzungen der Studie	24
2. Methode und Beteiligte	25
3. Internationale Trends datengetriebener Wissenschaft und Status quo in Österreich	29
3.1..Datenzugänge - Daten sind Voraussetzung exzellenter Forschung	31
3.2..Datenkompetenzen – Neue Aufgaben und Berufsbilder prägen die Wissenschaft	34
3.3..Interdisziplinarität – Datengetriebene Forschung überschreitet Grenzen	37
3.4..Datenstandards - Einheitliche Vorgehensweisen machen Daten teil- und nutzbar	38
3.5..Dateninfrastrukturen – Technologien ermöglichen Zugänglichkeit und Wertschöpfung	39
3.6..Rechtliche Rahmenbedingungen – Sicherheit und Klarheit bilden die Grundlage datengetriebener Wissenschaft	41
3.7..Artificial Intelligence – Die Arbeit mit Daten verändert die Wissenschaft von Grund auf	43
4. Vision: Datengetriebene Wissenschaft 2030	46
5. Handlungsempfehlungen	47
5.1 Handlungsfeld 1: Nationale Rahmenbedingungen für datengetriebene Wissenschaft	49
Handlungsempfehlung 1.1: Entwicklung und Umsetzung einer nationalen Forschungsdatenstrategie	50
Handlungsempfehlung 1.2: Aufbau einer Kommission für Datenethik und Foresight zu datengetriebener Wissenschaft	51
5.2 Handlungsfeld 2: Incentivierung eines professionellen Datenumgangs	52
Handlungsempfehlung 2. 1: Verankerung eines professionellen Datenumgangs an den Hochschulen	53
Handlungsempfehlung 2.2: Verankerung eines professionellen Datenumgangs in der Forschungsförderung	54
5.3 Handlungsfeld 3: Datenzugänge für die Wissenschaft	55
Handlungsempfehlung 3.1: Gesetzliche Voraussetzungen für die Nutzung von Government Data und Registerdaten in der Wissenschaft schaffen	56
Handlungsempfehlung 3.2: Sichere und einfache Remote-Zugänge zu Government Data und Registerdaten für die Wissenschaft bereitstellen	57
Handlungsempfehlung 3.3: BürgerInnen ermöglichen, ihre Daten freiwillig der Wissenschaft zur Verfügung zu stellen	58

Handlungsempfehlung 3.4: Etablieren von Vertragsbausteinen für eine sichere gemeinsame Datennutzung zwischen Wissenschaft und Wirtschaft.....	59
Handlungsempfehlung 3.5: Aktive Mitgestaltung des europäischen Rechtsrahmens für die Nutzung von Daten durch Österreich.....	60
5.4 Handlungsfeld 4: Datenkompetenz in Forschungseinrichtungen	62
Handlungsempfehlung 4.1: Entwicklung und Institutionalisierung von Datenstrategien in den Forschungseinrichtungen	63
Handlungsempfehlung 4.2: Anstellung (und Förderung) von Data Science ExpertInnen an allen Forschungseinrichtungen	64
Handlungsempfehlung 4.3: Einsatz von Data Stewards und/oder Datamanagementteams an allen Forschungseinrichtungen	65
Handlungsempfehlung 4.4: Angebot eines Supports für Datenrecht an jeder Forschungseinrichtung	66
5.5 Handlungsfeld 5: Ausbildung und Karrieremodelle	67
Handlungsempfehlung 5.1 Verankerung von Datenkompetenzen in allen Studiengängen	68
Handlungsempfehlung 5.2: Aus- und Weiterbildung von Data Scientists und Data Stewards spezifisch für die Wissenschaft	69
Handlungsempfehlung 5.3: Entwicklung von attraktiven unbefristeten Karrieremodellen und Aufgabenbereichen in Data Science, Data Stewardship und Datenrecht	70
Handlungsempfehlung 5.4: Neue Modelle der Leistungsmessung, Bewertung und Rekrutierung von WissenschaftlerInnen	71
5.6 Handlungsfeld 6: Dateninfrastrukturen und europäische Kooperation	72
Handlungsempfehlung 6.1 Verstärkte aktive Teilnahme Österreichs an internationalen Initiativen für Dateninfrastruktur	73
Handlungsempfehlung 6.2 Ausbau österreichischer Dateninfrastrukturen und Services	73
Handlungsempfehlung 6.3 Aufbau von Data Trusts, welche sensitive Daten hosten und deren sichere Nutzung ermöglichen.....	75
6. Resümee	76
Abbildungsverzeichnis	78

Big Data, Algorithmen und AI:

**Datengetriebene Forschung
in der Wissenschaft 2030**

Welche Herausforderungen kommen auf die österreichische Wissenschaft durch datengetriebene Forschung zu?

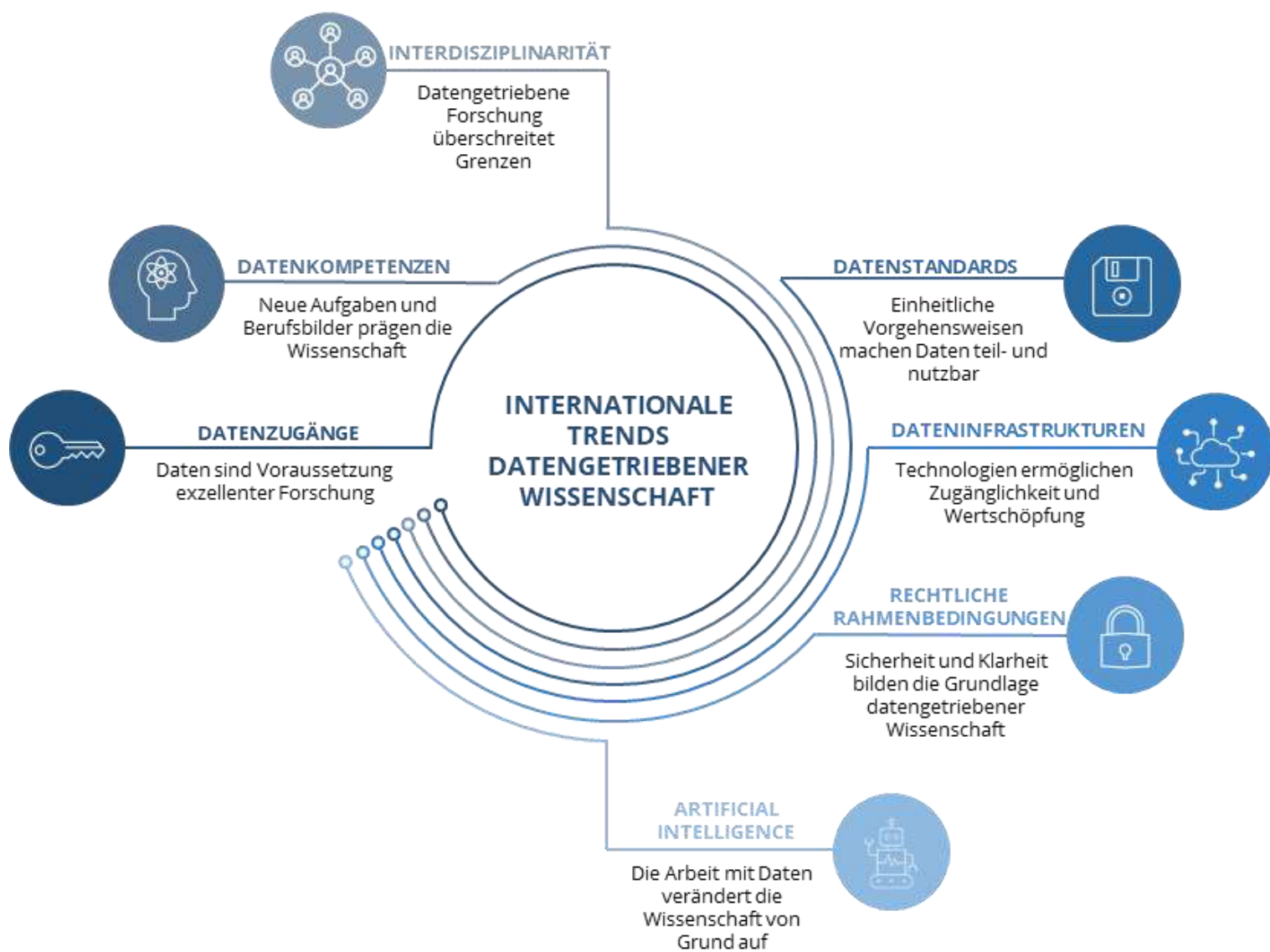
Executive Summary

Durch die fortschreitende Digitalisierung wächst die Menge der Daten täglich an. Algorithmen und Artificial Intelligence (AI) werden in Zukunft immer mehr Bereiche des Lebens durchdringen. Dies hat nicht nur drastische Auswirkungen auf das alltägliche Leben, sondern auch auf die Art und Weise, wie wissenschaftliche Forschung betrieben wird. Disziplinen, Forschungsprozesse und die Forschungsinstitutionen selbst stehen vor oder bereits mitten in einem Transformationsprozess, der sich noch weiter beschleunigen wird. Mehrere europäische Strategien adressieren diese bereits, etwa das geplante EU-Forschungsrahmenprogramm Horizon Europe, die „European Strategy for Data“, der „European Green Deal“ und auf nationaler Ebene die „Austrian ERA Roadmap“ sowie auch das Regierungsprogramm 2020-2024.

Doch **welche Herausforderungen kommen durch die rasant stattfindenden technologischen, organisationalen und inhaltlichen Entwicklungen im Bereich datengetriebener Wissenschaft auf die österreichischen Forschungseinrichtungen zu?** Und was muss in weiterer Folge getan werden, damit das österreichische Wissenschaftssystem, vornehmlich Hochschulen und außerhochschulische Forschungseinrichtungen, auf die Veränderungen, welche die digitale Datenrevolution mit sich bringt, nicht nur vorbereitet ist, sondern auch international wettbewerbsfähig bleibt?

Diese Fragestellungen sind für den Wissenschaftsstandort von hoher Relevanz, wurden jedoch bisher noch kaum konkret beleuchtet. Sie stehen daher im Mittelpunkt der vom Rat für Forschung- und Technologieentwicklung (RFTE) 2020 beauftragten Studie. Unter Anwendung von Open Innovation Methoden, die eine Beteiligung von ExpertInnen und konkret Betroffenen im Wissenschaftsbereich ermöglichten, erarbeitete das Projektteam von winnovation sieben große internationale Trends und Anforderungen datengetriebener Forschung, die in der folgenden Grafik im Überblick gezeigt werden.

Abbildung 1 **Die sieben internationalen Trends datengetriebener Wissenschaft**



20 Handlungsempfehlungen für die Zukunft der datengetriebenen Wissenschaft in Österreich

In weiterer Folge wurden 20 konkrete Handlungsempfehlungen für Politik und Forschungseinrichtungen in Österreich in sechs Handlungsfeldern erarbeitet.

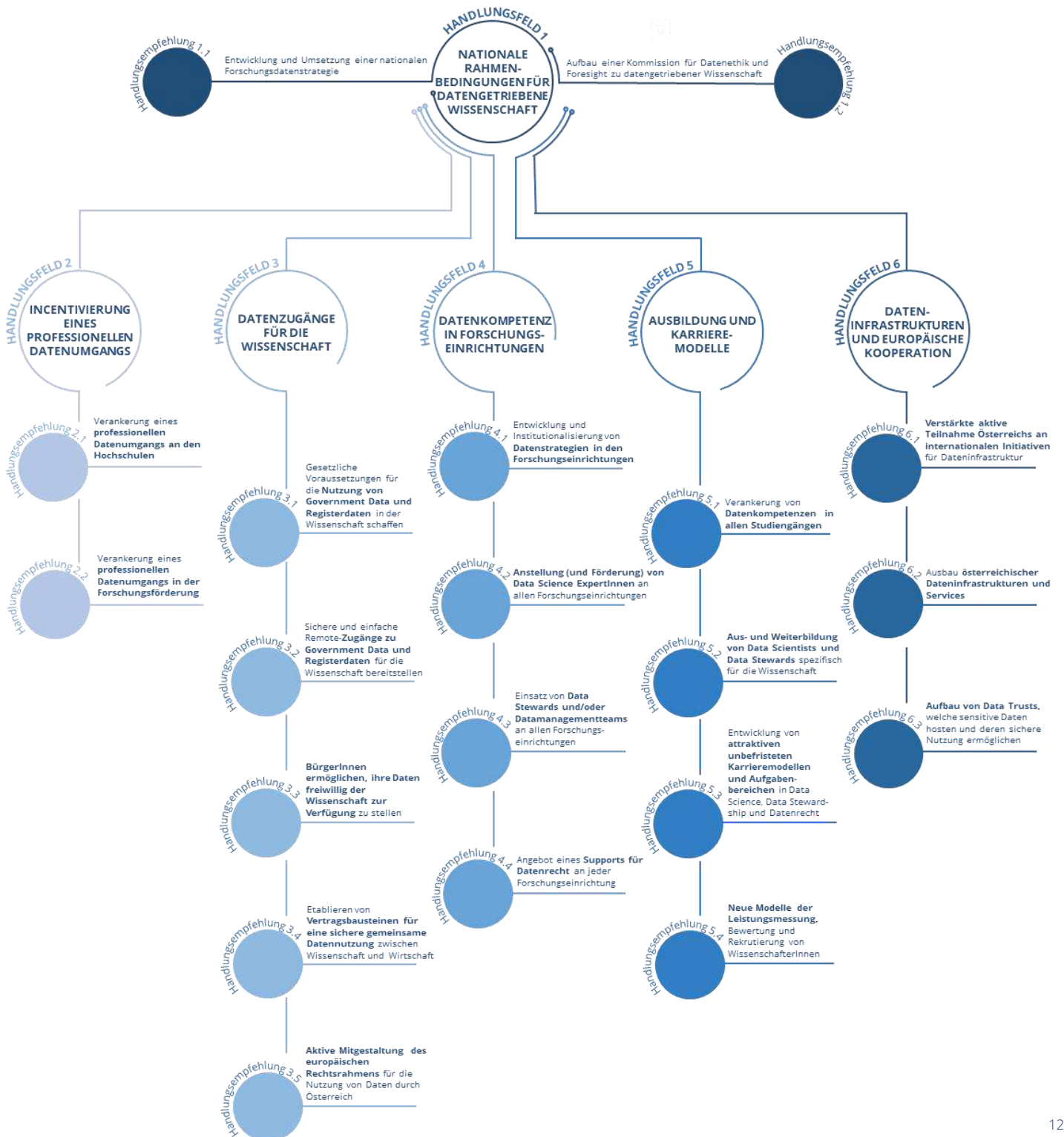
Erste positive Entwicklungen und Initiativen für datengetriebene Wissenschaft sind in Österreich vorhanden. Diese müssen allerdings verstärkt werden, damit Österreich angesichts der dynamischen Entwicklung als Wissenschaftsstandort international konkurrenzfähig bleiben kann.

Zentral ist hierbei die Erarbeitung einer nationalen Strategie zu datengetriebener Wissenschaft und zur Abstimmung der Aktivitäten, um Kräfte zu bündeln. Hierunter fällt auch eine durchgehende Beobachtung der aktuellen und zukünftigen technologischen Entwicklungen. Denn datengetriebene Forschung entwickelt sich in einem sehr schnellen Tempo, insbesondere in Bezug auf Artificial Intelligence (AI) werden große Auswirkungen auf das Wissenschaftssystem und die Art, wie Forschung betrieben wird, erwartet. Um diese Entwicklungen konkret zu adressieren, sind zum einen die Forschungsförderungen ein großer Stellhebel und zum anderen auch die Forschungsinstitutionen selbst. Doch auch die rechtlichen Rahmenbedingungen spielen eine große Rolle in Bezug auf datengetriebene Wissenschaft. Daten müssen und sollen vor Missbrauch geschützt werden, ErstellerInnen entsprechend entlohnt werden und zeitgleich Datenzugänge für die Wissenschaft ermöglicht werden.

Denn Daten sind die Grundlage und Ausgangspunkt für jegliche datengetriebene Forschung. Um mit diesen Daten arbeiten zu können, müssen in weiterer Folge auch die Kompetenzen an den Institutionen vorhanden sein. Dies kann einerseits durch den Ausbau von wissenschaftlichen Stellen und Entwicklung neuer, zukünftiger Karrierebilder ermöglicht werden und andererseits durch Services, die Beratung und Unterstützung bieten. Hierfür müssen dringend die richtigen SpezialistInnen ausgebildet werden und langfristig auch die Kompetenzlevel aller WissenschaftlerInnen gehoben werden. Damit diese datengetriebene Wissenschaft betreiben können, sind die richtigen Technologien von Nöten, leistungsstarke Datenverarbeitung und insbesondere Zugänge zu nationalen und internationalen Datenspeichern und Repositorien.

Damit diese Voraussetzungen geschaffen werden können, folgen nun die erarbeiteten Handlungsempfehlungen, welche konkret durch österreichische Forschungspolitik adressiert werden sollen.

Abbildung 2 Die Handlungsfelder und -empfehlungen für datengetriebene Wissenschaft



Handlungsfeld 1: Nationale Rahmenbedingungen für datengetriebene Wissenschaft

Auf Bundesebene soll eine **nationale Forschungsdatenstrategie** unter Federführung des Bundesministeriums für Bildung, Wissenschaft und Forschung (BMBWF) und unter Einbeziehung aller relevanter StakeholderInnen entwickelt und umgesetzt werden. Diese soll gemeinsame nationale Ziele im Umgang mit Daten definieren, prioritäre Kooperationsbereiche identifizieren und Ressourcen zwischen Forschungseinrichtungen / Universitäten bündeln.

Durch eine **Kommission für Datenethik und Foresight zu datengetriebener Wissenschaft** soll ein laufendes Monitoring der raschen technologischen Weiterentwicklungen und des Einsatzes von Daten in Wissenschaft und Forschung durchgeführt werden. Aus diesen Beobachtungen heraus sollen dann in Folge konkrete Empfehlungen für die österreichische Politik und Forschungslandschaft entstehen.

Handlungsfeld 2: Incentivierung eines professionellen Datenumgangs

Zur **Verankerung des professionellen Umgangs mit Forschungsdaten an den Hochschulen** sollen innovative Metriken zum Umgang mit Forschungsdaten auf institutioneller Ebene, als integrativer Bestandteil von Leistungsvereinbarungen von Hochschulen und Forschungsinstituten und auf individueller Ebene als zusätzliche Bewertungskriterien in Karrierevereinbarungen bzw. Berufungsverfahren an Hochschulen implementiert werden.

Der **professionelle Umgang mit Forschungsdaten** soll vermehrt **in der Forschungsförderung** gefördert und gefordert werden. Hierfür soll die Forschungsförderung in Hinblick auf einen professionellen Datenumgang weiterentwickelt und angepasst werden, um Anreizstrukturen in der Forschungsförderung zu verankern.

Handlungsfeld 3: Datenzugänge für die Wissenschaft

Um Datenzugänge für die Wissenschaft zu ermöglichen, sollen die **gesetzlichen Voraussetzungen für die Nutzung von Government Data und Registerdaten für die österreichische Wissenschaft geschaffen** werden. Hierfür sollen nationalstaatliche rechtliche Barrieren für die Nutzung unterschiedlicher Daten der öffentlichen Hand beseitigt werden. Weiters sollen rechtliche Standards für Zugänge zu Daten der öffentlichen Hand erstellt werden und die EU-Richtlinie zu offenen Daten und der Weiterverwendung von Informationen des öffentlichen Sektors (PSI) in Österreich möglichst breit umgesetzt werden.

Forschung. Deshalb soll eine **praktische und einfache Möglichkeit für BürgerInnen erarbeitet werden, der Wissenschaft Daten freiwillig zur Verfügung zu stellen** und die Erlaubnis zur Nutzung persönlicher Daten zu erteilen. Weiters sollen Prozesse und eine Plattform aufgebaut werden, um diese Daten für die Wissenschaft zugänglich zu machen.

Zudem sollen **Government Data und Registerdaten für die Wissenschaft über sichere und einfache Remote-Zugänge bereitgestellt** werden. Deshalb soll ein Datenzentrum für Mikrodaten und eine entsprechende Rechtsgrundlage geschaffen werden, umgesetzt durch das im Regierungsprogramm vorgesehene „Austrian Micro Data Center“.

Um den Prozess sehr aufwändiger Vertragsverhandlungen zwischen Unternehmen und Wissenschaft für den Zugang zu Daten zu beschleunigen und zu vereinfachen, sollen **Vertragsbausteine für einen sicheren Datenaustausch zwischen Wissenschaft und Wirtschaft etabliert** und ein entsprechender leicht verständlicher Entscheidungsbaum erarbeitet werden.

Persönliche Daten sind in Europa umfassend durch die Datenschutzgrundverordnung (DSGVO) geschützt. Dennoch sind sie eine potenziell immens wertvolle Datenquelle für die

Der **europäische Rechtsrahmen für die Nutzung von Daten soll aktiver durch die österreichische Forschungspolitik mitgestaltet** werden. Unter anderem soll dadurch eine EU-weite Entwicklung und Umsetzung von standardisierten Machine-to-Machine fähigen Ownership- und Lizenzmodellen für Daten aus unterschiedlichen Disziplinen und Kooperationsformen ermöglicht werden.

Handlungsfeld 4: Datenkompetenz in Forschungseinrichtungen

Forschungseinrichtungen sollen institutionelle Datenstrategien für den Umgang mit Daten entwickeln und etablieren. Zudem sollen sie klare Zielsetzungen, Maßnahmen und Werkzeuge etablieren, um WissenschaftlerInnen bei Aktivitäten im Bereich der datengetriebenen Wissenschaft zu unterstützen.

Scientists sollen je nach Bedarf in den Forschungsgruppen direkt integriert werden oder institutionsübergreifend zur Verfügung stehen und als Servicestelle für alle Fragen zur Anwendung von datengetriebener Forschung dienen. Hierfür sollen sowohl neue Stellenangebote geschaffen als auch entsprechende Personen explizit von den Institutionen hierfür ausgebildet werden.

An allen Forschungseinrichtungen sollen Data Science ExpertInnen angestellt und gefördert werden. Diese Data

Um die kontinuierliche Datenpflege und -haltung an den Forschungseinrichtungen und die Datenmanagement-kompetenz zu gewährleisten, sollen **an allen Forschungseinrichtungen Data Stewards und/oder Datenmanagementteams eingesetzt** werden, welche mit den entsprechenden Ressourcen und Verantwortlichkeiten ausgestattet sein sollen, um WissenschaftlerInnen ideal beim professionellen Forschungsdatenmanagement unterstützen zu können.

Handlungsfeld 5: Ausbildung und Karrieremodelle

Es soll eine flächendeckende Grundkenntnisvermittlung **von Datenkompetenzen in allen Studiengängen verankert** werden und der richtige Umgang mit Daten in allen wissenschaftlichen Disziplinen bereits in der Ausbildung (curricular und extra-curricular) gelehrt werden.

An den Hochschulen, aber auch in anderen Einrichtungen sollen verstärkt **Aus- und Weiterbildungen im Bereich der Data Science und Data Stewardship spezifisch für die Wissenschaft angeboten** und kommuniziert werden. Hierfür sollen neue Aus- und Weiterbildungen für WissenschaftlerInnen als auch QuereinsteigerInnen entwickelt werden, um den wachsenden Bedarf an Fachkräften zu decken.

Um Data Scientists, Data Stewards und DatenrechtsexpertInnen in den Forschungsinstitutionen einzusetzen und langfristig an den Forschungsinstitutionen zu halten, sollen **attraktive,**

Es sollen an den Forschungseinrichtungen **Supports für Datenrecht zu datengetriebener Wissenschaft angeboten** werden. Diese sollen als leicht zugängliche Anlaufstelle dienen, wenn im Rahmen von datengetriebener Forschung rechtliche Fragen und Unsicherheiten existieren (etwa zu Datennutzung, -speicherung, -weiternutzung aber auch Lizenzierung).

unbefristete Karrieremodelle und Aufgabenbereiche in Data Science, Data Stewardship und Datenrecht an Forschungseinrichtungen entwickelt werden.

Datengetriebene Forschung fordert neue Arbeitsweisen an den Institutionen und damit auch eine Vielzahl neuer Kompetenzen. Dafür wird auch ein zeitgemäßes Anreizsystem benötigt. Daher sollen **neue Modelle der Leistungsmessung, Bewertung und Rekrutierung von WissenschaftlerInnen** erarbeitet werden. WissenschaftlerInnen sollen anhand neuer Metriken bewertet werden und entsprechend ihrer individuellen Fähigkeiten interdisziplinär in Teams zusammenarbeiten, um einen Datenfluss über die Disziplinengrenzen hinweg zu erleichtern.

Handlungsfeld 6: Dateninfrastrukturen und europäische Kooperation

Der **österreichische Forschungsstandort soll aktiver an internationalen Initiativen für Dateninfrastruktur teilnehmen**, diese mitgestalten und wann immer möglich diese anstelle von nationalen Lösungen nutzen, und in weiterer Folge auch mitfinanzieren.

Abgestimmt mit der nationalen Forschungsdatenstrategie, soll Österreich eigene **Dateninfrastrukturen und Services** dort **ausbauen**, wo die europäischen Angebote nicht ausreichen, oder wo sie als resiliente Ausfallstrukturen dienen können.

Um sichere Bedingungen für die Nutzbarmachung sensibler Daten für die Forschung (z.B. aus der Wirtschaft) zu schaffen, sollen **Data Trusts, welche sensitive Daten hosten und deren sichere Nutzung ermöglichen, aufgebaut** werden. Diese sollen als vertrauenswürdige Intermediäre bei der Nutzbarmachung von Daten zwischen Vertragsparteien agieren, Daten hosten und die vertragskonforme Nutzung überwachen.

Resümee

Datengetriebene Forschung bietet ein nie dagewesenes Potenzial zur Adressierung der großen gesellschaftlichen Probleme unserer Zeit, und transformiert drastisch die globale Wissenschaft. In Österreich wurden bereits erste Initiativen gesetzt, um diese Potenziale nutzen zu können, diese müssen jedoch konsequent ausgebaut werden, um exzellente datengetriebene Wissenschaft in Österreich langfristig sicherzustellen. Erfolgsentscheidend ist hier ein konzertiertes und koordiniertes

Vorgehen auf allen Ebenen, um optimale Synergien zu schaffen und den Wissenschaftsstandort signifikant zu stärken. Werden die hier vorgeschlagenen Handlungsempfehlungen konsequent umgesetzt, dann finden WissenschaftlerInnen in Österreich optimale Bedingungen für datengetriebene Forschung vor, wodurch diese entscheidend zur Lösung gesellschaftlicher Herausforderungen beitragen können, global wie auch in Österreich.

Glossar: Wichtige Begriffe dieser Studie

Anonymisierte Daten	Daten ohne Personenbezug, sie lassen sich nicht auf eine (in ihrer Identität bestimmte) Person zurückführen ¹ .
Artificial Intelligence	Künstliche Intelligenz. Beschäftigt sich mit Methoden zur Automatisierung von intelligentem Verhalten, insbesondere durch Computer. Artificial Intelligence erlaubt es Computern Aufgaben durchzuführen, für die Menschen Intelligenz aufwenden müssten. ² Hierunter kann unter anderem Machine Learning, wie auch Deep Learning fallen.
Big Data	Bezeichnet Daten, die aufgrund ihrer enormen Menge, Komplexität oder geringen Strukturiertheit nicht mit herkömmlichen Methoden der Datenverarbeitung bearbeitet werden können ³ .
Datengetriebene Forschung	Extraktion von Wissen aus großen Datenmengen für wissenschaftliche Zwecke. Dabei werden Techniken und Methoden aus der Mathematik, Statistik, Stochastik und Informatik angewandt. ⁴ Daten, die in datengetriebener Forschung verwendet werden, können aus verschiedensten Quellen stammen. Vielfach handelt es sich um Daten die originär in Forschungsprozessen entstehen (z.B. Daten aus wissenschaftlichen Experimenten, Erhebungen o.ä.), jedoch können auch Daten aus anderen Quellen (Wirtschaftsdaten, Daten aus öffentlicher Hand, Registerdaten etc.) als Grundlage für datengetriebene Forschung dienen.
Datengetriebene Wissenschaft	Datengetriebene Wissenschaft bezieht sich im Rahmen dieser Studie auf die Nutzung datengetriebener Forschung im Kontext der akademischen Forschung.
Datenkompetenzen	Fähigkeiten und Wissen, um aktiv komplexe Anwendungen mit Daten selbst durchführen zu können, und diese für die eigene Forschungsarbeit nutzen zu können. Diese Fähigkeiten ermöglichen es WissenschaftlerInnen eigenständig tiefergehende Anwendungen (z.B. komplexe Analysen) von Daten selbst durchzuführen. Wird in dieser Publikation von Data Literacy (siehe Glossar) abgegrenzt.
Data Literacy	Grundverständnis für den Umgang mit Daten, für mathematisch/statistische Möglichkeiten und Einschränkungen, auch „Daten-Basisskills“ ⁵ . Eine Abgrenzung des Begriffes ist bisher nicht vollständig geklärt ⁶ . Im Kontext dieser Studie wird der Begriff für ein Wissenslevel verwendet, das es Personen ermöglicht Prinzipien von Datenverwendung zu verstehen und einfachere Anwendungen selbst durchzuführen. Komplexere Anwendungen (wie sie z.B. im Bereich Data Science durchgeführt werden) sind hier nicht eingeschlossen.
Data Science /Data Scientist	Data Science ist eine interdisziplinäres Feld, in dem statistische Methoden, Prozesse und Algorithmen verwendet werden, um Wissen aus verschiedenen strukturierten sowie nicht-strukturierten Daten zu extrahieren. Methoden wie Data Mining, Deep Learning und Machine Learning sowie Big Data kommen im Bereich Data Science zur Anwendung. Data Scientists kombinieren dazu Daten aus verschiedenen Quellen, und versuchen Erkenntnisse und Trends auf breiterer Ebene zu finden.

¹ Die Clearing Stelle der Medizinischen Universität Wien (2016).

https://www.meduniwien.ac.at/web/fileadmin/content/serviceeinrichtungen/evaluierung/Mission_Statement_Clearing_Stelle_20122016.pdf

² Definition künstliche Intelligenz (2018), <https://wirtschaftslexikon.gabler.de/definition/kuenstliche-intelligenz-ki-40285/version-263673>

	<p>Schwerpunkt-Kompetenzen von Data Scientists inkludieren z.B. Expertenwissen zu Mathematik, Statistik, maschinellem Lernen, AI, Algorithmik, Big Data Technologien, High-Performance Computing, Security und/oder Visualisierung. Der Begriff wird in dieser Studie breit verwendet und inkludiert auch andere Aufgaben, die komplexe Methoden verwenden (z.B. Data Exploration, Advanced Data Analysis etc.).</p>
Data Stewardship / Data Steward	<p>Person innerhalb einer Organisation, die für die Qualität und Einhaltung der Standards / Governance der Daten und Datenquellen verantwortlich ist. Data Stewards sind für die Umsetzung der strategischen Vorgaben der Data Governance zur Datenqualität verantwortlich.</p> <p>Der Fokus liegt u.a. auf Datenkuratierung, Metadaten, ethischen und rechtlichen Aspekten, Langzeitarchivierung und Zugänglichkeit von Daten. Im Kontext dieser Studie wird der Begriff Data Steward breit verwendet und inkludiert alle Aufgaben von professionalisiertem Datenmanagement. Der Begriff Data Steward wird hier nicht von anderen Rollen in diesem Bereich (z.B. Data Custodian, Data Hygienist etc.) abgegrenzt.</p>
Interdisziplinarität	<p>Unter Interdisziplinarität wird die Zusammenarbeit über die wissenschaftlichen Disziplingrenzen hinaus verstanden. Ziel ist es, das Wissen von verschiedenen Disziplinen in Hinblick auf ein gemeinsames Ziel zu verbinden und damit ein besseres Verständnis oder eine Lösung auf eine Frage zu bekommen, die zu komplex ist, um mit dem Wissen und der Methodik einer einzelnen Disziplin beantwortet werden zu können⁷</p>
Metadaten	<p>Daten über Forschungsdaten, die das Finden, Durchsuchen, Nutzen und Wiederverwenden von Forschungsdaten ermöglichen. Sie stellen bei richtiger Anwendung die Interpretierbarkeit und Wiederverwendbarkeit von Daten sicher.⁸</p>
Machine-actionable / maschinenverarbeitbar	<p>Informationen, die in konsistenter Weise so strukturiert werden, dass sie von Maschinen oder Computern gelesen und verarbeitet werden können.</p>
Pseudonymisierte Daten	<p>Daten, in welchen der Name und andere Identifikationsmerkmale durch ein Kennzeichen ersetzt wurden, um die Bestimmung der/des Betroffenen auszuschließen oder wesentlich zu erschweren⁹.</p>
Small Data	<p>Bezeichnet Daten, die in Umfang und Komplexität einfach, überschaubar und klein genug sind, um vom Menschen begriffen- und somit bearbeitet werden zu können.</p>

³ Definition Big Data. https://de.wikipedia.org/wiki/Big_Data

⁴ Data Science. <https://www.informatik-verstehen.de/lexikon/data-science/>

⁵ Future Skills: Ein Framework für Data Literacy (2019). https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_AP_Nr_47_DALI_Kompetenzrahmen_WEB.pdf

⁶ Data Literacy: A Systematic Review (2019). <https://www.stifterverband.org/download/file/fid/7908>

⁷ Inter- und Transdisziplinarität in der Nachhaltigkeits-orientierten Forschung (2013).

https://boku.ac.at/fileadmin/data/H99000/H99100/nachhaltigkeit/Forschung/23122013_Inter_Transdisziplin_NH.pdf

⁸ Open Science Network Austria OANA: Empfehlungen für eine nationale Open Science Strategie in Österreich. Version 1.1: Entwurf für die öffentliche Konsultation

⁹ Die Clearing Stelle der Medizinischen Universität Wien (2016).

https://www.meduniwien.ac.at/web/fileadmin/content/serviceeinrichtungen/evaluierung/Mission_Statement_Clearing_Stelle_20122016.pdf

Abkürzungsverzeichnis

AI	Artificial Intelligence
AIT	Austrian Institute of Technology
BMBWF	Bundesministerium für Bildung, Wissenschaft und Forschung
DOI	Digital Object Identifier
DSGVO	Datenschutzgrundverordnung
DMP	Datenmanagementplan
EOSC	European Open Science Cloud
ESFRI	Europäisches Strategieforum für Forschungsinfrastrukturen
ERA	European Research Area
EK	Europäische Kommission
EU	Europäische Union
FTI	Forschung, Technologie und Innovation
FOG	Forschungsorganisationsgesetz
IKT	Informations- und Kommunikationstechnik
IoT	Internet of Things
RDM	Research Data Management
RFTE	Rat für Forschungs- und Technologieentwicklung
SDG	Sustainable Development Goals

“Those who will control the world in the twenty-first century are those who will control data”

Yuval Noah Harari

1. Einleitung

Daten sind einer der wichtigsten Rohstoffe unserer Zeit und ihre Menge steigt stetig an. Durch neue Technologien, Sensoren und Internet-of-Things (IoT)-Anwendungen werden tagtäglich große Datenmengen gesammelt. So wird erwartet, dass die Menge an weltweit existierenden Daten von 33 Zettabytes im Jahr 2018 bis 2025 auf 175 Zettabytes ansteigen wird¹⁰. Neue Speichertechnologien, leistungsstarke Daten-verarbeitung und AI ermöglichen es, diese großen Daten-mengen sinnvoll zu nutzen. Daraus resultierende datengetriebene Angebote, Services und Produkte beeinflussen das Leben vieler Menschen schon jetzt auf eine nie dagewesene Art und Weise, seien es alltägliche Suchanfragen, Smart Home Anwendungen oder individualisierte Mobilitätsangebote. Im Gegensatz zu anderen, herkömmlichen Ressourcen sind Daten jedoch nicht-konkurrierend: Sie können von verschiedenen NutzerInnen zeitgleich für verschiedene Zwecke verwendet werden, ohne dass sie dabei „verbraucht werden“ – im Gegenteil, der Wert von Daten steigt oftmals dadurch, dass sie wiederverwendet werden. Daten sind somit nicht das „neue Öl“, sondern ähneln bei entsprechender Nutzung eher einem „Ölfass, das niemals leer wird“.

Auch in der wissenschaftlichen Forschung ermöglichen neue Technologien der Datensammlung, -speicherung und -analyse grundlegend neue Erkenntnisse und Beiträge mit hohem gesellschaftlichem Impact. Auch hier steigen die existierenden Datenmengen massiv an, sei es durch Simulationen (so-genannte In-Silico-Daten), durch neue Sensortechniken für Real-Life-Daten und Near-Real-Time-Datenströme und Echtzeitdaten oder durch die fortschreitende Digitalisierung von bereits bestehendem Wissen. Durch diese Entwicklungen kündigt sich schon jetzt ein Paradigmenwechsel in der globalen Wissenschaft an, der bereits drastisch Forschungsprozesse transformiert, Rollen von WissenschaftlerInnen neu definiert und enorme Potenziale für Wissenschaft und Gesellschaft bietet. Diese Potenziale gilt es durch systematische, sinnstiftende Datenanalysen und -nutzung im Kontext der akademischen Forschung, also **datengetriebener Wissenschaft**, zu heben.

Die Datenexplosion in der Wissenschaft

Täglich werden Genomdaten im Ausmaß von 10 Terrabyte produziert. Die Geschwindigkeit der Genom-Sequenzierung ist bereits so hoch, dass es technisch möglich wäre, bis 2025 die Genome von zwei Milliarden Menschen zu analysieren.¹¹

In der Astronomie produzieren einzelne Teleskope bis zu 7,5 Terrabyte an Daten pro Sekunde, das geplante Square Kilometer Array - Teleskop in Südafrika wird über 700 Terrabyte pro Sekunde produzieren.¹²

Die Geistes- und Sozialwissenschaften greifen über Social Media auf noch nie dagewesene Datenpools zu: Auf Twitter werden nahezu 500 Millionen Tweets täglich gepostet, auf Facebook sind täglich 1,5 Milliarden Nutzer aktiv.

¹⁰ The European Data Strategy (2020). https://ec.europa.eu/commission/presscorner/api/files/attachment/862109/European_data_strategy_en.pdf

¹¹ Big Data: Astronomical or Genomical? (2015). <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002195>

¹² Big Data Management, Technologies, and Applications (2013). <https://www.igi-global.com/book/big-data-management-technologies-applications/77404>

“Vor 15 Jahren waren 90 Prozent der Arbeitsplätze in einem Forschungszentrum im Labor. Baut man heute ein neues biomedizinisches Forschungszentrum, sind 90 Prozent Computer-Arbeitsplätze und nur noch 10 Prozent Labor-Arbeitsplätze.”¹³

Gil McVean, Director (Big Data Institute, University of Oxford)

1.1. Definition und Bedeutung datengetriebener Wissenschaft

Datengetriebene Wissenschaft findet in allen Disziplinen Anwendung

Datengetriebene Wissenschaft nutzt Daten, um daraus Wissen zu extrahieren. Darunter fallen Techniken und Methoden aus der Mathematik, Statistik, Stochastik und Informatik. Es handelt sich dabei nicht um einen Teilbereich einer bestimmten Disziplin, sondern um Querschnittsmethoden, welche in fast allen Forschungsdisziplinen an Bedeutung gewinnen und in der gesamten Wissenschaft zu einem erhöhten Erkenntnisgewinn

führen können. Dabei ist zu beachten, dass **datengetriebene Wissenschaft** kein grundsätzlich neues Phänomen ist, sondern bereits lange existiert, ihre Bedeutung jedoch aufgrund der Verfügbarkeit unterschiedlicher digitaler Daten als auch Verarbeitungsmöglichkeiten quer durch die Disziplinen stark zunimmt.

Datengetriebene Wissenschaft verändert die Wissenschaft tiefgreifend

Durch datengetriebene Wissenschaft verändert sich die Art und Weise, wie Forschung betrieben wird, inklusive Forschungsprozessen und Methoden fundamental. Bestand Forschungsarbeit vor wenigen Jahren in vielen Bereichen noch größtenteils darin, Experimente oder Datenerhebungen durchzuführen, so ist die Analyse von Big Data und Small Data heute in vielen Disziplinen zur wissenschaftlichen Haupt-

aufgabe geworden¹⁴. Selbst in jenen Disziplinen, die traditionell stark mit qualitativen Methoden arbeiten, wird die Auswertung digitaler Daten wichtiger. Dadurch sind neue Forschungsfelder entstanden, z.B. Computational Linguistics, Digital History (z.B. durch Massendigitalisierung von Quellen) oder auch Computational Archaeology.

Datengetriebene Wissenschaft beschleunigt den Wissensfluss zwischen den Disziplinen und führt zu neuartigen Erkenntnissen und Lösungen

Weil zunehmend mehr Daten digital und offen zur Verfügung stehen und kombinierbar sind, können sie disziplinübergreifend-, und im besten Fall durch eine verstärkte Zusammenarbeit interdisziplinär, genutzt werden. Hierdurch wird ein Wissensfluss ermöglicht, der völlig neue Erkenntnisse generiert. Anwendungsbeispiele finden sich in der Klima-

forschung oder etwa in der Zusammenarbeit von Informationstechnologie, Geologie und Sprachwissenschaft, um über die Analyse von Twitter-Meldungen eine genauere und schnellere Lokalisierung von Erdbeben-Epizentren zu bewerkstelligen als dies über herkömmliche Methoden erzielbar ist.

Covid-19, ein Beschleuniger für datengetriebene Forschung?

Die Covid-19-Pandemie hat die Wissenschaft massiv gefordert. Innerhalb kürzester Zeit mussten neue Wege eingeschlagen werden, um möglichst schnell Lösungen für diese globale Herausforderung zu finden. In dieser Zeit zeigten sich klar die Vorteile von datengetriebener Wissenschaft und was in Zukunft in diesem Bereich möglich sein wird.

So wurden sehr schnell Daten aus unterschiedlichen Disziplinen und unterschiedlicher Herkunft miteinander kombiniert und hierdurch neue Erkenntnisse gewonnen. Hierfür arbeiteten nicht nur WissenschaftlerInnen zusammen, sondern auch Unternehmen aus der Wirtschaft erkannten die Dringlichkeit der Situation und brachten sich und ihre Fähigkeiten mit ein. So konnten unter anderem anhand von Bewegungsdaten Muster der Ausbreitung der Pandemie, wie auch die Lokalisation von Infektionsherden schneller erkannt und entsprechend gehandelt werden.

Eine der Grundlagen dieser erfolgreichen Kollaborationen war die zeitnahe Bereitstellung der entsprechenden Daten und die Überwindung von Hürden, die dieses Zugänglichmachen von Daten zuvor verhindert haben. Kurz nach Erfassung der Daten wurden diese von WissenschaftlerInnen (z.B. über Preprint-Server) bereits online offen zur Verfügung gestellt, vor dem Durchlaufen herkömmlicher Reviewprozesse, teilweise ohne an konkrete Publikationen gebunden zu sein. Auch wurden Daten, die zuvor nicht für die Wissenschaft zugänglich waren, wie beispielsweise Patientendaten, zugänglich gemacht und dadurch ein schnellerer und größerer Erkenntnisgewinn, z.B. über die Wirksamkeit von Medikamenten und Impfstoffen, ermöglicht.

Hierdurch wurde die Wissenschaft innerhalb kürzester Zeit schneller, kollaborativer, interdisziplinärer, internationaler und offener. All diese Fortschritte wären ohne datengetriebene Wissenschaft nicht möglich gewesen und zeigen das Potenzial, welches aufgegriffen und langfristig genutzt werden sollte. Denn neben Covid-19 gibt es viele weitere große gesellschaftliche Herausforderungen (z.B. den Klimawandel), für deren Lösung datengetriebene Wissenschaft eingesetzt werden kann.

¹³ Studying genetics in the age of big data (2018), <https://www.spectrumnews.org/news/studying-genetics-age-big-data/>

¹⁴ Forschung & Lehre: Datengetriebene Forschung (2018), <https://www.forschung-und-lehre.de/datengetriebene-forschung-1005/>

1.2. Zunehmende politische Bedeutung von datengetriebener Wissenschaft international und national

Die wichtige Rolle von Daten in der Transformation von Wissenschaftssystemen und -politik

Daten, ihre Auswertung und Nutzung haben einen stetig wachsenden Einfluss auf das Wissenschaftssystem. Neue Möglichkeiten der Datenverarbeitung und -speicherung entwickeln sich exponentiell, die eingesetzten Technologien werden günstiger und sind sowohl schneller als auch einfacher zu bedienen. Algorithmen unterstützen ForscherInnen schon jetzt bei ihrer Arbeit, ihre Entwicklung und ihr Einsatz für die Wissenschaft werden zunehmen. Künftig ist zu erwarten, dass AI eine immer

wichtigere und auch verantwortungsvollere Rolle im Wissenschaftssystem einnehmen wird. Algorithmen werden zwar auf absehbare Zeit kein gleichberechtigtes Teammitglied sein, entwickeln sich jedoch mehr und mehr zu einem Aushängeschild von Forschungsgruppen. Diese Entwicklungen werden die Wissenschaftslandschaft als auch die Wissenschaftspolitik verändern.

Internationale politische Strategien

Diese hohen Potenziale datengetriebener Forschung für Wissenschaft und Gesellschaft werden international bereits erkannt und in wichtigen politischen Strategien forciert. Im

Folgenden werden die für Österreich einflussreichsten internationalen Strategien und Entwicklungen erläutert

Horizon Europe

Im neuen EU-Forschungsrahmenprogramm **Horizon Europe** werden Daten (insbesondere Open Data) einen noch höheren Stellenwert einnehmen als im Vorgängerprogramm Horizon2020. Insbesondere die Erweiterung des Piloten zu Open Access von Research Data ist hier zu nennen, aber auch erweiterte Anreize zum Teilen von Daten, Empfehlungen zu nachhaltigem Datenmanagement (z.B. durch Forderung von

Datenmanagementplänen (DMPs)) und ein starkes Bekenntnis zu „Open Access“ Publikationen sind Maßnahmen, mit denen sich die Europäische Kommission (EK) zu einem offenen und nachhaltigen Umgang mit Forschungsdaten bekennt. Somit dient die Ausgestaltung von Horizon Europe direkt und indirekt der datengetriebenen Forschung¹⁵.

A European Strategy for Data

In der „**European Strategy for Data**“¹⁶ drückt die EK den Anspruch der EU aus, eine datengetriebene Gesellschaft zu sein, in der Daten als zentrale Grundlage für Entscheidungen dienen. Als Hindernisse dafür identifiziert die EK u.a. einen Mangel an zur Verfügung stehender Daten, Abhängigkeiten von außereuropäischen Cloud-Lieferanten und einen Mangel an digitalen Skills. Sie schlägt eine umfassende Strategie vor, diese zu adressieren.

Die Strategie besteht aus vier Grundpfeilern:

- Etablierung eines sektorübergreifenden Rechtsrahmens

- Stärkung der Fähigkeiten und Infrastruktur der Speicherung, Verarbeitung und Nutzung von Daten
- Stärkung von Skills und Kompetenzen im Bereich Daten sowie
- Etablierung von gemeinsamen europäischen Datenräumen in strategischen Sektoren.

Obwohl sich die EK nicht auf Daten aus- oder für die Wissenschaft beschränkt, kommt datengetriebener Forschung hier ein besonderes Augenmerk zu, insbesondere da die European Open Science Cloud (EOSC) als Vorbild für weitere europäische Datenräume in anderen Sektoren dienen soll.

European Green Deal

Daten, insbesondere Forschungsdaten, bieten ein enormes Potenzial in der Bekämpfung des Klimawandels. Aus diesem Grund haben Daten auch einen hohen Stellenwert im **European Green Deal der Europäischen Kommission**, in welchem zugängliche und interoperable Daten in Kombination mit der entsprechenden Infrastruktur als zentrale Grundlagen für Forschung und Innovation zur Erreichung der Klimaziele

genannt werden¹⁷. Andererseits wird auch der Beitrag großer Datenmengen zum Klimawandel selbst erkannt, wie z.B. durch den enormen Energieverbrauch durch Datenzentren und Rechner. Darum setzt sich die EK zum Ziel konsequent neue Möglichkeiten der nachhaltigen Datennutzung (z.B. durch energieeffiziente Datacenter, IKT Dienstleistungen, Breitbandnetzwerke etc.) zu entwickeln.

¹⁵ Horizon Europe – the next research and innovation framework programme (2020). https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme_en

¹⁶ A European Strategy for Data (2020). https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

¹⁷ European Green Deal (2020). https://ec.europa.eu/info/sites/info/files/european-green-deal-communication_en.pdf

Nationale Strategien

Neben diesen internationalen Strategien, welche die Wichtigkeit von datengetriebener Wissenschaft betonen und bereits aktuell und umso mehr in Zukunft die Entwicklungen

verstärken werden, gibt es auch auf nationaler Ebene in Österreich bereits Strategien, die datengetriebene Wissenschaft adressieren.

Regierungsprogramm 2020–2024

Das aktuelle Regierungsprogramm 2020-2024 der österreichischen Bundesregierung¹⁸ sieht eine Reihe von Maßnahmen mit Implikationen für datengetriebene Forschung vor. Insbesondere die Schaffung der notwendigen Datenzugänge für die österreichische Forschung ist vorgesehen. Dafür sollen vermehrt Registerdaten für die Wissenschaft zur Verfügung gestellt werden, dies soll durch die Schaffung des „Austrian Micro Data Center“ unter Einbeziehung relevanter StakeholderInnen (wie z.B. Statistik Austria) realisiert werden. Der rechtliche Rahmen dafür soll durch eine Novellierung des Bundesstatistikgesetzes geschaffen werden. Auch die wissenschaftliche Nutzung anonymisierter Pflege- und Gesundheitsdaten soll geprüft werden. Darüber hinaus bekennt sich die Bundesregierung zu Open Government Data, und zu einer umfassenden Umsetzung der Open Data-PSI

Richtlinie der EU. Ergänzend sieht sie eine Umsetzungsstrategie vor, um das Prinzip „Open by Default“ für nicht-personalisierte Daten des Bundes zu etablieren. Weiters sind Erweiterungen von wichtigen Forschungsinfrastrukturen vorgesehen. Zum einen wird die Wichtigkeit des Zugangs zu europäischen Infrastrukturen betont. Doch auch national sollen entsprechende Infrastrukturen und Rechenkapazitäten ausgebaut werden, insbesondere bezüglich AI (wie z.B. dem Vienna Scientific Cluster¹⁹ oder dem Complexity Science Hub²⁰). Ein besonderes Augenmerk wird auf die Weiterentwicklung von AI in Österreich gelegt. Dazu ist die Entwicklung einer AI-Strategie vorgesehen, sowie in weiterer Folge die Entwicklung von Richtlinien für die Nutzung von AI und die Setzung entsprechender Forschungsschwerpunkte.

Austrian ERA Roadmap

Das ERA Austria Portal dient als Informations-Hub für europäische Forschungspolitik und ihre Umsetzung in Österreich, und veröffentlicht im Rahmen der Österreichischen ERA Roadmap Empfehlungen für die zukünftige Entwicklung des österreichischen Forschungsraums²¹. Unter anderem empfiehlt ERA die konsequente Ausrichtung des österreichischen Forschungsbetriebs auf Europa, beispielsweise in Bezug auf Forschungsinfrastrukturen (z.B. ESFRI

Roadmap Infrastrukturen), die verstärkte Adressierung großer gesellschaftlicher Herausforderungen durch die Wissenschaft sowie den Ausbau des offenen Zugangs zu Forschungsdaten (Open Access to Research Data). ERA betont weiters die Wichtigkeit der notwendigen Lizenzmodelle und Sicherstellung der Verwertungsrechte an geistigem Eigentum für WissenschaftlerInnen (inklusive Forschungsdaten).

¹⁸ Aus Verantwortung für Österreich. Regierungsprogramm 2020-2024 (2020). <https://www.bundeskanzleramt.gv.at/bundeskanzleramt/die-bundesregierung/regierungsdokumente.html>

¹⁹ Vienna Scientific Cluster. <https://vsc.ac.at/home/>

²⁰ Complexity Science Hub. <https://www.csh.ac.at/>

²¹ Österreichische ERA Roadmap (2016). https://era.gv.at/object/document/2581/attach/oesterreichische_ERA_Roadmap.pdf

1.3. Zielsetzungen der Studie

Um in Österreich zukunftsorientierte FTI-Politik betreiben und den Forschungs- und Innovationsstandort weiterentwickeln zu können, muss der Einfluss datengetriebener Wissenschaft auf

das Wissenschaftssystem im Detail bekannt sein. Daher wurden im Rahmen dieser Studie zwei zentrale Fragen bearbeitet:

1. **Welche Herausforderungen** kommen durch die **rasant stattfindenden technologischen, organisationalen und inhaltlichen Entwicklungen** im Bereich datengetriebener Wissenschaft **auf die österreichischen Forschungseinrichtungen zu?**
2. **Was sollte getan werden, damit das Wissenschaftssystem**, vornehmlich Hochschulen und außerhochschulische Forschungseinrichtungen, auf die Veränderungen, welche die digitale Datenrevolution mit sich bringt, **nicht nur vorbereitet ist, sondern auch international wettbewerbsfähig bleibt?**

Ein wesentliches Augenmerk lag im Rahmen der zweiten Frage auf der Entwicklung von konkreten Handlungsempfehlungen, mit deren Hilfe Politik und Forschungseinrichtungen rechtzeitig Voraussetzungen für den Erfolg in einem zunehmend datengetriebenen, internationalen Umfeld schaffen können.

Um nicht die aktuelle Entwicklung als Maßstab der erforderlichen Schritte zu nehmen, wurde als Zeithorizont für den Veränderungsbedarf das Jahr 2030 definiert. Auf diese Weise können die hohe Dynamik und rasche Weiterentwicklung der Themenstellung in abzuleitenden Schlussfolgerungen und Empfehlungen antizipiert werden.

Unter anderem wurde erarbeitet, auf welche Weise digitale Daten als essenzieller Zukunftsrohstoff spezifisch für die österreichische Forschungslandschaft dienen können. Dabei wurden verschiedene Arten von Daten berücksichtigt, wie Small Data und Big Data aus unterschiedlichsten Quellen, sowie Synthetische Daten. Zudem wurde der Einfluss der Datenexplosion, neuer technologischer Möglichkeiten und regulatorischer Rahmenbedingungen auf das Sammeln, Speichern, Warten, Teilen und Löschen von Datenbeständen berücksichtigt. Besonderes Augenmerk lag darauf, durch welche Voraussetzungen die Datensouveränität sowie Datenhoheit verschiedener StakeholderInnen der österreichischen Forschungslandschaft sichergestellt werden kann. Des Weiteren wurde analysiert, welche Humanressourcen (v.a. Kompetenzen und Skills) sowie Infrastrukturen und Organisationseinheiten (z.B. Big-Data- oder AI Zentren) an den Forschungseinrichtungen zentral oder in den einzelnen Disziplinen verfügbar gemacht werden sollten.

Zusätzlich wurde behandelt, welche Datenmanagement- und Datentreuhandsysteme für die österreichische Hochschullandschaft zur Verfügung gestellt werden sollten, und welche

Standards und Qualitätskriterien geschaffen werden müssen, um Interoperabilität verschiedener Datenbestände sicherzustellen.

Berücksichtigt wurde zudem, wie man internationale, interdisziplinäre Forschung vorantreiben kann, insbesondere unter Berücksichtigung der Fragmentierung zahlreicher Regelungen zur Nutzung von Datenbeständen in der Forschung (z.B. auf institutioneller-, Bundes-, Länder und EU-Ebene und zwischen den Mitgliedsstaaten).

Völlig neue, datengetriebene Forschung erfordert auch neue Ansätze zur strategischen Ausrichtung von Forschungsstätten. Ziele und Erfolgsindikatoren von WissenschaftlerInnen sowie Institutionen müssen aufgrund der Datenrevolution neu gedacht werden. Weiters wurde betrachtet, wie adäquate Unterstützungsangebote (z.B. Kompetenzzentren, die Beratungsfunktionen für Forschungseinrichtungen ausüben) gestaltet werden können.

In diesem Zusammenhang wurde weiters erhoben, ob die österreichischen Forschungseinrichtungen Rahmenbedingungen vorfinden, die es ihnen erlauben, notwendige Daten zu erhalten und damit zu agieren. Es wurde erläutert, welche potenziellen Hindernisse (z.B. im Zugang zu Registerdaten, amtlicher Statistik, personenbezogenen Daten) herrschen und wie diese überwunden werden können.

Folgend wurde herausgearbeitet, welche Themen und Handlungsfelder österreichische Hochschulen und außerhochschulische Forschungseinrichtungen einerseits und die österreichische Forschungspolitik andererseits bearbeiten sollten, um für die Herausforderungen von datengetriebener Forschung gerüstet zu sein und optimalen Nutzen daraus ziehen zu können.

2. Methode und Beteiligte

Um diese Ziele zu erreichen wurde unter Anwendung von Open Innovation Methoden, unter Beteiligung von internationalen und österreichischen führenden ExpertInnen aus Wissenschaft, Forschungspolitik und Governance eine umfassende Analyse datengetriebener Forschung in- und außerhalb von Österreich durchgeführt und konkrete Handlungsempfehlungen erarbeitet.

Als Ausgangspunkt wurde eine internationale Desktop Recherche zur Identifikation künftiger Entwicklungen im Bereich der datengetriebenen Forschung und daraus resultierenden Herausforderungen für die österreichische Forschungslandschaft durchgeführt. Hierdurch konnte ein Überblick über die internationalen und nationalen Entwicklungen, welche künftig an Relevanz gewinnen werden, generiert werden. In weiterer Folge wurden semistrukturierte individuell an die Expertise und Vorkenntnisse angepasste Tiefeninterviews mit internationalen und – international vernetzten österreichischen ExpertInnen im Bereich der datengetriebenen Forschung durchgeführt.

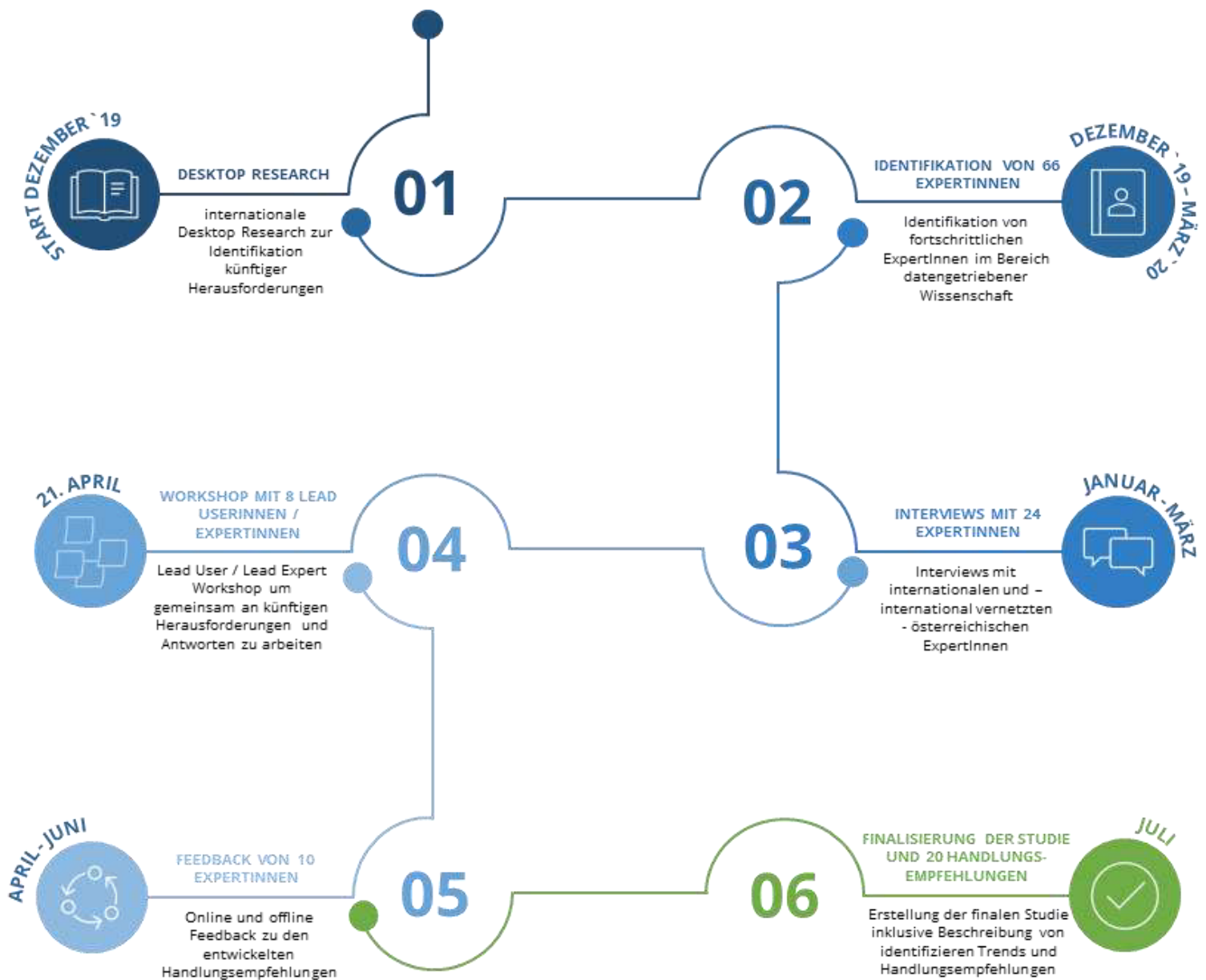
Durch eine durchgehende deduktiv-induktive Analyse konnte ein klares Bild der Lage Österreichs gezeichnet werden, aus dem sich erste konkrete Herausforderungen und Anforderungen für zukünftige datengetriebene Wissenschaft ableiten ließen. Diese wurden im Rahmen eines Lead User / Lead Expert

Workshops mit visionären WissensgeberInnen und innovativen DenkerInnen konkretisiert und vertieft, sowie konzise Lösungsansätze für die österreichische Wissenschaft identifiziert. Hierfür wurden erste Visionen zur Zukunft der datengetriebenen Wissenschaft im Jahre 2030 erarbeitet und nötige Voraussetzungen hierfür abgeleitet. Ausgehend von diesen Voraussetzungen wurden in Folge in drei vertiefenden Themengebieten Handlungsempfehlungen für datengetriebene Forschung in Österreich für Politik und Forschungseinrichtungen erarbeitet. Die drei Themengebiete waren die folgenden:

- Organisation und Governance
- Humanressourcen und Skills
- Infrastruktur und Ressourcen

Nachfolgend wurden die in diesen Themengebieten erarbeiteten Handlungsempfehlungen ausformuliert und den ExpertInnen für tiefergehendes Feedback zur Verfügung gestellt, sowie zu spezifischen Themen weitere ExpertInnen in Form von spezifischen Feedback-Interviews zu Rate gezogen. Aus den gesammelten Ergebnissen aus dem Workshop, Interviews und der Recherche wurden abschließend zukünftige Entwicklungen im Bereich der datengetriebenen Forschung definiert und abschließend konkrete und praxistaugliche Handlungsempfehlungen für die Forschungspolitik abgeleitet, die den Kern der vorliegenden Studie bilden.

Abbildung 3 **Prozess der Studienerstellung**



Wir möchten den beteiligten ExpertInnen herzlich für ihren Input und ihr Engagement danken!

Folgende ExpertInnen haben ihre Expertise in Form von qualitativen Tiefeninterviews beigetragen:

Susanne Blumesberger, Leitung Abteilung Repositorienmanagement PHAIDRA-Services | Universität Wien

Raman Ganguly, Leiter Software Design and Development Zentraler Informatikdienst | Universität Wien

Allan Hanbury, Professor of Data Intelligence | TU Wien

Nikolaus Hautsch, Professor of Finance and Statistics; stv. Leiter - Forschungsplattform Data Science | Universität Wien

Tobias Hell, Leiter des Universitätslehrgangs Data Science - From Mathematical Foundations to Applications | Universität Innsbruck

Monika Henzinger, Head of Research Group Theory and Applications of Algorithms | Universität Wien

Devika Madalli, Professor and Head - Documentation Research and Training Center (DRTC) | Indian Statistical Institute (ISI)

Torsten Möller Head of Research Group Visualization and Data Analysis | Universität Wien

Silvana Muscella, CEO Trust-IT Services Ltd; Expert Evaluator/Reviewer DG Connect, DG RTD, EASME, H2020 SME Instruments & (FTI) Programmes | European Commission

Harald Oberhofer stv. Institutsvorstand des Instituts für Internationale Wirtschaft | WU Wien; Senior Economist | WIFO; Plattform Registerforschung

Andreas Rauber, Head of Information and Software Engineering Group | TU Wien

Falk Reckling, Abteilungsleiter Strategie - Policy, Evaluation, Analyse | FWF

Katharina Rieck, Open Science Management und Publikationskosten Strategy - Policy, Evaluation, Analysis | FWF

Tony Ross-Hellauer, Forschungsgruppenleiter | TU Graz und Know-Center GmbH

Sarah Spiekermann-Hoff, Professorin, Leitung Institute for Information Systems & Society | WU Wien

Stefan Strauß, Senior Scientist, Institut für Technikfolgenabschätzung | ÖAW

Achim Streit, Professor, Direktor Steinbuch Center for Computing (SCC) | Karlsruher Institut für Technologie

Gabi Schneider, Universitätsbibliothek Basel; ehem. Projektleiterin Open Access | swissuniversities

Andreas Trügler, Senior Researcher and Project Leader | Know-Center GmbH

Michalis Tzatzanis, Experte europäische und internationale Programme | FFG

Wolfgang Wagner, Head of Research Group Microwave Remote Sensing | TU Wien; Mitgründer und wissenschaftlicher Berater | EODC Earth Observation Data Center

Barbara Weitgruber, Sektionschefin Wissenschaftliche Forschung und internationale Angelegenheiten | Bundesministerium für Bildung, Wissenschaft und Forschung

Sonja Zillner, Key Expert | Siemens AG

Folgende ExpertInnen haben ihre Expertise im Rahmen eines Co-Creation Workshops in die Studie einfließen lassen:

Susanne Blumesberger, Leitung Abteilung
Repositorienmanagement PHAIDRA-Services | Universität Wien

Manfred Halver, Nationale Kontaktstelle für
Forschungsinfrastrukturen Experte Informations- &
Kommunikationstechnologien, Europäische und internationale
Programme | FFG

Allan Hanbury, Professor of Data Intelligence | TU Wien

Tobias Hell, Leiter des Universitätslehrgangs Data Science - From
Mathematical Foundations to Applications | Universität Innsbruck

Andreas Rauber, Head of Information and Software Engineering
Group | TU Wien

Katharina Rieck, Open Science Management und
Publikationskosten Strategy - Policy, Evaluation, Analysis | FWF

Gabi Schneider, Universitätsbibliothek Basel; ehem.
Projektleiterin Open Access | swissuniversities

Folgende ExpertInnen haben Feedback zu einzelnen Kapiteln der Studie oder Zwischenergebnissen des Workshops abgegeben:

Susanne Blumesberger, Leitung Abteilung
Repositorienmanagement PHAIDRA-Services | Universität Wien

Raman Ganguly, Leiter Software Design and Development
Zentraler Informatikdienst | Universität Wien

Harald Oberhofer, stv. Institutsvorstand des Instituts für
Internationale Wirtschaft | WU Wien; Senior Economist | WIFO;
Plattform Registerforschung

Andreas Rauber, Head of Information and Software Engineering
Group | TU Wien

Katharina Rieck, Open Science Management und
Publikationskosten Strategy - Policy, Evaluation, Analysis | FWF

Gabi Schneider, Universitätsbibliothek Basel; ehem.
Projektleiterin Open Access | swissuniversities

Gerhard Schwarz, Koordinator - Kompetenzzentrum
Konjunktur- und Unternehmensbefragungen | WIFO; Plattform
Registerforschung

Michael Strassnig, Programmmanager | Wiener Wissenschafts-,
Forschung- und Technologiefonds; Plattform Registerforschung

Christiane Wendehorst, Professorin für Zivilrecht, Präsidentin
des European Law Institute, Stellvertretende Institutsvorständin
am Institut für Digitalisierung und Recht | Universität Wien

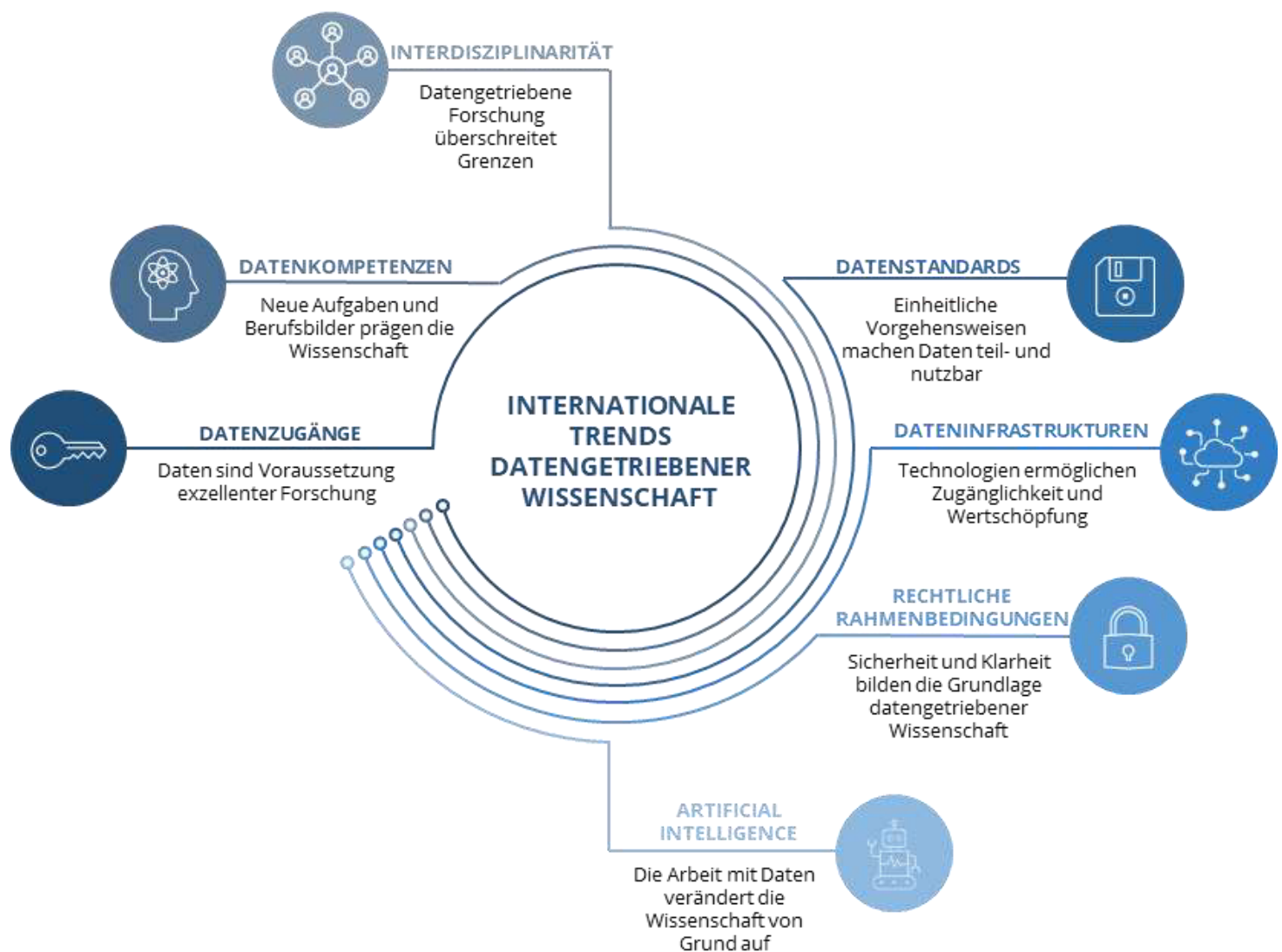
3. Internationale Trends datengetriebener Wissenschaft und Status quo in Österreich

Die Nutzung von Daten kann zu großen Erkenntnisssprüngen innerhalb der Wissenschaft führen und radikale Innovationen hervorbringen. Komplexe Problemstellungen wie etwa im Bereich des Klimawandels sind nur durch datengetriebene Wissenschaft zu bearbeiten und zu lösen. Um digitale Daten sinnvoll einsetzen zu können, müssen jedoch verschiedene Anforderungen erfüllt werden, wie die folgende Analyse internationaler Trends und Entwicklungen bis 2030 zeigt.

Im Rahmen dieser Studie wurden **sieben zentrale Bereiche** genauer beleuchtet und dabei aktuelle internationale Entwicklungen und die daraus resultierenden Anforderungen dargestellt. Weiters wurde der aktuelle Status quo in Österreich analysiert.

In Summe liegt Österreich nach Meinung der interviewten ExpertInnen im Mittelfeld, was Rahmenbedingungen und Voraussetzungen für datengetriebene Wissenschaft betrifft. Einige Anforderungen werden bereits adressiert oder es bestehen schon fortschrittliche Projekte, doch um zu den führenden Nationen aufzuholen und langfristig eine exzellente Qualität der datengetriebenen Forschung in Österreich zu gewährleisten, gibt es in bestimmten Bereichen Nachholbedarf.

Abbildung 4 **Die sieben internationalen Trends datengetriebener Wissenschaft**



3.1. Datenzugänge -

Daten sind Voraussetzung exzellenter Forschung

Um exzellente Forschung betreiben zu können, sind Zugänge zu hochqualitativen Daten aus verschiedensten Quellen (öffentliche Hand, Wirtschaft, Wissenschaft und Gesellschaft) absolut notwendig.

Leider steigt jedoch in den letzten Jahren das für die Wissenschaft nicht zugängliche Wissen (sogenanntes Dark Knowledge). Umso mehr sind vorhandene Datenzugänge ein zentraler Standortfaktor für die Wissenschaft.

Daten der öffentlichen Hand

Unter anderem deshalb ermöglichen fortschrittliche Nationen den Zugang zu Registerdaten für die Forschung. In Dänemark wurden die rechtlichen Grundlagen dafür bereits 2008 geschaffen und der Zugang durch Statistics Denmark gewährleistet²².

Denn Datenzugänge und -verfügbarkeit von Daten aus der öffentlichen Hand sind Grundlage kompetitiver Forschung und daher auch wichtiger Standortfaktor. Zudem kann der Zugang und die Analyse solcher Daten zu einem Erkenntnisgewinn über das Gesamtsystem Wissenschaft führen.

Daten aus der Wirtschaft

Aufholbedarf herrscht in Bezug auf Daten aus der Wirtschaft, die in ungenügendem Maße in der Wissenschaft genutzt werden können. Dies wurde nicht nur von den InterviewpartnerInnen in Österreich bedauert, sondern auch auf EU-Ebene als großes ungenutztes Potenzial identifiziert²³.

Unter anderem werden die großen Aufwände, Daten rechtskonform und in entsprechender Qualität bereitzustellen und die nicht vorhandenen Anreizstrukturen für eine Datenbereitstellung aus der Wirtschaft als Gründe genannt, dass Daten oftmals nicht zur Verfügung stehen.

Open Data aus der Wissenschaft

Doch nicht nur Daten der öffentlichen Hand und der Wirtschaft sind für die Wissenschaft von Interesse. Auch Datensätze aus der Wissenschaft selbst werden vermehrt öffentlich zugänglich gemacht. International ist diese Open Data Bewegung bereits sehr aktiv, denn hierdurch kann der Nutzen von Daten aus der Forschung stark gesteigert werden. Für die weltweite Öffnung von Forschungsdaten setzt sich beispielsweise die Non-Profit Organisation Open Data Watch ein²⁴. Und auch auf europäischer Ebene wird bereits unter

Horizon 2020 die European Open Science Cloud gefördert (EOSC), ein Portal, welches unter anderem Open Data zugänglich macht. Unter Horizon Europe wird dieses Projekt weiter gefördert und Open Science, also die Öffnung von Forschungsprozessen und wissenschaftlichen Erkenntnissen²⁶ zu einem der Grundsätze für Förderungen erklärt²⁷. Dies ist bereits unter dem Open Research Data Pilot im Programm Horizon 2020 der Fall²⁸.

Open Data Directive 2019/1024 (Open Data und PSI Richtlinie)

Um weitreichend Daten und Informationen aus dem öffentlichen Sektor u.a. der Forschung zugänglich zu machen, wurde 2018 eine Novelle der Open Data und Public Sector Information (PSI) Richtlinie beschlossen. Datengetriebener Forschung kommt in dieser Neufassung ein besonders hoher Stellenwert zu, da ihr Geltungsbereich nun, mit Ausnahmen, auf öffentlich finanzierte Forschungsdaten ausgeweitet wurde. Diese waren bisher zu einem großen Teil ausgenommen. Somit wird klar das Ziel verfolgt, das Potenzial von Daten aus der Wissenschaft einerseits für datengetriebene wissenschaftliche Weiternutzung, andererseits für datengetriebene Innovation zu heben²⁵. Um diese Ziele zu erreichen wird jedoch eine möglichst offene und breite Ausgestaltung auf nationalstaatlicher Ebene ein wesentlicher Erfolgsfaktor sein. Im Rahmen der Umsetzung werden derzeit „hochwertige Datensätze“ identifiziert, welche kostenlos verfügbar und maschinenleserlich zugänglich sein sollen. Die EK plant dies über gesonderte Rechtsakte durchzuführen.

²² The danish system for access to micro data (2014). <https://www.dst.dk/ext/645846915/0/forskning/Access-to-micro-data-at-Statistics-Denmark-2014--pdf>

²³ The economics of Business-to-Government data sharing (2020). <https://ec.europa.eu/jrc/sites/jrcsh/files/jrc119947.pdf>

²⁴ Open Data Watch. www.opendatawatch.com

²⁵ Directive (EU) 2019/1024 (2019). <https://eur-lex.europa.eu/eli/dir/2019/1024/oj>

²⁶ Open Science Network Austria OANA: Empfehlungen für eine nationale Open Science Strategie in Österreich. Version 1.1: Entwurf für die öffentliche Konsultation

²⁷ European Open Science Cloud. <https://ec.europa.eu/digital-single-market/en/european-open-science-cloud>

²⁸ Open Data in Horizon 2020. https://www.ffg.at/europa/recht-finanzen/h2020-open_data

Doch auch nationale Förderagenturen haben den Mehrwert von Open Data erkannt und fördern vermehrt die Öffnung und Bereitstellung von Daten in hoher Qualität als auch Aktivitäten, die dies erleichtern. Irland fördert so seit 2015 Engagement im Bereich der Open Data, die Entwicklung von Anwendungen, Visualisierungen oder Use Cases, welche die Vorteile von Open Data illustrieren²⁹.

Die Unterstützung von Open Data führt langfristig zu einer Veränderung der Publikationspraxis. Längst sind nicht mehr Publikationen die ultima ratio, sondern vermehrt werden komplette Datensätze in Kombination mit Publikationen, Datensätze allein als auch Algorithmen veröffentlicht. So nimmt die Bedeutung von sogenannten Data Papers, also der Veröffentlichung von Datensätzen mit entsprechender Dokumentation³⁰ laut den befragten ExpertInnen aktuell stark zu. Problematisch ist nach wie vor die mangelnde Veröffentlichung von Negativdaten, also Daten, die im Rahmen von nicht erfolgreichen Experimenten bzw. Forschungsprojekten entstanden.

Open Access

Open Access beschreibt den unbeschränkten und kostenlosen Zugang zu wissenschaftlicher Information im Internet unter Verwendung offener Lizenzen. Der Begriff Open Access wird oftmals in Bezug auf wissenschaftliche Publikationen („Open Access to Publications“) verwendet, jedoch bezieht sich Open Access auch auf Primär- und Metadaten („Open Access to Research Data“), Quellentexte, und digitale Reproduktion.

Publikationen können entweder ausschließlich in dezidierten Open Access Medien erfolgen, mit weitreichenden Weiternutzungsmöglichkeiten und meist unter Wahrung aller Rechte der UrheberInnen („Gold Open Access“). Alternativ können Publikationen zusätzlich zur Veröffentlichung in herkömmlichen Medien, meist nach einer Embargoperiode, auf Repositorien, privaten Websites der Urheber o.ä. veröffentlicht werden („Green Open Access“).

In Europa wird ab 2021 der Plan S in Kraft treten, welcher zum Ziel hat, den freien Zugang (Open-Access) zu wissenschaftlichen Erkenntnissen aus öffentlichen Mitteln voranzutreiben.

Die Situation in Österreich

Von einem großen Teil der InterviewpartnerInnen wurde angemerkt, dass in Österreich für die Wissenschaft der Zugang zu Daten noch nicht ausreichend gewährleistet wird. Dies betrifft sowohl Daten der öffentlichen Hand, Daten aus der Wirtschaft und

den Zugang zu Open Data. Der Zugang zu hochqualitativen und hochaktuellen Daten bildet jedoch eine wesentliche Voraussetzung für einen konkurrenzfähigen Wissenschaftsstandort.

Öffnung von Daten der öffentlichen Hand nötig

Besonders im Vergleich zu anderen Ländern mangelt es laut den InterviewpartnerInnen in Österreich massiv an der Möglichkeit, Daten der öffentlichen Hand (z.B. Registerdaten, aber auch Open Government Data) in der Wissenschaft verwenden zu können. Hierdurch können einerseits wichtige Erkenntnisse über das österreichische System nicht gewonnen werden (z.B. hinsichtlich Gesundheitsversorgung), andererseits wird dadurch die Forschung selbst erschwert. Erfolgreiche WissenschaftlerInnen in datengetriebenen Disziplinen siedeln sich nicht in Österreich an, da ihnen hier die entsprechenden Datenzugänge nicht geboten werden. Die Interviews haben folglich ergeben, dass Österreich dadurch langfristig an Wettbewerbsfähigkeit verliert.

Zudem wird von befragten ExpertInnen kritisiert, dass die mangelnden Zugänge zu sensiblen Daten verschiedener Arten (aus öffentlicher Hand, Wissenschaft und Wirtschaft) konkret das Erarbeiten von Lösungen für große gesellschaftliche Herausforderungen (u.a. adressiert in den fünf Missionen von Horizon Europe³¹ und den SDGs) und die Beteiligung an internationalen Initiativen etwa im Klimaschutz oder Gesundheitsbereich erschwert.

Diese Problematik sollte zwar 2018 durch die Novelle des Forschungsorganisationsgesetzes (FOG)³² teilweise gelöst werden, indem diese beispielsweise Daten aus Melderegistern, Sozialversicherungen, Standesämtern und medizinischen Datenbanken durch eine Einschränkung des Datenschutzes im Interesse

²⁹ Open Data Engagement Fund. <https://data.gov.ie/pages/open-data-engagement-fund>

³⁰ TU Wien, Data Paper. <https://www.tuwien.at/forschung/fti-support/forschungsdaten/aufbewahren-und-publikieren/data-paper/>

³¹ European Commission: Horizon Europe – the next research and innovation framework programme (2019). https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme_en

³² Bundesgesetz über allgemeine Angelegenheiten gemäß Art. 89 DSGVO und die Forschungsorganisation (Forschungsorganisationsgesetz – FOG) (2018). <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009514>

der Wissenschaft³³ prinzipiell zugänglich gemacht hat. Jedoch fehlen hier nach ExpertInnenmeinung die entsprechenden Verordnungen der Ministerien, um die Daten tatsächlich für die Wissenschaft zugänglich und nutzbar zu machen. Im Regierungsprogramm 2020-2024 ist, um der Problematik der geringen Verfügbarkeit der Daten entgegen zu wirken, ein so-

ge nanntes „Austrian Micro Data Center“ geplant. Dieses soll akkreditierten WissenschaftlerInnen Zugang zu den Datenbeständen der Statistik Austria gewähren, ohne dass eine Rückführung auf Einzelfälle möglich ist¹⁸.

Mangelnde Open Data und Zugriff zu Daten aus der Wissenschaft

In Bezug auf Open Data gibt es in Österreich bereits einige Initiativen und begonnene Projekte, die diese Problematik beheben wollen. So fördert beispielsweise der FWF seit 2019 Datenmanagementpläne (DMP) und übernimmt unter anderem die Kosten der Aufbereitung der Daten und der Archivierung von Forschungsdaten in Repositorien³⁴. Auch das Open Science Network Austria (OANA), welches sich seit 2012 mit den Themen Open Science, Open Access und Open Research Data befasst, setzt sich für die vermehrte Zurverfügungstellung und Ermöglichung von Open Data ein.³⁵ Diverse Institutionen bieten zudem einerseits finanzielle Unterstützung von Open Access-Publikationskosten im Rahmen von Publikationsfonds an. Andererseits stellen sie institutionelle Repositorien zur Archivierung und Weiternutzung von Daten zur Verfügung³⁶, unter anderem sind hier die Universität Wien (Repositorium Phaidra) und die Johannes Kepler Universität Linz zu nennen. Diese Entwicklungen werden von ExpertInnen als sehr positiv bewertet, da Forschungsprojekte, aus denen Daten für andere WissenschaftlerInnen zugänglich gemacht werden, einen wesentlich höheren Nutzen stiften als solche, in denen die Daten nicht nachgenutzt werden.

Jedoch wird von den befragten ExpertInnen bemängelt, dass in Österreich strategisch initiierte Projekte und Zusammenschlüsse fehlen. Die Datennutzbarmachung wird von der Forschungspolitik nicht durchgehend eingefordert und ist nicht ausreichend durch Förderungen oder andere Incentivierungsmaßnahmen in der österreichischen Wissenschaftslandschaft verankert. Zudem wird der Aufwand für das Research Data Management, also Datenhaltung und -pflege (RDM), stark unterschätzt und ist in den Finanzierungsinstrumenten in zu geringem Maße abgebildet.

Einige InterviewpartnerInnen haben die Unsicherheit der WissenschaftlerInnen bezüglich der Folgen der Öffnung von Daten als Hemmnisse für das Teilen von Daten genannt. Hier fehlt es ihnen zufolge an Aufklärung bei WissenschaftlerInnen und verbesserten Rahmenbedingungen, doch verhindert aktuell auch oft die herrschende Kultur innerhalb der Wissenschaft und Wirtschaft das Teilen und Öffnen von Daten, da es keine positive Reputation für individuelle ForscherInnen bzw. Forschungsgruppen bringt.

³³ Wo das Forschungs- datenschutzgesetz aussticht (2018). <https://www.derstandard.at/story/2000077962946/wo-das-forschungs-das-datenschutz-gesetz-aussticht>

³⁴ FWF: Forschungsdatenmanagement (2019) <https://m.fwf.ac.at/de/forschungsfoerderung/open-access-policy/forschungsdatenmanagement/>

³⁵ OANA: <https://www.oana.at/>

³⁶ Open Science Network Austria OANA: Empfehlungen für eine nationale Open Science Strategie in Österreich. Version 1.1: Entwurf für die öffentliche Konsultation

3.2. Datenkompetenzen –

Neue Aufgaben und Berufsbilder prägen die Wissenschaft

Der Umgang mit Daten erfordert von WissenschaftlerInnen spezifische Fähigkeiten und Kenntnisse sowie den Einsatz von Fachkräften in Forschungseinrichtungen. Je nach Art der

Tätigkeit und der dadurch erforderlichen Kompetenzen, wird in dieser Studie zwischen **Data Science** und **Data Stewardship** unterschieden.

Data Science Kompetenzen

Dabei bezieht sich **Data Science** auf die Extraktion von Wissen aus Daten, also vor allem auf die Analyse von Daten. Dazu gehört es, Daten aus verschiedenen Quellen zu kombinieren und daraus Erkenntnisse und Entwicklungen ableiten zu können.

Basis dafür sind Kompetenzen aus Mathematik, Statistik, Maschinellem Lernen, AI, Algorithmik, Big Data Technologien, High-Performance Computing, Security und Visualisierung.

Data Stewardship Kompetenzen

Forschungseinrichtungen benötigen zudem Kompetenzen im Bereich der **Data Stewardship**. Hierunter fallen im Wesentlichen Aufgaben im Bereich der Qualitätssicherung und Einhaltung von Standards / Data Governance sowie der Pflege und Erschließung von Datenquellen. Der Fokus liegt hierbei auf der Daten-Kuratierung, Metadaten, ethischen und rechtlichen Aspekten, der Archivierung und Accessibility von Daten sowie der Verantwortung für das Teilen von Daten.

Während sich also Data Science mit der konkreten Durchführung datengetriebener Forschung befasst, stellt Data Stewardship sicher, dass die richtigen Rahmenbedingungen (z.B. an einzelnen Forschungseinrichtungen) herrschen und die Daten langfristig den höchsten Nutzen generieren können. In diesen beiden Schlüsselbereichen der datengetriebenen Forschung gibt es eine Vielzahl an verschiedenen, sich gerade definierenden Berufs- und Aufgabenfeldern sowie Spezialisierungen.

Grundkenntnisse in beiden Bereichen: Data Literacy in der wissenschaftlichen Forschung

Auf Ebene der Forschungseinrichtungen sind Data Science und Data Stewardship in zweierlei Hinsicht von Relevanz: Erstens werden bei ForscherInnen in allen wissenschaftlichen Disziplinen gewisse Grundlagenkompetenzen, also eine generelle **Data Literacy**, in beiden Bereichen nötig sein, zweitens braucht es **Spezialisierungen**. Der Umgang mit Daten ist zentral für den Erkenntnisgewinn innerhalb der Wissenschaften. Data Literacy, also ein Grundverständnis zu statistischen- bzw. mathematischen Möglichkeiten von Daten und Forschungsdatenmanagement, wird in den kommenden Jahren zu einem grundlegenden Bestandteil von wissenschaftlichen Ausbildungen werden. In Deutschland wird Data Literacy beispielsweise seit Januar 2020 durch das Programm Data Literacy Education.nrw unterstützt. Dabei werden zehn Hochschulen gefördert, welche fachübergreifende Data Literacy Programme anbieten³⁷. Doch kann das Wissen zu Data Literacy bereits in der Primär- und Sekundärstufe gestärkt werden: An der US-amerikanischen Carnegie Library of Pittsburgh werden spezielle Data Literacy Programme und Events für Kinder und Jugendliche durchgeführt³⁸.

DORA Declaration

Die 2012 erstellte San Francisco Declaration on Research Assessment, listet eine Reihe von Empfehlungen zur Bewertung von Forschungsergebnissen, abseits von oft gängigen Publikationsbasierten Metriken (Journal Impact Factor). Unter anderem wird die Bewertung von anderen Forschungsleistungen, einschließlich der Erstellung von Datensätzen und Software empfohlen. Die DORA Declaration wurde inzwischen bereits von ca. 1980 Organisationen unterzeichnet³⁹, in Österreich unterstützt unter anderem der FWF die DORA Deklaration bereits seit 2013.⁴⁰

Das Öffnen und Zurverfügungstellen von Daten sollte auch in der Karriereentwicklung und Leistungsbewertung von WissenschaftlerInnen bewertet und honoriert werden, etwa indem Metriken und Kennzahlen für Leistungsbewertung weiterentwickelt und nicht primär publikationsbasierte Metriken genutzt werden⁴¹. Dies fordern internationale WissenschaftlerInnen im Leiden Manifesto⁴² oder auch in der DORA Deklaration³⁹, zu denen sich bereits viele führende

³⁷ Land und Stifterverband fördern Hochschulen mit drei Millionen Euro für die Vermittlung von Datenkompetenzen (2019). <https://www.land.nrw/de/pressemitteilung/land-und-stifterverband-foerdern-hochschulen-mit-drei-millionen-euro-fuer-die>

³⁸ Why Kids Need Data Literacy, and How You Can Teach It (2017).

<https://www.slj.com/?detailStory=why-kids-need-data-literacy-and-how-you-can-teach-it>

³⁹ San Francisco Declaration on Research Assessment (DORA) (2013).

<https://sfidora.org/signers/>

⁴⁰ FWF unterstützt San Francisco Declaration on Research Assessment (2013)

<https://www.fwf.ac.at/de/news-presse/news/nachricht/nid/20130521-1694/>

⁴¹ Open Science Network Austria OANA: Empfehlungen für eine nationale Open Science Strategie in Österreich, Version 1.1: Entwurf für die öffentliche Konsultation

⁴² Leiden Manifesto for research metrics (2015). <https://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351>

internationale Forschungseinrichtungen und Förderinstitutionen bekennen. Führende Universitäten implementieren bereits solche neuen Qualitätskriterien und Systeme, wie z.B. für die Berücksichtigung von gesellschaftlichem Impact, interdisziplinärer Kollaboration und Diversität (z.B. UC Berkeley⁴³, CIRAD⁴⁴, Universität Gent⁴⁵). Um datengetriebene

Forschung in Zukunft optimal zu ermöglichen ist es zentral den pro-aktiven Umgang mit Daten (Nutzbarmachung, Weiter-nutzung, Pflege etc.) in der Qualitätsbewertung auf institu-tioneller sowie individueller Ebene zu berücksichtigen. Auch InterviewpartnerInnen in Österreich betonen, dass in diesem Bereich Maßnahmen gesetzt werden sollten.

Spezialisierte Fachkräfte und Fachservices

Neben generellen Grundkenntnissen bei allen Wissen-schafterInnen benötigen Forschungsinstitutionen spezia-lisierte Fachkräfte im Bereich der Data Science und der Data Stewardship. Ihr Einsatz ist aktuell bereits zu beobachten. Langfristig wird es breiter aufgestellte ExpertInnen, welche in vielen Disziplinen Unterstützung bieten können, ebenso wie auf spezielle Aufgabengebiete fokussierte ExpertInnen geben müssen, die sich z.B. in den Bereichen AI, Datenrecht oder Datenarchivierung besonders engagieren und die Wissen-schaft unterstützen. Eine große Herausforderung ergibt sich dadurch, dass nicht nur in der Wissenschaft datengetriebene Forschung einen immer höheren Stellenwert einnimmt,

sondern auch in der Wirtschaft der Bedarf an Data Scientists groß ist. Bis 2025 wird erwartet, dass sich die Anzahl der DatenexpertInnen in Europa auf der Basis des Werts von 2018 verdoppelt haben wird⁴⁶, jedoch wurde in den Interviews mehrfach betont, dass sich die Wissenschaft bewusst sein sollte, dass weiterhin mit einem massiven Mangel an Data Scientists (und verwandten Kompetenzen) in der Wirtschaft zu rechnen ist. Im Gegensatz zur Wissenschaft kann die Wirtschaft wesentlich bessere Arbeitskonditionen bieten. Somit wird sich der Wettbewerb um die besten Köpfe im Datenbereich zwischen akademischer Forschung und Wissenschaft in den nächsten Jahren weiter zuspitzen.

Die Situation in Österreich

Aktuell nimmt in Österreich die Zahl der Ausbildungsangebote und Weiterbildungsmöglichkeiten im Bereich Data Science zu. Mittlerweile bietet eine Reihe von österreichischen Universitäten (unter anderem Universität Wien, TU Wien, TU Graz, Universität Salzburg, Universität Linz, Universität Inns-bruck, WU Wien...) und Fachhochschulen (u.a. FH Technikum Wien, FH St. Pölten, FH Kufstein, FH Wiener Neustadt...) Vollzeit-sowie berufsbegleitende Bachelor-, und Masterstudiengänge sowie Weiterbildungsmöglichkeiten im Bereich Data Science an. Ein großer Teil dieser Studiengänge wurde innerhalb der letzten Jahre etabliert. Ein weiterer Ausbau der Studienplätze als auch außercurricularer Weiterbildungsmöglichkeiten ist

jedoch angesichts des großen Bedarfs insgesamt als auch des Sogs der Wirtschaft in diesem Bereich dringend nötig.

Gleichzeitig ist die Verankerung von Data Science und Data Literacy in den regulären Studienangeboten laut den befragten ExpertInnen verstärkt einzufordern: Unter den Wissen-schafterInnen fehle es an Wissen und Bewusstsein für die Möglichkeiten und Limitationen von hochqualitativer daten-getriebener Forschung und den entsprechenden Wissens-grundlagen zum professionellen Umgang mit Daten. Im inter-nationalen Vergleich sind laut den befragten ExpertInnen die Budgets für die Aus- und Weiterbildung spezifisch in diesem Bereich in Österreich zu gering.

Zu enge Bewertungskriterien für wissenschaftliche Qualität und Leistung

Wissen im Bereich der Data Literacy, des zielführenden Einsatzes von datengetriebener Forschung und dem Zurverfügungstellen von Daten sowie die konkrete Umsetzung und Anwendung hiervon wird zudem aktuell laut den ExpertInnen an österreichischen Hochschulen nicht aus-reichend incentiviert. Etwa wird die Bewertung der wissenschaftlichen Leistungen in Berufungsverfahren oder bei

der Besetzung von wissenschaftlichen Stellen an den österreichischen Hochschulen nach wie vor meist vorwiegend auf Basis bibliometrischer Daten (Impact Faktoren) und ggf. erfolgreicher Projektanträge durchgeführt. Andere Qualitäten von WissenschaftlerInnen werden in Hiring Policies und der Leistungsmessung der Universitäten und Förderungen oftmals vernachlässigt und entsprechend nicht forciert.

⁴³ Support for Faculty Search Committees. <https://ofew.berkeley.edu/recruitment/contributions-diversity/support-faculty-search-committees>

⁴⁴ Impact of research: CIRAD publishes the two guides on its ImpresS method (2018). <https://www.cirad.fr/en/news/all-news-items/articles/2018/institutionnel/impact-of-research-guides-on-the-impress-method>

⁴⁵ Evaluation of research Gent University (2018). <https://www.ugent.be/en/research/research-strategy/research-evaluation.htm#Responsibleuseofmetricsandindicators>

⁴⁶ The European Data Strategy (2020). https://ec.europa.eu/commission/presscorner/api/files/attachment/862109/European_data_strategy_en.pdf.pdf,

Daten-Services an Forschungseinrichtungen können Unterstützung bieten

Die befragten ExpertInnen betonen dabei, dass nicht jede/r WissenschaftlerIn ein/e ExpertIn in Data Science, Datenrecht, Data Management etc. sein muss. Abhilfe schaffen können in Bezug auf mangelnde Kompetenzen und Wissen hochqualitative Daten-Support- und -Serviceleistungen auf institutioneller sowie nationaler Ebene. Dies ist besonders wichtig, wenn beispielsweise viele unterschiedliche StakeholderInnen involviert sind, ethisch oder juristisch kritische Fragestellungen beantwortet werden sollen oder Disziplinen mit einem wenig entwickelten Wissensstand datengetriebene Forschung be-

treiben möchten. Vereinzelt werden solche Services in Österreich, z.B. im Bereich der Rechtsberatung bereits an Institutionen angeboten. Ein Beispiel hierfür ist die Medizinischen Universität Wien mit ihrer Daten-Clearingstelle⁴⁷. Aus den Interviewergebnissen geht aber klar hervor, dass diese Angebote insgesamt noch nicht ausreichend vorhanden sind. Solche Services sollten bedarfsorientiert institutionenübergreifend installiert werden, denn nicht jede Institution in Österreich kann ein Service aufbauen.

Positionen und Karriere innerhalb der Wissenschaft

Entsprechend des Bedarfs an Ressourcen und neuen Fähigkeiten innerhalb der Forschungsteams, aber auch der Services, werden zudem neue Positionen für diese Funktionen an den Instituten benötigt.

Dies stellt aktuell noch eine große Herausforderung dar. InterviewpartnerInnen berichten von der Problematik, an den Forschungseinrichtungen attraktive Karrieremodelle für gefragte DatenexpertInnen anbieten zu können.

⁴⁷ Daten-Clearingstelle der MedUni Wien <https://www.meduniwien.ac.at/web/ueberuns/organisation/gremien/daten-clearingstelle/>

3.3. Interdisziplinarität -

Datengetriebene Forschung überschreitet Grenzen

Multi- bzw. interdisziplinär zur Verfügung gestellte Daten können zu einem großen Erkenntnisgewinn führen. Die Formen der Zusammenarbeit und Kollaboration und somit auch interdisziplinärer Arbeit bekommen durch datengetriebene Forschung eine völlig neue Dimension. Datengetriebene Forschung ermöglicht damit nicht nur neue Formen der Zusammenarbeit und Kollaboration, sondern fordert sie auch ein.

Die Forschung der Zukunft wird sich daher strukturell ändern und wesentlich stärker als heute an Missionen orientiert und über Disziplinengrenzen hinweg arbeiten, wie dies beispielsweise im Programm Horizon Europe eingefordert wird⁴⁸. Dies betrifft nicht nur lokale Arbeitsgemeinschaften. Durch neue Netzwerke und gemeinsame Datenspeicher können weltweit verteilte

Teams an Forschungsprojekten arbeiten. Problematisch ist hierbei oftmals, dass laut ExpertInnen die einzelnen Disziplinen und Institutionen aktuell noch in sehr unterschiedlichen Entwicklungsstadien der datengetriebenen Forschung stehen und zudem kulturelle Einstellungen in Bezug auf die Bereitschaft Datensets zu öffnen oder interdisziplinär zu arbeiten innerhalb der unterschiedlichen Disziplinen stark variieren. Beispielsweise im Bereich der Meteorologie, Astronomie wie auch der Medizin sind bereits große Fortschritte und eine offenerere Kultur des Datenteilens und Zusammenarbeitens erkennbar. Die sehr unterschiedlichen Kulturen wurden sehr häufig von den befragten ExpertInnen als Hemmnis genannt. Sie erschweren die interdisziplinäre Kommunikation und Arbeit.

Die Situation in Österreich

Die befragten InterviewpartnerInnen bedauern, dass in Österreich die Möglichkeiten interdisziplinärer Zusammenarbeit, insbesondere mit großen Datensätzen, und deren Potential aktuell noch nicht ausreichend genutzt werden. In vielen Disziplinen wird noch sehr stark innerhalb bekannter Strukturen gearbeitet und zudem Daten eher selten in entsprechender Qualität zur Verfügung gestellt, damit sie auch von disziplinfremden WissenschaftlerInnen weiterverwendet werden können. Dies wird zum einen durch die schon erwähnten zu eng gefassten Metriken und nicht ausreichenden

Incentivierungsmechanismen zum Datenteilen in der österreichischen Wissenschaft verstärkt, zum anderen bewerten die ExpertInnen die Organisationsstrukturen an vielen Institutionen als zu starr und traditionell, wodurch interdisziplinäre Arbeit erschwert wird. In einzelnen Institutionen innerhalb von Österreich wurden jedoch schon erste Initiativen gesetzt, wie beispielsweise an der Universität Wien, wo die Lehre durch interdisziplinäre Erweiterungscurricula und Forschungsplattformen verbessert werden soll⁴⁹.

⁴⁸European Comission: Missions in Horizon Europe (2020). https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme/missions-horizon-europe_en

⁴⁹ Österreichischer Forschungs- und FTI-Bericht (2020). [https://www.bmdw.gv.at/dam/jcr:6e68b78f-d170-49d9-b8ba-](https://www.bmdw.gv.at/dam/jcr:6e68b78f-d170-49d9-b8ba-2e2c8a9093c3/%C3%96sterreichischer%20Forschungs-%20und%20Technologiebericht%202020_barrierefrei.pdf)

[2e2c8a9093c3/%C3%96sterreichischer%20Forschungs-%20und%20Technologiebericht%202020_barrierefrei.pdf](https://www.bmdw.gv.at/dam/jcr:6e68b78f-d170-49d9-b8ba-2e2c8a9093c3/%C3%96sterreichischer%20Forschungs-%20und%20Technologiebericht%202020_barrierefrei.pdf)

3.4. Datenstandards -

Einheitliche Vorgehensweisen machen Daten teil- und nutzbar

Um Daten teilen und insbesondere auch interdisziplinär nutzen zu können, werden gewisse allgemein anerkannte Standards und Schnittstellen für Daten aus den unterschiedlichsten Quellen benötigt. Diese müssen von allen NutzerInnen gleichermaßen verstanden werden, untereinander interoperabel, also auch miteinander kombinierbar und mithilfe von Open Source Formaten nutzbar sein. International kristallisieren sich hier die sogenannten FAIR Principles als

grundlegende Richtlinien für den Umgang mit Daten heraus, um disziplin- und länderübergreifend von der Wissenschaft als auch von Maschinen les- und bearbeitbar zu sein. Diese Prinzipien finden immer mehr Verbreitung, nicht zuletzt aufgrund von Bottom-up-Initiativen wie z.B. GO FAIR.⁵⁰ Für die Erfüllung von FAIR, insbesondere der Punkte Interoperable und Reusable, sind gemeinsame Standards (z.B. für Datenformate) unverzichtbar.

FAIR Data Principles⁵¹

FAIR Daten sind die Grundlage für disziplin- und länderübergreifende Nachnutzung von Daten. Diese wurden 2016 erstmals veröffentlicht und besagen, dass Daten folgende Eigenschaften erfüllen sollen¹:

- **Findable** (z.B. durch entsprechende Metadaten und Identifikatoren, sowie Registrierungen auf Plattformen und Repositorien)
- **Accessible** (z.B. durch einfachen Abruf und standardisierte Lizenzen zu den Nutzungsbedingungen)
- **Interoperable** (z.B. durch gemeinsame, standardisierte Datenformate und Metadaten)
- **Reusable** (durch ausreichende Beschreibung ihres Untersuchungszusammenhangs und der Entstehung)

Die Situation in Österreich

Die befragten ExpertInnen sehen es als kritisch an, dass aufgrund der ausbaufähigen Data Literacy unter den WissenschaftlerInnen in Österreich und dem Mangel an entsprechenden unterstützenden Services zu wenig Bewusstsein für die Standards für Daten und ihrer Erstellung vorhanden ist. Einheitliche Vorgehensweisen sind jedoch wichtig, um die Nachvollziehbarkeit von Daten zu gewährleisten. Die un-

zureichenden Standards sind ein Problem das oftmals bei der interdisziplinären Nutzung von Daten auftritt, da die Daten nicht interdisziplinär beschlagwortet sind und dadurch nicht von unterschiedlichen Disziplinen genutzt werden können. Nötig sind solche entsprechenden Standards auch, damit Daten in Repositorien gefunden werden können.

⁵⁰Initiative Go FAIR. www.go-fair.org

⁵¹ The FAIR Guiding Principles for scientific data management and stewardship (2016). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/>

3.5. Dateninfrastrukturen –

Technologien ermöglichen Zugänglichkeit und Wertschöpfung

Die Wissenschaft muss in Infrastrukturen zur Datenverarbeitung, Datenspeicherung und Datenübertragung investieren, um das Potenzial von datengetriebener Forschung nutzen zu können. Der zusätzliche Nutzen daraus ist, dass aufgrund von Innovationen und der permanent steigenden Leistungsfähigkeit von Technologien auch neue methodische Möglichkeiten für die Wissenschaft entstehen (z.B. durch Quantencomputing).

EOSC

Die **European Open Science Cloud (EOSC)**, die 2018 unter der österreichischen EU-Ratspräsidentschaft gestartet wurde, stellt eine zentrale, föderierte digitale Datenplattform für die europäische Forschungslandschaft dar, die europäischen ForscherInnen den einfachen und nahtlosen Zugang zu Forschungsdaten sowie relevanten Services zur Datenverarbeitung bieten soll. Dabei werden konsequent die „FAIR“ Prinzipien¹ zum Datenumgang verfolgt und sichergestellt. Somit ist die EOSC nicht nur ein zentrales Instrument, um datengetriebene Forschung in Europa zu ermöglichen, sondern auch um die europäische Forschung im internationalen Vergleich wettbewerbsfähig zu halten. Die EK betont die Rolle von datengetriebener Forschung als Vorbild für andere Arten von datengetriebenen Aktivitäten in Wirtschaft und Gesellschaft, indem Sie die Rolle der EOSC in der European Strategy for Data als Vorbild für andere „common European data spaces“ herausstreicht. Auch auf Ebene der Forschungsdisziplinen wurde bereits die Notwendigkeit gemeinsam genutzter Dateninfrastrukturen erkannt. Ein Beispiel hierfür ist die **Elixir** Initiative, ein Zusammenschluss führender europäischer Forschungseinrichtungen aus dem Life Sciences Bereich, der die gemeinsame Speicherung und Nutzung von biologischen Daten verfolgt.

International wurde dieser Investitionsbedarf von federführenden Forschungseinrichtungen bereits erkannt: So verfügt das deutsche Krebsforschungszentrum (DKFZ) über eine Speicherkapazität von aktuell siebenundzwanzig Petabyte (=27 000 Terrabyte).⁵² Das 2019 eröffnete Cambridge Centre for Data Driven Discovery (C2D3) mit vorerst fünfzig MitarbeiterInnen versteht sich als zentrale Ressource für acht führende englische Universitäten und ermöglicht institutionenübergreifend datenintensive Simulationen und Advanced Analytics⁵³. Doch im internationalen Vergleich fällt Europa zurück, beispielsweise in Bezug auf Supercomputer: Nur einer der 10 besten Supercomputer ist in Europa zu finden, dieser findet sich zudem auf Platz 9. Auch in Bezug auf Entwicklung von Quantencomputing fällt die EU im Vergleich zu China weit ab⁵⁴.

Entsprechend dem steigenden Datenvolumen werden auch Repositorien benötigt, die Daten bzw. Datensätze leicht auffindbar in den richtigen Formaten zur Verfügung stellen. In Europa wird hierfür bereits an der European Open Science Cloud (EOSC) gearbeitet, welche zusätzlich zu nutzerzentrierten Datenzugängen auch Services für die Wissenschaftslandschaft in den Mitgliedsstaaten anbietet⁵⁵. In Deutschland wird bis 2028 die Nationale Forschungsdateninfrastruktur (NFDI) aufgebaut, welche als digitaler, regional verteilter und vernetzter Wissensspeicher Forschungsdaten nachhaltig sichern und nutzbar machen soll und von Anfang an anschlussfähig an die EOSC gestaltet werden soll⁵⁶.

⁵² DKFZ: Einblick – Big Data und Krebs (2018). https://www.dkfz.de/de/presse/veroeffentlichungen/einblick/download/2018/einblick_03_2018_WEB.pdf

⁵³ Cambridge establishes new centre for data science (2019). <https://www.cam.ac.uk/news/cambridge-establishes-new-centre-for-data-science>

⁵⁴ The Digital Economy and Society Index (DESI) (2020). <https://ec.europa.eu/digital-single-market/en/desi>

⁵⁵ European Open Science Cloud EOSC. <https://www.eosc-portal.eu/>

⁵⁶ BMBF, Nationale Forschungsdateninfrastruktur (2018). <https://www.bmbf.de/de/nationale-forschungsdateninfrastruktur-8299.html>

Die Situation in Österreich

Es gibt in Österreich bereits einige Vorreiterbeispiele im Bereich der Daten-Infrastrukturen. Der Supercomputer am Vienna Scientific Cluster ist hier zu nennen. An diesem wird eine sehr hohe Rechenleistung von bis zu 2,7 PFlop/s ausgewählten österreichischen Forschungsprojekten zur Verfügung gestellt. Betrieben wird der Vienna Scientific Cluster in einer Kooperation zwischen Technischer Universität Wien, Universität Wien, Universität für Bodenkultur Wien, Technischer Universität Graz und der Universität Innsbruck⁵⁷.

Auch in Bezug auf einzelne Disziplinen gibt es bereits fortschrittliche Dateninfrastrukturen, wie etwa das Austrian Social Science Data Archive⁵⁸. Weiters gründeten bereits 2015 die Technische Universität Wien, Technische Universität Graz, Medizinische Uni Wien und das AIT den Complexity Science Hub Vienna, ein Verein mit der Zielsetzung, Big-Data Science in österreichischen Institutionen zu bündeln und „sinnvolles Wissen aus Big Data zu gewinnen“⁵⁹.

Im zweiten e-Infrastructures Austria Projekt („e-Infrastructures Austria Plus“) arbeiten seit 2017 neun österreichische Universitäten an der Vorbereitung und Implementierung digitaler Forschungsinfrastrukturen (u.a. Weiterentwicklung digitaler

Research Data Management (RDM)-Policies, Datenmanagementpläne, institutionelle Repositorien, Standards für Metadaten und Digital Object Identifier (DOI)).⁶⁰

Solche Initiativen sind Vorbilder für die gesamte österreichische Forschungslandschaft. Allerdings fehlt es laut den befragten ExpertInnen an einer nationalen Vorgehensweise und Harmonisierung solcher Initiativen ähnlich wie der geplanten deutschen Nationalen Forschungsdateninfrastruktur (NFDI). Dies wird von den ExpertInnen auch in Bezug auf die Anschlussfähigkeit an europäische Bestrebungen beobachtet. Obwohl Österreich sich hier bereits teilweise involviert, so fehlt eine durchgängige und systematische Anbindung an leistungsstarke europäische Forschungsinfrastrukturen und damit auch an Forschungsprojekte. Ein weiterer Kritikpunkt der ExpertInnen bezieht sich nicht nur auf die mangelnde Bereitstellung von Datenverarbeitung und -speicherung, sondern auch auf den Mangel an Forschungsrepositorien. Nationale, interdisziplinäre Repositorien fehlen aktuell, es fällt den österreichischen WissenschaftlerInnen laut ExpertInnenaussagen damit nicht nur schwer, ihre eigenen Daten zu speichern, sondern auch Daten aus anderen Disziplinen zu finden und zu nutzen.

Bedrohung der Unabhängigkeit durch kommerzielle Anbieter

Die österreichischen ExpertInnen verweisen auch auf die Problematik, dass WissenschaftlerInnen aufgrund mangelnder Alternativen oder der nicht vorhandenen Bewertungskompetenz auf (kostenfreie) Angebote für die Datenverarbeitung und -speicherung in der Cloud zurückgreifen, die von kommerziellen, internationalen Unternehmen angeboten werden. Dies betrifft bekannte allgemeine Cloud Anbieter sowie spezialisierte Angebote von großen wissenschaftlichen Verlagen. Dadurch droht jedoch der Kontrollverlust über die Daten und oftmals in weiterer Folge die (unerwünschte) Kommerzialisierung von Forschungsdaten. Langfristig befürchten die ExpertInnen Lock-in-Effekte, die den österreichischen Wissenschaftsstandort in eine schädliche Abhängigkeit von internationalen Anbietern bringen.

Denn es ist damit zu rechnen, dass diese Services langfristig nicht kostenfrei nutzbar sein werden und der österreichischen Forschungslandschaft keine Alternative zu diesen Services bleiben wird, wenn nicht rechtzeitig unabhängige Infrastrukturen als Alternativen angeboten werden. Im Extremfall müssen Forschungseinrichtungen ihre eigens generierten Daten teuer zurückkaufen. Auch wurde in den Interviews Bedauern darüber geäußert, dass der österreichischen Forschungspolitik ein Wissensgewinn entgeht, denn aus der Analyse der eingereichten Gesamtdaten lassen sich wertvolle Schlüsse über die nationale Forschung ziehen.

⁵⁷ Vienna Scientific Cluster. <https://vsc.ac.at/home/>

⁵⁸ AUSSDA. <https://aussda.at/>

⁵⁹ „Complexity Science Hub Vienna“ nimmt Arbeit auf (2015). <https://www.derstandard.at/story/2000019836995/complexity-science-hub-vienna-nimmt-arbeit-auf>

⁶⁰ E-Infrastructures Austria. <https://www.e-infrastructures.at/de/>

3.6. Rechtliche Rahmenbedingungen -

Sicherheit und Klarheit bilden die Grundlage datengetriebener Wissenschaft

Auch wenn Daten beliebig oft geteilt werden können, ohne an Wert zu verlieren, so ist aufgrund unterschiedlicher Aspekte von einer unregulierten Nutzung abzusehen. Daten können sensitive Informationen, insbesondere personenbezogene Informationen oder Informationen, die bei einer Veröffentlichung

geschäftsschädigend für Unternehmen wirken könnten, beinhalten. Zudem werden die Daten oftmals unter großem Ressourcenaufwand gesammelt und bereitgestellt. Bei einer unregulierten Bereitstellung würden die ErzeugerInnen der Daten nicht ausreichend entlohnt

Datenschutzgrundverordnung DSGVO

Die Datenschutzgrundverordnung, die seit nun mehr zwei Jahren in Kraft ist, stellt den weitreichenden Schutz personenbezogener Daten sicher. Gleichzeitig legt sie den rechtlichen Grundstein für den freien Datenverkehr im europäischen Binnenmarkt. Die DSGVO ist somit einerseits ein zentraler Enabler für datengetriebene Forschung, gleichzeitig wird sie oftmals als Hindernis zu Datennutzung auch im Forschungsbereich angesehen, insbesondere durch übermäßig abschreckende Wirkung und daraus folgende „Chilling Effects“.

Schutz personenbezogener Daten

Personenbezogene Daten werden in Europa seit 2018 durch die DSGVO⁶¹ geregelt, die alle EU-BürgerInnen in einer zunehmend datengestützten Welt vor Verletzungen der Privatsphäre und Datenschutzverstößen schützt und gleichzeitig einen klareren und kohärenteren Rahmen für Unternehmen schafft. Doch sind auch hier gewisse Spielräume unter Wahrung des Datenschutzes im Interesse der Wissenschaft möglich (in Österreich z.B. im Rahmen des Forschungsorganisationsgesetzes⁶²). Befragte ExpertInnen kritisieren hier jedoch eine mangelnde Harmonisierung unter den einzelnen Ländern innerhalb der EU, wie auch mit Ländern außerhalb der EU.

Lizenzen

Die Weitergabe und Verwendung von Daten kann durch die Vergabe von Lizenzen festgelegt werden. Ein Kritikpunkt der ExpertInnen betrifft dabei die mangelnden Möglichkeiten, bzw. das Fehlen einheitlicher, einfach verwendbarer Lizenzmodelle,

die es WissenschaftlerInnen erlauben, die konkreten Bedingungen der Nutzung von Daten durch andere WissenschaftlerInnen oder auch NutzerInnen außerhalb der Wissenschaft festzulegen.

Schutzrechte und Lizenzen

Prinzipiell unterliegen (Einzel-)Daten nicht dem Urheberrecht und sind somit gemeinfrei. Allerdings können Sammlungen und Zusammenstellungen von Daten in Österreich und Deutschland in manchen Fällen dennoch urheberrechtlich geschützt sein, insbesondere wenn zu ihrer Erhebung intellektuelle- und/oder kreative Leistung notwendig ist.⁶³ Lizenzen sind Befugnisse, die Dritten eingeräumt werden, die dem Rechtsinhaber zustehenden gewerblichen Schutzrechte oder urheberrechtliche Verwertungsrechte, auszuüben (Nutzungsrechte).

Diese Befugnis kann unter anderem auf folgende Nutzungsrechte erteilt werden: Patente, Gebrauchsmuster, Marken oder Geschmacksmuster⁶⁵. Neben kostenpflichtigen Modellen gibt es Lizenzen, die das kostenfreie öffentliche Einsehen, die Änderung oder Nutzung von Dritten erlauben. Darunter fallen beispielsweise sogenannte „Open Source Lizenzen“, die erlauben, dass beispielsweise Software frei nutzbar, modifizierbar und teilbar ist⁶⁶. Ein weiterer Sonderfall, der im Rahmen von Daten zu nennen ist, sind creative commons Lizenzen. Diese geben in ihren verschiedenen Ausprägungen unterschiedliche Nutzungsrechte unter unterschiedlichen Bedingungen, jedoch immer mit dem Ziel die Daten der Öffentlichkeit zur Verfügung stellen zu können⁶⁷.

⁶¹ Datenschutz-Grundverordnung (EU) 2016/679 (2016). <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

⁶² Bundesgesetz über allgemeine Angelegenheiten gemäß Art. 89 DSGVO und die Forschungsorganisation (Forschungsorganisationsgesetz – FOG) (2018). <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009514>

⁶³ Rechtliche Aspekte beim Umgang mit Forschungsdaten – Joachim Losehand – FFG (2016). https://www.ffg.at/sites/default/files/160628_ffg_forschungsdaten_losehand.pdf

⁶⁴ Wem gehören Forschungsdaten (2018). <https://www.forschung-und-lehre.de/wem-gehoeeren-forschungsdaten-1013/>

⁶⁵ Definition Lizenzen: <https://wirtschaftslexikon.gabler.de/definition/lizenz-38383>

⁶⁶ Open Source Licenses. <https://opensource.org/licenses>

⁶⁷ Creative Commons Licenses. <https://creativecommons.org/about/cclicenses/>

Die Situation in Österreich

Abschreckungseffekte wegen rechtlicher Unsicherheit beim Teilen und Nutzen von Daten, wie sie international, insbesondere innereuropäisch zu beobachten sind (sogenannte chilling effects), werden von den befragten ExpertInnen auch in Österreich gesehen. Oftmals herrscht ihrer Meinung nach Unklarheit, wie im konkreten Fall mit personenbezogenen Daten in der Wissenschaft in rechtskonformer

Weise (DSVGO) umzugehen ist. Darum sehen viele StakeholderInnen überhaupt davon ab, eigene Daten zur Verfügung zu stellen oder fremde Daten zu nutzen. ExpertInnen fordern leichter verständliche Erklärungen und Handbücher zur Rechtsauslegung, um die Weiternutzung von Daten zu ermöglichen und dabei Nutzungszwecke wie auch Rechte an Daten klar zu definieren.

3.7. Artificial Intelligence -

Die Arbeit mit Daten verändert die Wissenschaft von Grund auf

Wie kaum eine andere Technologie revolutioniert AI derzeit sämtliche Bereiche menschlichen Zusammenlebens, inklusive der Wissenschaft. Im Rahmen von datengetriebener Wissenschaft nimmt AI eine entscheidende Rolle ein. Hinsichtlich des Einsatzes von AI können dabei zwei Ebenen unterschieden werden. Einerseits kann AI so programmiert sein, dass etwa Datenanalysen innerhalb bestimmter festgelegter Grenzen, in abgeschlossenen Systemen durchgeführt werden und AI zur Unterstützung beispielsweise für die Analyse riesiger Datenbestände eingesetzt wird („schwache AI“). AI kann aber auch so gestaltet sein, dass „nur“ die Voraussetzungen für Lernprozesse programmiert werden und selbständiges Lernen ermöglicht wird. Die Gestaltung sogenannter künstlicher

neuronaler Netze ermöglicht kognitive Fähigkeiten, die es gestatten, ohne Vorgabe von Lösungsmöglichkeiten oder Zielen, neue Erkenntnisse zu entwickeln. Man spricht dabei von „starker AI“. Die Möglichkeiten und Entwicklungen sind dabei aus heutiger Sicht nur abschätzbar.

Das Spektrum von AI umfasst grundsätzlich alle Ebenen und für die Nutzung von AI haben Daten als „Grundressource“ eine zentrale Bedeutung.

Im Zuge dessen wirft AI jedoch auch diverse Fragestellungen auf, z.B. Fragen der Urhebererschaft und Intellectual Property: 2019 wurde beispielsweise das erste wissenschaftliche Textbuch, das von einer AI geschrieben wurde, unter dem Namen: „Lithium-Ion Batteries: A Machine-Generated Summary of Current Research“ veröffentlicht⁷⁰.

Wie die ExpertInnenbefragungen gezeigt haben, sind die konkreten Entwicklungen und Folgen innerhalb der Wissenschaft aktuell noch stark umstritten. In Zukunft könnten jedoch durch AI eigene Forschungsfragen aufgeworfen werden oder AI als maschinelles Teammitglied angesehen werden, welches sich zwar massiv von menschlichen Arbeitskräften unterscheidet, aber hohen Wert hat. Auch ist es gut denkbar, dass in Zukunft von AI strategische Entscheidungen getroffen werden, die direkt das Leben der Menschen beeinflussen. Diese Entwicklungen können zu einem großen Erkenntnisgewinn, einer beschleunigten und effizienteren Wissenschaft und einem geringeren Arbeitsaufwand für die WissenschaftlerInnen führen, doch ergeben sich hierdurch auch neue gesellschaftliche, juristische und ethische Fragestellungen. Diese Fragestellungen werden langfristig von der Gesellschaft und Politik gelöst werden müssen, um einen menschenzentrierten und nachhaltigen Umgang mit Daten und Algorithmen sicherzustellen, ohne dabei auf die Vorteile von AI verzichten zu müssen. Dies wurde bereits auf europäischer Ebene und in einigen Staaten erkannt, beispielsweise im Falle der Datenethikkommission in Deutschland⁷¹ oder über das Verfassen des White Paper on AI durch die EK⁶⁹.

AI auf europäischer Ebene

Die Europäische Kommission setzte 2018 die High Level Expert Group on AI (AI HLEG) als unabhängiges Expertengremium ein. Die AI HLEG veröffentlichte die „Ethics guidelines for trustworthy AI“, die StakeholderInnen dazu dient zu überprüfen, ob AI-Systeme „vertrauenswürdig“ sind. Weiters veröffentlichte die AI HLEG die „Policy and Investment Recommendations for trustworthy AI“, eine Liste von 33 Empfehlungen, die an politische EntscheidungsträgerInnen auf nationalstaatlicher und europäischer Ebene gerichtet sind und die Entwicklung und Nutzung von vertrauenswürdiger AI in Europa fördern sollen. Dabei soll einerseits sichergestellt werden, dass AI eine positive Wirkung in Europa entfalten kann, und andererseits die für AI (und somit für Daten als auch deren Grundlage) notwendigen Infrastrukturen und Skills aufzubauen⁶⁸.

Im „White Paper on AI – a European approach to excellence and trust“ baut die Europäische Kommission auf den Empfehlungen der AI HLEG auf, und ruft vor allem zu einer einheitlichen europäischen Vorgehensweise auf, um ein von Exzellenz und Vertrauen geprägtes AI-Ökosystem zu schaffen. Insbesondere wird vermehrt Forschung und Innovation im Bereich AI eingefordert, um verstärkte Synergien zwischen Forschungseinrichtungen und nationalen Standorten zu schaffen. Als Grundlage dafür verlangt die EK jedoch die konsequente Sicherstellung des Zugangs zu Daten in wiedernutzbarer Form (durch nachhaltiges Datenmanagement und Einhaltung der FAIR Prinzipien), sowie die Schaffung der notwendigen Infrastrukturen.⁶⁹

⁶⁸ High-Level Expert Group on Artificial Intelligence, „Ethics Guidelines for trustworthy AI“ (2019). <https://ec.europa.eu/futurium/en/ai-alliance-consultation>

⁶⁹ White Paper on Artificial Intelligence: a European approach to excellence and trust (2020). https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

⁷⁰ Lithium-Ion Batteries_A_Machine-Generated_Summary_of_Current_Research (2019). https://www.researchgate.net/publication/332761919_Lithium-Ion_Batteries_A_Machine-Generated_Summary_of_Current_Research

⁷¹ Gutachten der Datenethikkommission (2019). https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?__blob=publicationFile&v=6

Die Situation in Österreich

In Österreich fehlen laut den ExpertInnen aktuell oftmals die nötigen Infrastrukturen und Ressourcen, um AI entwickeln und einsetzen zu können. Die Wissenschaft ist häufig abhängig von Investitionen großer Unternehmen, z.B. durch die Weiterentwicklung und -entwicklung von in der Wirtschaft entwickelten pre-trained Netzwerken. Um solche Netzwerke allein zu entwickeln, fehlen den österreichischen Standorten die entsprechenden Ressourcen. Doch auch ein mangelnder Wissensstand wird von den befragten ExpertInnen genannt, dieser sei oftmals nicht ausreichend, um AI zielgerichtet zu nutzen. Es wird berichtet, dass viele WissenschaftlerInnen, aber auch Wirtschaftsunternehmen AI einsetzen, ohne ein konkretes Ziel zu verfolgen oder die richtigen Methoden zu nutzen, da ihnen die grundlegenden Kenntnisse dieser Technologien nicht bekannt sind.

Insgesamt beobachten die ExpertInnen, dass AI oftmals mit Angst begegnet wird und sie große Unsicherheiten hervorruft. Da viele ethische und juristische Fragestellungen in Bezug auf AI noch ungeklärt sind, wird generell der Einsatz von AI als problematisch angesehen. Doch gibt es hier erste Bestrebungen, die richtigen Rahmenbedingungen zu schaffen. So wurde 2018 die Artificial Intelligence Mission Austria 2030 veröffentlicht⁷² und bereits 2017 wurde der Österreichische Rat für Robotik und Künstliche Intelligenz⁷³ etabliert. Im Regierungsprogramm 2020-2024 bekennt sich die Bundesregierung dazu, die Rahmenbedingungen für die Entwicklung und den Einsatz von AI-Systemen und Algorithmen zu ermöglichen und betont dabei den Schutz der Menschenwürde. Unter anderem sollen relevante Beiräte der Ethik im Bereich der AI zusammengelegt werden¹⁸.

⁷² AIM AT 2030 Artificial Intelligence Mission Austria 2030 (2019).
https://www.bmkgv.at/dam/jcr:8acef058-7167-4335-880e-9fa341b723c8/aimat_ua.pdf

⁷³ Österreichischer Rat für Robotik und Künstliche Intelligenz. www.acrai.at

4. Vision: Datengetriebene Wissenschaft 2030

Im Jahr 2030 hat sich die Wissenschaft grundlegend geändert. Viele Veränderungen, die sich zehn Jahre zuvor in ihren Anfangsstadien abzeichneten oder als Wünsche in den Köpfen der WissenschaftlerInnen gebildet hatten, sind 2030 in Österreich Realität. Durch strategisches, abgestimmtes Vorgehen konnte Österreich international eine führende Rolle in datengetriebener Forschung erlangen. Die besten internationalen WissenschaftlerInnen siedeln sich darum bevorzugt in Österreich an, hier finden sie die optimalen Rahmenbedingungen für datengetriebene Wissenschaft: verfügbare Daten, eine hohe Datenkompetenz in der Wissenschaft und angepasste Belohnungssysteme.

Um die großen Probleme der Menschheit, wie z.B. den Klimawandel, zu adressieren, sind viele interdisziplinäre Kooperationen entstanden, die ihre Forschungsergebnisse in Echtzeit teilen und auf diese Weise zur Entwicklung konkreter Lösungen für die Gesellschaft beitragen. Es haben sich große Repositorien unter Beteiligung mehrerer Forschungseinrichtungen gebildet, von denen WissenschaftlerInnen jederzeit und von überall aus Zugriff auf Daten unterschiedlicher Herkunft haben und zusammenarbeiten können. Das sichere Zusammenführen von Datenbeständen des öffentlichen Sektors, der Wissenschaft und Wirtschaft unter Wahrung von Datenschutz und Eigentümernrechten ist zu einem Alleinstellungsmerkmal für den österreichischen Forschungsstandort geworden.

Leistungsstarke Supercomputer ermöglichen WissenschaftlerInnen in Österreich die Bewältigung großer Datenmengen. Insbesondere AI hat mittlerweile Forschungsprozesse revolutioniert. AI wird als maschinelles Teammitglied in Forschungsgruppen gesehen. Zahlreiche Datensets und Publikationen wurden von AIs veröffentlicht und seit einigen Jahren ent-

wickeln sie nicht nur selbst neue Forschungsfragen und -hypothesen, sondern treffen auch strategische Entscheidungen. Durch allumfassende Sensorik, sozusagen einer „Vermessung der Welt“, entstand eine neue Empirie, die es erlaubt, in allen Bereichen der Wissenschaft daten- und evidenzbasiert zu arbeiten. Hierdurch wurden die Grenzen zwischen den Disziplinen immer stärker verwischt und durchlässiger, wenngleich klar abgrenzbare Themenfelder bearbeitet werden. Auch in der akademischen Ausbildung haben sich klassischen Disziplinengrenzen in Österreich bereits aufgelöst, denn unter dem neuen Forschungsparadigma der datengetriebenen Wissenschaft dominiert bereichsübergreifendes Lernen.

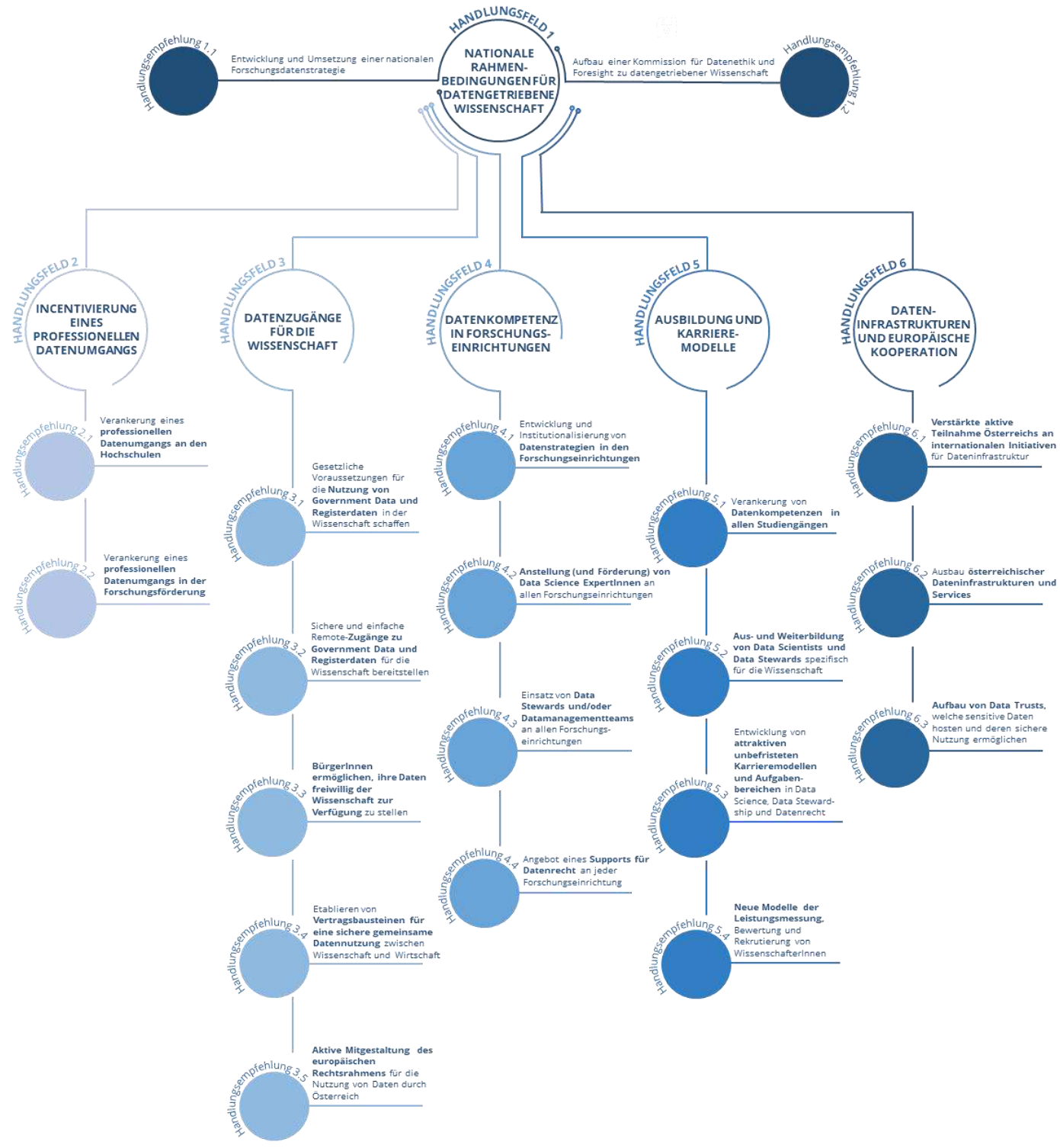
Daten wurden somit endgültig zum Zentrum wissenschaftlichen Handelns, was sich auch in den Wertigkeiten innerhalb des wissenschaftlichen Systems niederschlägt: Längst wurde erkannt, dass der Wert gut gepflegter und kontinuierlich gewarteter Datensätze den Wert von wissenschaftlichen Einzelerkenntnissen in Form von Publikationen weit übersteigt, was sich auch in neuen gelebten Definitionen wissenschaftlicher Qualität und der Belohnung eines professionellen Umgangs mit Daten im Karriereverlauf niederschlägt. Durch die Berücksichtigung vielfältiger Aspekte in Definitionen wissenschaftlicher Qualität konnte Österreich herausragende WissenschaftlerInnen aufbauen, die durch unkonventionelle Arbeitsweise bahnbrechende Entdeckungen machen. Da der Großteil des Forschungsprozesses ohnehin digital passiert, sind räumliche Grenzen weniger wichtig geworden. Es wird international gemeinsam und projektbasiert an den großen Herausforderungen und Missionen der Menschheit geforscht. Globale Pandemien und der Klimawandel haben der Menschheit schon vor Jahren gezeigt, dass kleinere Wissenschaftsstandorte von einer Vernetzung überproportional profitieren können. Dies gilt insbesondere für Österreich.

5. Handlungsempfehlungen

Die Analyse der Trends, Anforderungen und des Status quo in Österreich hat gezeigt, wo aktuell Defizite in der österreichischen Forschungslandschaft bezüglich der datengetriebenen Wissenschaft bestehen, die Vision hat ein Bild gezeichnet, wohin sich Österreich und die internationale Wissenschaft entwickeln können, wenn sie die Herausforderungen und Chancen der datengetriebenen Wissenschaft richtig nutzen. So stellt

sich nun die Frage, wie dies konkret umgesetzt werden kann. Hierfür wurden **sechs verschiedene Handlungsfelder** identifiziert, mit insgesamt **20 untergeordneten Handlungsempfehlungen**, welche bei einer idealen Umsetzung dazu beitragen können, Österreich und die österreichische Wissenschaft auf das Jahr 2030 vorzubereiten.

Abbildung 5 Handlungsfelder und Empfehlungen für datengetriebene Wissenschaft



5.1 Handlungsfeld 1:

Nationale Rahmenbedingungen für datengetriebene Wissenschaft

Die stetig anwachsenden Datenbestände und steigende Bedeutung von datengetriebener Forschung erfordern mehr Aufmerksamkeit und eine grundsätzliche strategische Auseinandersetzung auf allen Ebenen der Forschung, des Forschungsmanagements und der Forschungspolitik.

Damit für die Wissenschaft in Österreich Daten in ausreichender Qualität und Quantität zur Verfügung stehen und zielgerichtet eingesetzt werden können, sind definierte Zielsetzungen und unterstützende Rahmenbedingungen auf nationaler Ebene nötig.

Handlungsempfehlung 1.1:

Entwicklung und Umsetzung einer nationalen Forschungsdatenstrategie

Zielsetzungen:

- Festlegung gemeinsamer nationaler Zielsetzungen im Umgang mit Daten
- Identifikation prioritärer Kooperationsbereiche und Bündelung von Ressourcen zwischen Forschungseinrichtungen /Universitäten
- Ausbau von Dateninfrastrukturen und Services für WissenschaftlerInnen

Wirkung:

- > Strategisches nationales Vorgehen und dadurch Erhöhung der Wettbewerbsfähigkeit Österreichs bei datengetriebener Wissenschaft
- > Verhinderung von Insellösungen in einzelnen Bereichen und Institutionen
- > Effizientere Nutzung von Forschungsgeldern und Generierung von gestiegenem Mehrwert für den gleichen Einsatz
- > Stärkung der Unabhängigkeit der österreichischen Wissenschaft von internationalen kommerziellen Speicher- und Rechenleistungsanbietern und Plattformen

Maßnahme:

Auf Bundesebene soll eine **nationale Forschungsdatenstrategie** unter Führung des Bundesministeriums für Bildung, Wissenschaft und Forschung und unter Einbeziehung aller relevanter StakeholderInnen erarbeitet werden. Diese soll neben

übergeordneten Zielsetzungen folgende Teilbereiche beinhalten, wobei stets auch die europäische Anschlussfähigkeit und Zukunftsplanung miteinbezogen werden muss:

Dateninfrastruktur – Ein gemeinsames Vorgehen und Strategie bezüglich Infrastrukturen und Services schafft Synergien

Forschungsdaten sind ein wertvolles Gut, welches aktuell jedoch oftmals nicht vollends ausgenutzt wird. Lokal gespeicherte Daten gehen häufig verloren, indem sie nicht in entsprechende Datenspeicher und -repositorien hochgeladen und geteilt werden. Daher soll eine nationale Forschungsdatenstrategie unter anderem bewirken, **Einzelinitiativen, einzelne Infrastrukturen und Repositorien zu erfassen, zu bündeln** und wo nötig und sinnvoll **gemeinsame nationale Forschungsdateninfrastrukturen und -**

services aufzubauen. Dies erfordert sowohl einen bedarfsorientierten Ausbau der Services auf institutioneller Ebene als auch eine Gesamtkoordination auf nationaler Ebene, um Synergien zu schaffen und **international anschlussfähig** zu sein. Nationale Dateninfrastrukturen sollen u.a. auch als Alternativen zu kommerziellen Anbietern von Datenspeichern und Services dienen und entsprechende Exitstrategien aus kommerziellen Angeboten für Forschungsinstitutionen entwickelt werden.

Datenmanagement – Einheitliche Standards ermöglichen die gemeinsame Arbeit an Daten und einen erhöhten Erkenntnisgewinn

Zusätzlich soll die nationale Forschungsdatenstrategie übergreifende Zielsetzungen und Vorgehensweisen in allen Forschungsdisziplinen und -institutionen bezüglich des Öffnens und Nutzbarmachens von Daten entwickeln. Hierunter fallen u.a. Vorgehensweisen zur Entwicklung und zum Einsatz von maschinenverarbeitbaren Datenmanagementplänen und Lizenzmodellen (siehe auch *Handlungsempfehlung 3.4 Etablieren von Vertragsbausteinen für eine sichere gemeinsame Datennutzung zwischen Wissenschaft und Wirtschaft* sowie *Hand-*

lungsempfehlung 3.5 Aktive Mitgestaltung des europäischen Rechtsrahmens für die Nutzung von Daten durch Österreich) sowie disziplinübergreifendes Vorgehen für die Erstellung von offenen Metadaten, damit Daten für unterschiedliche Nutzer auffindbar sind. Daten sollen, soweit möglich, offengelegt werden und mit Open Source Software genutzt werden können. Grundsätzlich gilt im Umgang mit Daten **so interdisziplinär und international wie möglich zu arbeiten und so lokal und disziplinspezifisch wie nötig**.

Datennutzung und -zugang: Datengetriebene Forschung führt auch im Strategieentwicklungsprozess zu Vorteilen

Unterschiedliche Erkenntnisse hinsichtlich Datenverfügbarkeit und -nutzung in Forschung und Innovation in Österreich, etwa von Bundesministerien, Forschungsinstitutionen und Fördergebern, sollen in die Vorbereitung und Begleitung des Strategieentwicklungsprozesses einfließen bzw. im Vorfeld erhoben werden. Auf diese Weise kann die Entwicklung der Forschungs-

strategie evidenzbasiert und ressourceneffizient gestaltet und **strategische Entscheidungen für die nationale Forschungsdatenstrategie datenbasiert** getroffen werden. Eine datenbasierte Steuerung der Forschungspolitik wird bereits im aktuellen Regierungsprogramm gefordert¹⁸.

Handlungsempfehlung 1.2:

Aufbau einer Kommission für Datenethik und Foresight zu datengetriebener Wissenschaft

Zielsetzungen:

- Laufendes Monitoring der raschen technologischen Weiterentwicklung und des Einsatzes von Daten in Wissenschaft und Forschung
- Erarbeitung von Empfehlungen für die österreichische Politik und Forschungslandschaft

Wirkung:

- > Stärkung des Bewusstseins und Bereitstellung von Informationen über zukünftige Entwicklungen im Bereich datengetriebener Forschung, AI und Algorithmen bei EntscheidungsträgerInnen in Politik und Wissenschaft
- > Ermöglichen des frühzeitigen Antizipierens kommender Entwicklungen für alle AkteurInnen
- > Bewusste Begleitung der ethischen und juristischen Auswirkungen von AI auf die Gesellschaft

Maßnahme:

Neue technologische Entwicklungen im Bereich der datengetriebenen Forschung werden in Zukunft die Art wie Forschung betrieben wird noch stärker als bisher verändern und teils disruptive Auswirkungen entfalten. Diese Entwicklungen sollen durch eine **eigenständige Kommission für Datenethik und**

Foresight zu datengetriebener Wissenschaft (mit operativem Teil in Form einer dauerhaften Geschäftsstelle als auch einem hochkarätigen Beratungsgremium) **sorgfältig und laufend analysiert und antizipiert** werden.

Foresight und Innovation strategisch nutzen, um die Zukunft der österreichischen Forschung zu sichern

Durch institutionalisierte Foresight- und Innovationsprozesse sollen **Empfehlungen und Lösungen für die österreichische Politik** zur frühzeitigen Vorbereitung auf Veränderungen erarbeitet werden. Als Vorbild hierfür kann das britische Centre for Data Ethics and Innovation⁷⁴, wie auch das Centre for

Strategic Futures⁷⁵ in Singapur dienen. Dieser Prozess soll **laufend und langfristig aufgesetzt** sein, um nicht den Anschluss an die internationale Forschungsumgebung zu verlieren.

Datengetriebene Forschung wird in Zukunft auf viele Bereiche Auswirkungen haben, diese gilt es zu antizipieren

Folgende Aufgaben sollen durch die Kommission für Datenethik und Foresight zu datengetriebener Wissenschaft bearbeitet werden:

- **Laufende Beobachtung der technologischen Entwicklungen** mit Implikationen für datengetriebene Wissenschaft
- **Definition von neuen Rollenverteilungen, Aufgaben und Verantwortlichkeiten** z.B. zwischen menschlichen WissenschaftlerInnen und Algorithmen
- Frühzeitige Ausrichtung des **notigen Infrastrukturausbaus und der nötigen technologischen Investitionen** auf aufkommende technologische Disruptionen

- **Beobachtung und Berücksichtigung von ethischen Herausforderungen**, die sich auf die Gesellschaft auswirken
- **Erarbeitung von Möglichkeiten AI Technologien** gezielt und unter den richtigen Bedingungen zur Lösung der großen gesellschaftlichen Probleme einzusetzen
- **Entwicklung juristischer Grundsätze**, um ethischen Implikationen entsprechend Gesetze und Regelungen zu etablieren (bspw. Regelung, wer für AI haftet und wie die Qualität und Fairness von Algorithmen geprüft werden kann).

Um die Zukunft zu analysieren werden ExpertInnen aus vielen Bereichen gebraucht

Aufgrund dieser Vielzahl an Aufgaben sollen in der entsprechenden Kommission für Datenethik und Foresight zu datengetriebener Wissenschaft diverse Perspektiven und Expertisen inkludiert sein: **Rechtswissenschaften, Ethik,**

Soziologie, IT, Innovationsforschung, Data Science sowie Vertretungen der diversen Forschungsdisziplinen und der Gesellschaft sollen dort laufend eingebunden sein.

⁷⁴ Centre for Data Ethics and Innovation UK.
<https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation>

⁷⁵ Centre for Strategic Futures Singapore. <https://www.csf.gov.sg/>

5.2 Handlungsfeld 2:

Incentivierung eines professionellen Datenumgangs

In der aktuellen Bewertung von Forschung findet ein guter Umgang mit und das Zugänglichmachen von Daten wenig Beachtung und wird deshalb kaum durchgeführt. Nur wenn WissenschaftlerInnen und Forschungsinstitutionen auch entsprechend incentiviert werden, professionell mit Daten

umzugehen (z.B. Forschungsdaten kontinuierlich zu pflegen, diese zu teilen und für andere nutzbar machen), ist damit zu rechnen, dass sich die Praktiken im Wissenschaftssystem ändern und wertvolle Daten in vermehrtem Ausmaß zur Verfügung stehen.

Handlungsempfehlung 2.1:

Verankerung eines professionellen Datenumgangs an den Hochschulen

Zielsetzungen:

- Entwicklung und Implementierung innovativer Metriken zum Umgang mit Forschungsdaten auf institutioneller Ebene, als integrativer Bestandteil von Leistungsvereinbarungen von Hochschulen und Forschungsinstituten sowie auf individueller Ebene als zusätzliche Bewertungskriterien in Karrierevereinbarungen bzw. Berufungsverfahren an Hochschulen (vgl. dazu *Handlungsempfehlung 5.4 Neue Modelle der Leistungs-messung, Bewertung und Rekrutierung von WissenschaftlerInnen*)
- Umsetzung der nationalen Forschungsdatenstrategie und Entwicklung eigener Datenstrategien innerhalb dieses Rahmens an den Hochschulen
- Erhebung des Investitionsbedarfs für datengetriebene Forschung an Hochschulen

Maßnahme:

Um einen fortschrittlichen Umgang mit Daten und datengetriebener Forschung an Österreichs Hochschulen und Forschungseinrichtungen zu incentivieren, sollen die **Leistungsvereinbarungen der Hochschulen entsprechend angepasst werden** (z.B. über gewisse Anteile

Wirkung:

- > Incentivierung von Aktivitäten der Universitäten im Bereich datengetriebener Forschung (siehe auch *Handlungsfeld 4: Datenkompetenz in Forschungseinrichtungen* sowie *Handlungsfeld 5: Ausbildung und Karrieremodelle*)
- > Erhöhung des Verantwortungsgefühls und Bewusstsein unter EntscheidungsträgerInnen in den Forschungseinrichtungen für die Wichtigkeit und den Wert von Forschungsdaten für die österreichische Forschung sowie die Notwendigkeiten, Ressourcen für datengetriebene Forschung bereitzustellen

Neue Kriterien sollen den richtigen Datenumgang an Hochschulen stärker incentivieren

Hierfür soll das Bundesministerium für Bildung, Wissenschaft und Forschung **neue Kriterien für einen professionellen Datenumgang** an Hochschulen einfördern und über die **Leistungsvereinbarungen steuern**. Hierunter kann beispielsweise die Entwicklung neuer Bewertungskriterien für die Arbeit von WissenschaftlerInnen fallen, welche alternativ oder zusätzlich zu bibliometrischen Analysen und Bewertungen für die Karriereentwicklung herangezogen werden sollen oder auch Kriterien für einen nachhaltigen Umgang in der Erstellung und Speicherung von Forschungsdaten (siehe auch *Handlungsempfehlung 4.1 Entwicklung und Institutionalisierung von Datenstrategien in den Forschungseinrichtungen*).

Auf diese Weise soll das Implementieren der nationalen Forschungsdatenstrategie, die entsprechende **Nutzung und Anbindung an**

von Open Data Veröffentlichung, Beteiligung an Dateninfrastrukturen, Auswahlkriterien von akademischem Personal, Datenstrategien und Standards zum Umgang mit Daten).

Dateninfrastrukturen, aber auch ein fortschrittliches **Forschungsdatenmanagement** und die **Bereitstellung von hochqualitativen Daten, Softwarecodes und Algorithmen entsprechend der nationalen Standards innerhalb der Institutionen umgesetzt werden**.

Insbesondere die Schaffung **neuer Stellen und Verantwortlichkeiten** und die **Aus- und Weiterbildung im Bereich der datengetriebenen Forschung** sollen in den Leistungsvereinbarungen verankert und den Institutionen damit die Verantwortung über die nötigen Kompetenzen verdeutlicht werden. Auch die **Umsetzung der DORA Erklärung**⁷⁶ an den Hochschulen sollte in den Leistungsvereinbarungen positiv bewertet werden.

Die Informationsbasis zu notwendigen Investitionsbedarfen für datengetriebene Forschung fehlt derzeit

Datengetriebene Forschung erfordert eine effektive Neugestaltung **des Ressourceneinsatzes für den notwendigen Ausbau von Infrastrukturen und zur Hebung des digitalen Kompetenzlevels an den Hochschulen**. Informationen zur Finanzierung des Status quo sowie zu den Investitionsbedarfen der Institutionen fehlen jedoch derzeit weitgehend. In einem ersten Schritt soll hierfür Wissen über die eingesetzten Ressourcen und Kosten der bestehenden

Dateninfrastrukturen an den Hochschulen und Forschungsinstitutionen in Österreich generiert werden. In Folge sollten die Bedarfe zur Schaffung internationaler Levels in datengetriebener Forschung erhoben werden und Richtwerte für notwendige Investitionen an den Institutionen erarbeitet und kommuniziert werden⁷⁶, um entsprechende Umschichtungen, und wenn nötig Erhöhungen der vorhandenen Budgets zu ermöglichen.

76 Disruption oder Evolution? Systemische Rahmenbedingungen der Digitalisierung in der Hochschulbildung (2019). https://www.hof.uni-halle.de/web/dateien/pdf/ab_109.pdf

Handlungsempfehlung 2.2:

Verankerung eines professionellen Datenumgangs in der Forschungsförderung

Zielsetzungen:

- Weiterentwicklung und Anpassung der Forschungsförderung in Hinblick auf einen professionellen Datenumgang
- Verankerung von Anreizstrukturen in der Forschungsförderung

Wirkung:

- > Erhöht den Stellenwert eines fortschrittlichen Umgangs mit Forschungsdaten bei WissenschaftlerInnen (z.B. Bereitstellen, Pflegen und Teilen von Forschungsdaten)
- > Ermöglicht Daten langfristig und in entsprechender Qualität zur Verfügung zu stellen und steigert den Nutzen, der aus einer Förderung gezogen werden kann

Maßnahme:

Ein fortschrittlicher Umgang mit Forschungsdaten nimmt **Zeit und Ressourcen in Anspruch, langfristig wird damit jedoch der Impact von Forschung nachhaltig gesteigert**. Entsprechend soll ein fortschrittlicher Umgang in Bezug auf datengetriebene Wissenschaft in den Programmen der **Forschungsförderung verstärkt gefördert und gefordert**

werden. Teilweise geschieht dies bereits, so fordert der FWF seit 2019 verpflichtend Datenmanagementpläne und fördert auch anfallende Kosten³⁴. Forderungen nach- und Förderung von fortschrittlichem Umgang mit Daten (wie z.B. der langfristigen Archivierung, Pflege und Öffnung) gilt es zu stärken und auszuweiten.

Fortschrittlicher Datenumgang soll in bestehende Förderprogramme aufgenommen werden

In einem ersten Schritt sollen, basierend auf einer Analyse des Status quo, ein fortschrittlicher Umgang mit Daten (so weit als möglich und sinnvoll) Teil aller Forschungs- und Innovationsprojekte werden, indem **Qualitätsanforderungen an Forschungsprojekte und Evaluierungsprozesse (durch entsprechende Kennzahlen) weiterentwickelt** werden. Dabei ist zu berücksichtigen, dass der bürokratische Aufwand

für die FördernehmerInnen und FördergeberInnen nicht unverhältnismäßig gesteigert werden soll. Insbesondere soll die **Erstellung von maschinenverarbeitbaren Datenmanagementplänen**, welche den gesamten Daten-Life-Cycle berücksichtigen (von der Datengenerierung bis zum Umgang mit den Daten nach Abschluss des Projektes), bei geförderten Projekten vorgegeben werden.

Die Veröffentlichung von Datensätzen erhöht den Nutzen von Förderungen

Die **Veröffentlichung und Nutzbarmachung von hochqualitativen Datensätzen und Algorithmen als Open Access und Open Data**, auch abseits von Publikationen und positiven Resultaten, soll im Rahmen der Forschungsförderung, etwa

durch einen Bonus, belohnt werden. Vor allem sollen auch die notwendigen Humanressourcen (z.B. Data Scientists und Data Stewards) sowie die entsprechenden Infrastrukturen langfristig finanziert werden.

Steigerung der Datenkompetenz in der Begutachtung von Förderanträgen ist nötig, um neue Evaluierungskriterien effektiv umzusetzen

Um diese neuen Kriterien entsprechend zu bewerten, soll die **Datenkompetenz im Gutachtersystem gestärkt** werden. Denn besonders neue Metriken im Bereich der datengetriebenen Forschung sind aktuell schwer zu gewichten. Zentral ist hier die **Weiterbildung von GutachterInnen bezüglich**

ihrer Datenkompetenz. Unterstützend müssen Lösungen gefunden werden, um den Begutachtungsaufwand für (datengetriebene) Forschungsprojekte nicht unverhältnismäßig zu steigern.

Neue Datenzugänge zu erschaffen und zu erhalten, soll entlohnt werden

Es sollen **neue Förderprogramme entwickelt** werden, die konkret das Wirkungsziel verfolgen, einen fortschrittlichen Datenumgang und die Bereitstellung von Daten zu fördern. Beispielsweise können die Aufarbeitung und Öffnung bereits lokal bestehender wertvoller Daten gefördert werden, oder

auch das Einpflegen wichtiger Informationen in Datenbanken, die aktuell nicht digital zur Verfügung gestellt werden. Auch sind Förderprogramme vorstellbar, die ForscherInnen nicht nur monetär fördern, sondern auch durch die Bereitstellung von schwer beziehbaren oder teuren Datenbeständen.

Handlungsempfehlung 3.1:

Gesetzliche Voraussetzungen für die Nutzung von Government Data und Registerdaten in der Wissenschaft schaffen

Zielsetzungen:

- Beseitigung der nationalstaatlichen rechtlichen Barrieren für die wissenschaftliche Arbeit mit unterschiedlichen Daten der öffentlichen Hand
- Schaffung von rechtlichen Standards für Datenzugänge zu Government Data inkl. Registerdaten unter Wahrung von Persönlichkeitsrechten und Datenschutz
- Breite Umsetzung der EU-Richtlinie zu offenen Daten und der Weiterverwendung von Informationen des öffentlichen Sektors in Österreich (Open Data und PSI-Richtlinie⁷⁷)

Wirkung:

- > Steigerung des wissenschaftlichen Erkenntnisgewinns durch die Nutzung von Government Data und Registerdaten
- > Stärkung der internationalen Wettbewerbsfähigkeit des österreichischen Forschungsstandortes
- > Steigerung der gesellschaftlichen Wirkung von Wissenschaft in Österreich durch die Nutzung relevanter Datenbestände

Maßnahme:

Daten aus dem öffentlichen Sektor sind ein zentraler „Rohstoff“ für viele Forschungsdisziplinen. Es sollen daher mehr **Govern-**

ment Data und Registerdaten der Wissenschaft in Österreich zur Verfügung gestellt werden.

Government Data und Registerdaten bieten großes Potenzial, welches gehoben werden sollte

Der Zugang der Wissenschaft zu ohnehin vorhandenen Datenbeständen der öffentlichen Hand (z.B. Daten des Melderegisters, der Sozialversicherungen, der Standesämter und medizinische Datenbanken, der Gesundheitsakte ELGA oder auch der Forschungsförderungen³³) wird in Österreich im internationalen Vergleich sehr restriktiv gehandhabt und ist je nach Materie in unterschiedlichen Gesetzen geregelt. Auch die Schaffung einer Forschungsförderungsdatenbank, die wertvolle wissenschaftliche Analysen zum österreichischen Forschungsstandort ermöglicht hätte, war zwar im Regierungsprogramm 2017-2022 enthalten, wurde jedoch bisher nicht umgesetzt⁷⁸. Das **Forschungsorganisationsgesetz** beinhaltet zwar seit 2018 einen Rahmen für die Nutzung persönlicher Daten durch die Wissenschaft, allerdings **fehlen die entsprechenden Verordnungen der zuständigen Ministerien**, um die Daten tatsächlich in der Wissenschaft nutzen zu können. Deshalb sollen jene Verordnungen der Ministerien erlassen werden, welche einen faktischen Zugang für die Wis-

senschaft zu verschiedenen, von Bundesbehörden erhobenen Datenbeständen ermöglichen. Darüber hinaus sollen relevante Gesetze (z.B. das Bundesstatistikgesetz, deren Novelle seit einiger Zeit diskutiert wird) so konkret und verbindlich wie möglich gestaltet werden.

Es sollen rechtliche **Standards und Regelungen geschaffen** werden, welche den **Zugang und die Nutzung** von unterschiedlich geregelten öffentlichen Daten für die Wissenschaft ermöglichen sowie vereinheitlichen. Derzeit existiert hierzu ein gewisser Ermessensspielraum bei den relevanten Stellen (z.B. Statistik Austria), der jedoch oftmals die konkrete Zurverfügungstellung relevanter Daten verhindert. Um Government Data in der Wissenschaft besser nutzen zu können, ist es weiters nötig, die **EU-Richtlinie zu offenen Daten und der Weiterverwendung von Informationen des öffentlichen Sektors (Open Data und PSI-Richtlinie) in Österreich möglichst breit und forschungsfreundlich** umzusetzen.

⁷⁷ Richtlinie (EU) 2019/1024 über offene Daten und die Weiterverwendung von Informationen des öffentlichen Sektors (2019). https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=uriserv:OJ.L_.2019.172.01.0056.01.DEU&toc=OJ.L.2019:172:TOC

⁷⁸ Zusammen. Für unser Österreich. Regierungsprogramm 2017-2022 https://www.oeh.ac.at/sites/default/files/files/pages/regierungsprogramm_2017-2022.pdf

Vereinfachung rechtlicher Regelungen, um Unsicherheiten zu vermeiden und zugleich Persönlichkeitsrechte zu wahren

Wesentlich ist, dass rechtliche **Vorgaben für WissenschaftlerInnen vereinfacht** werden und **praktisch durchführbar** sind. Dabei muss insbesondere der Schutz der Persönlichkeitsrechte, der ein starkes europäisches Asset ist, weiter gewährleistet werden. Daher sollen **Möglichkeiten unterschiedlicher „Schutzniveaus“ für verschiedene Arten von Daten** aus öffentlicher Hand evaluiert und eingeführt werden. Um hier Optionen in einem sicheren Rahmen zu erproben, können **„Regulatory Sandboxes“**, abgeschlossene Experimentierräume

zur Erprobung neuer gesetzlicher Rahmenbedingungen für konkrete, spezifische wissenschaftliche Fragestellungen (z.B. im Bereich Nutzung von Gesundheitsdaten für wissenschaftliche Forschung), für einen erhöhten und schnelleren Erkenntnisgewinn genutzt werden, wie es beispielsweise im Bereich der Finanzmarkt-aufsicht in Österreich bereits durchgeführt wird ⁷⁹

Handlungsempfehlung 3.2:

Sichere und einfache Remote-Zugänge zu Government Data und Registerdaten für die Wissenschaft bereitstellen

Zielsetzungen:

- Schaffung eines Datenzentrums und einer entsprechenden Rechtsgrundlage, welche der Wissenschaft Zugriff zu Government Data und Registerdaten bieten
- Erweiterung und Umsetzung des bereits im Regierungsprogramm vorgesehenen „Austrian Micro Data Centers“

Wirkung:

- Erhöhung des Erkenntnisgewinns durch bessere Zugänge zu öffentlichen Daten und Registerdaten
- Stärkung des österreichischen Forschungsstandortes
- Veredelung von Daten aus öffentlicher Hand durch Weiterverarbeitung in der Forschung

Maßnahme:

Um die Nutzbarkeit der zur Verfügung gestellten Government Data und Registerdaten zu erhöhen, sollen **Mikrodaten des Bundes für die Wissenschaft** über ein **Portal mit hoher**

Usability und klaren Prozessen remote zugänglich und nutzbar gemacht werden.

Nutzbarkeit der Daten und hohe Usability sind Schlüsselfaktoren für neue Datenplattformen

Ausschlaggebend für den Erfolg eines solchen Portals ist die **Übersichtlichkeit über die vorhandenen Daten**. Zudem soll es möglich sein, bei einem solchen Portal **neue, bisher nicht aktiv der Wissenschaft angebotene Datensets** anzufordern, damit diese in Zukunft nach Möglichkeit ebenso der Wissenschaft zur Verfügung gestellt werden.

Diese Plattform kann durch die konsequente Umsetzung des bereits im Regierungsprogramm¹⁸ vorgesehen „Austrian Micro

Data Center“ geschaffen werden, welches mit der bereits vorhandenen Plattform „data.gv.at“ verknüpft werden könnte. Als Vorbild kann das Portal „Statistics Denmark“⁸⁰ dienen. Denkbar ist auch eine Verknüpfung mit den in *Handlungsempfehlung 6.3 Aufbau von Data Trusts, welche sensitive Daten hosten und deren sichere Nutzung ermöglichen* beschriebenen Data Trusts.

⁷⁹ Bundesministerium Finanzen, Regulatory Sandbox (2020). <https://www.bmf.gv.at/rechtsnews/finanzmaerkte-rechtsnews/regulatory-sandbox.html>

⁸⁰ Statistics Denmark. <https://www.dst.dk/en>

Handlungsempfehlung 3.3:

BürgerInnen ermöglichen, ihre Daten freiwillig der Wissenschaft zur Verfügung zu stellen

Zielsetzungen:

- Aufbau von Prozessen und einer Plattform, die es BürgerInnen ermöglicht, ihre Daten freiwillig der Wissenschaft zur Verfügung zu stellen

Wirkung:

- > Stärkung des Bewusstseins zu Wert sowie Nutzung persönlicher Daten in der breiten Bevölkerung
- > Stärkung von Recht und Kultur der informationellen Selbstbestimmung für BürgerInnen
- > Erweiterung der Datenzugänge zu wertvollen personenbezogenen Daten für die österreichische Forschung

Maßnahme:

Persönliche Daten sind durch die Datenschutzgrundverordnung (DSGVO) in Europa⁶¹ umfassend geschützt. Dennoch sind sie eine potenziell immens wertvolle Datenquelle für die Forschung. Prinzipiell besteht die Möglichkeit für BürgerInnen, personenbezogene Daten freiwillig der Wissenschaft zur Verfügung zu stellen. Es fehlt jedoch derzeit eine praktische und einfache Möglichkeit für BürgerInnen, diese **Erlaubnis zu erteilen, sowie an einer Plattform, um diese Daten zur Verfügung zu stellen**. Dies ist beispielsweise auf der amerikanischen Plattform openhumans.org bereits der Fall, auf welcher BürgerInnen verschiedene Daten der Wissenschaft zur Verfügung stellen können, darunter befinden sich DNA Testergebnisse, Verläufe von Suchanfragen in Onlinesuchma-

schinen oder den Verlauf von getrackten persönlichen Standorten⁶¹. In Europa ist eine ähnliche Entwicklung (insbesondere seit der Corona-Pandemie) in Bezug auf Gesundheitsdaten sichtbar, so haben innerhalb kürzester Zeit über 500.000 Personen über die „Datenspende“-App des Robert Koch-Instituts⁶² ihre Informationen aus Fitnesstrackern der Wissenschaft zur Verfügung gestellt⁶³. In Bezug auf medizinische Daten forderte bereits 2017 der deutsche Ethikrat die Möglichkeit für Bürger „...die umfassende Nutzung ihrer Daten für die medizinische Forschung zu erlauben“⁶⁴. Auch die kassenärztliche Bundesvereinigung in Deutschland schlägt vor, den BürgerInnen die Option einer freiwilligen und strukturierten Datenspende ihrer Gesundheitsdaten als Bürgerservice anzubieten⁶⁵.

Durch die freiwillige Zurverfügungstellung von Daten der Bevölkerung wird hochqualitative Forschung erleichtert

Es soll die Möglichkeit geschaffen werden, dass **BürgerInnen freiwillig personenbezogene Daten der Forschung zur Verfügung stellen** (zum Beispiel im Bereich der Gesundheit). Hierfür sollen entsprechende **sichere Prozesse** und eine **Plattform** aufgebaut werden. Auf dieser soll es möglich sein, die Daten durch sichere Verschlüsselungsverfahren zu anonymisieren oder zu pseudonymisieren (siehe hierzu auch *Handlungsempfehlung 6.3 Aufbau von Data Trusts, welche sensitive Daten hosten und deren sichere Nutzung ermöglichen*) und dabei

stets die Datenhoheit über die eigenen Daten zu behalten. Eine Umsetzung dieser Plattform ist als eigenständiges Projekt, jedoch auch als Erweiterung des „Austrian Micro Data Center“ (siehe auch *Handlungsempfehlung 3.2 Sichere und einfache Remote-Zugänge zu Government Data und Registerdaten für die Wissenschaft bereitstellen* sowie *Handlungsempfehlung 6.3 Aufbau von Data Trusts, welche sensitive Daten hosten und deren sichere Nutzung ermöglichen*) umsetzbar.

⁶¹ Open humans. www.openhumans.org

⁶² Corona-Datenspende. <https://corona-datenspende.de/>

⁶³ Apa Science: Coronavirus – 500.000 Teilnehmer bei „Datenspende“-App in Deutschland (2020). https://science.apa.at/rubrik/politik_und_wirtschaft/Coronavirus_-500_000_Teilnehmer_bei_Datenspende_-App_in_Deutschland/SCI_20200506_SCI40111351054464448

⁶⁴ Deutscher Ethikrat: Big Data und Gesundheit – Datensouveränität als informationelle Freiheitsgestaltung, Stellungnahme (2017).

<https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-big-data-und-gesundheit.pdf>

⁶⁵ Deutscher Ethikrat: Big Data und Gesundheit – Bericht über die öffentliche Befragung des Deutschen Ethikrates (2018). <https://www.ethikrat.org/fileadmin/Publikationen/Studien/befragung-big-data-und-gesundheit.pdf>

Handlungsempfehlung 3.4:

Etablieren von Vertragsbausteinen für eine sichere gemeinsame Datennutzung zwischen Wissenschaft und Wirtschaft

Zielsetzungen:

- Schaffung einer Grundlage für die einfache Erarbeitung von Verträgen zur Nutzbarmachung von Forschungsdaten und Algorithmen zwischen verschiedenen AkteurInnen (v.a. aus Wissenschaft und Wirtschaft)
- Entwicklung von Vertragsbaukästen zur rechtssicheren und einfachen Klärung des Rechtsrahmens zur Datennutzung

Wirkung:

- > Schaffung einer sicheren Basis für gegenseitige Datennutzung zwischen Unternehmen und Wissenschaft
- > Vereinfachung und Beschleunigung von Forschungsprozessen und der Weiternutzung von Daten
- > Ressourcenschonung bei allen beteiligten StakeholderInnen

Maßnahme:

Damit Unternehmen eine sichere Basis haben, um bestimmte Daten der Wissenschaft zur Verfügung stellen zu können und umgekehrt, braucht es häufig sehr aufwändige und heikle Vertragsverhandlungen. Um diesen Prozess zu beschleunigen und zu vereinfachen, ist ein **Vertragsbaukasten mit einem entsprechend leicht verständlichen Entscheidungsbaum** nötig,

durch den **rechtliche Fragestellungen effizienter und klar definiert beantwortet** werden können. Als Vorbild kann hier die umfangreiche „Contract Guidance on Utilization of AI and Data“ des japanischen Ministeriums für Wirtschaft, Handel und Industrie (METI) dienen⁸⁶.

Verträge zwischen Wissenschaft und Wirtschaft sollen zum Enabler für gemeinsame Datennutzung werden

Es sollen vorgefertigte Vertragsbausteine zur Verfügung gestellt werden, welche die **rechtssichere Erstellung von Verträgen erleichtern und beschleunigen**. Diese sollen u.a. die **rechtssichere Verfügbarmachung, Nutzung von Daten und Algorithmen inklusive Haftung sowie Übertragung von Rechten regeln**.

Sie sollen als **Grundlage für die Zusammenarbeit zwischen verschiedenen Kollaborateuren**, wie bspw. zwischen **Wissenschaft und Wirtschaft**, dienen und für die schnellere Weiterverarbeitung maschinenverarbeitbar sein. Die Erarbeitung sollte unter Einbeziehung aller relevanten StakeholderInnen stattfinden – insbesondere Forschungseinrichtungen und wirtschaftliche Interessensvertretungen sollen dabei involviert werden.

⁸⁶ Contract Guidelines on Utilization of AI and Data (2019).
https://www.meti.go.jp/english/press/2019/0404_001.html

Handlungsempfehlung 3.5:

Aktive Mitgestaltung des europäischen Rechtsrahmens für die Nutzung von Daten durch Österreich

Zielsetzungen:

- Aktive Mitgestaltung des europäischen Rechtsrahmens durch die österreichische Forschungspolitik, um den Austausch und die Nutzung von Forschungsdaten durch verschiedene AkteurInnen in der Wissenschaft und zwischen Wissenschaft und Wirtschaft voranzutreiben
- Anstoß einer EU-weiten Entwicklung und Umsetzung von standardisierten Machine-to-Machine fähigen Ownership- und Lizenzmodellen für Daten aus unterschiedlichen Disziplinen und Kooperationsformen

Wirkung:

- > Erhöhte Sicherheit beim Teilen und Nutzen von Daten
- > Erleichterung der interdisziplinären und internationalen Zusammenarbeit innerhalb der Wissenschaft und der Wissenschaft mit der Wirtschaft
- > Erleichterung der Weiternutzung von Forschungsdaten
- > Verringerung von Abstimmungsaufwänden und Unsicherheiten („Chilling Effects“) beim Zurverfügungstellen von Forschungsdaten

Maßnahme:

Die aktuelle europäische Rechtslage stiftet oftmals Verwirrung unter WissenschaftlerInnen bezüglich der Bedingungen, unter denen Daten geteilt und genutzt werden können und welchem Schutz sie unterliegen. Für viele WissenschaftlerInnen ist unklar, wie der Rechtsrahmen im konkreten Fall auszulegen ist, und wie sie sich rechtskonform verhalten sollen (z.B. ob Daten in konkreten Fällen personenbezogen sind oder nicht). Dies führt dazu, dass aus Angst vor Rechtsverstößen (die sogenannten „Chilling Effects“ - abschreckende Wirkungen)

Daten oftmals nicht veröffentlicht oder anderen NutzerInnen zur Verfügung gestellt werden.⁸⁷ Um den Austausch von Daten aus Wissenschaft und Gesellschaft und somit eine mehrfache Nutzung zu forcieren, sind **praxisnahe Definitionen und Absicherungen von möglichen Nutzungszwecken sowie von Data Ownership Modellen** nötig. Dabei müssen Rechte und Pflichten von DatenlieferantInnen sowie DatennutzerInnen klar geregelt werden.

Österreich sollte in Europa bei der Entwicklung gemeinsamer Lizenzmodelle aktiv werden

Um die (Wieder-)Nutzung von Daten zu forcieren, sollen **EU-weite, standardisierte Lizenzmodelle** erarbeitet werden, welche **Data Ownership von Einzelpersonen sowie von interdisziplinären und internationalen Teams** in Bezug auf personenbezogene und nicht-personenbezogene Daten definieren, da Daten häufig kollaborativ erzeugt werden. Diese Lizenzmodelle sollen auf Initiative von Österreich EU-weit entwickelt werden, wobei die Wissenschaftscommunity, Förderorganisationen und die Wirtschaft stark eingebunden und die entwickelten Lizenzen **international anschlussfähig** sein sollen.

Die künftigen Lizenzmodelle sollen **möglichst praktikabel an den Bedürfnissen der ForscherInnen** ausgerichtet sein und

so die Veröffentlichung von Daten und das gemeinsame Arbeiten mit Daten beschleunigen. Weiters sollen sie sicherstellen, dass die Urheber- und Nutzungsrechte an **Forschungsdaten klar geregelt** sind, dass sich einzelne WissenschaftlerInnen über ihre Rechte bewusst sind und diese leichter einfordern können.

Des Weiteren sollen diese Lizenzmodelle von Maschinen gelesen und verarbeitet werden können, um den Arbeitsaufwand in der Weiternutzung für die WissenschaftlerInnen zu verringern. Angelehnt werden diese Modelle an Creative Commons und Open Data Commons Modelle, verbreitet werden sie durch die Verankerung der Nutzung dieser Lizenzmodelle in Leistungsvereinbarungen und Förderungen.

⁸⁷ Response to the public consultation on "A European strategy for data" (2020).
https://europeanlawinstitute.eu/fileadmin/user_upload/p_elii/Projects/Data_Economy/ELI_Respo_nse_European_Strategy_for_Data.pdf

Vereinfachung des Rechtsrahmens kann die Bereitstellung und Nutzung von Daten erleichtern

In einem nachfolgenden Schritt soll die **Neugestaltung, bzw. Überarbeitung der Rechtsgrundlage von österreichischer Seite auf EU-Ebene aktiv betrieben** werden. Beispielsweise ergeben sich hier in Bezug auf große, gemischte Datenpools Unsicherheiten innerhalb der Wissenschaft und Wirtschaft, einige befragte ExpertInnen schildern Verbesserungspotenziale der DSGVO in Bezug auf die Bereiche Big Data und AI⁷⁶. Deshalb soll der europäische Rechtsrahmen klarer ausdefiniert und leichter anwendbar gestaltet werden, um unter den WissenschaftlerInnen und innerhalb der Wirtschaft Hemmnisse abzubauen, Daten zu teilen und kollaborativ zu arbeiten. Auch in Hinblick auf eventuelle Verstöße fehlt es noch an bei-

spielhaften Fällen. Wünschenswert wäre es hier, Bagatelverstöße nachsichtig zu behandeln und Möglichkeiten zu finden, die aktuell großen Ängste vor der Datennutzung und -teilung abzubauen. Des Weiteren soll der europäische Rechtsrahmen harmonisiert werden und über Grenzen hinweg anwendbar sein. Aktuell wird dahingehend oftmals kritisiert, dass durch die verschiedenen rechtlichen Gegebenheiten, in Europa beispielsweise in Bezug auf nationale Ausnahmeregelungen der DSGVO (in Österreich beispielsweise das Forschungsorganisationsgesetz)³², kein einheitlicher Rechtsrahmen besteht und dadurch die transnationale Datennutzung und Zusammenarbeit erschwert wird.

Hemmnisse in der Wirtschaft abbauen

Aus Sicht der befragten WissenschaftlerInnen gibt es Handlungsbedarf im Bereich der **Öffnung von Daten aus der Wirtschaft**. Hier liegt ein großer Wissensschatz begraben, welcher durch eine juristische Klarstellung im Bereich des **Schutzes vor Drittwirkungen** gehoben werden könnte. Der aktuelle Rechtsrahmen regelt hauptsächlich die Rechte und Pflichten bei der Weitergabe und Nutzung von Daten zwischen VertragspartnerInnen, jedoch besteht kein wirksamer Schutz vor widerrechtlicher Nutzung von Daten durch Dritte. Österreich

könnte dahingehend beispielsweise auf nationaler Ebene einen solchen Rechtsschutz schaffen, der an die Geschäftsgeheimnisrichtlinie angelehnt werden könnte, in weiterer Folge könnte dies einen „Push“ auf europäischer Ebene auslösen. Dadurch könnte unter den Unternehmen die Bereitschaft erhöht werden, ihre Daten mit der Wissenschaft zu teilen, da Unternehmen umfassender vor dem Missbrauch ihrer Daten geschützt wären⁸⁷.

5.4 Handlungsfeld 4:

Datenkompetenz in Forschungseinrichtungen

Österreichische Forschungseinrichtungen verfolgen meist noch keine klare Strategie in Hinblick auf den Umgang mit Daten, wodurch es der Entscheidung einzelner WissenschaftlerInnen obliegt, wie sie Daten sammeln, speichern bzw. nutzen. Diese fehlende strukturelle Unterstützung ist einer der Gründe, warum aktuell viele WissenschaftlerInnen datengetriebene Forschung nicht oder nur unzureichend betreiben können.

Insbesondere fehlt an vielen Forschungseinrichtungen der Zugang zu einschlägigem Knowhow im Bereich Data Science und Forschungsdatenmanagement. Sowohl ein strategisches, institutionelles Vorgehen als auch der Aufbau der notwendigen Kompetenzen sind zentrale Erfolgsfaktoren für österreichische Institutionen, um international konkurrenzfähig zu bleiben.

Handlungsempfehlung 4.1:

Entwicklung und Institutionalisierung von Datenstrategien in den Forschungseinrichtungen

Zielsetzungen:

- Entwicklung und Umsetzung von Forschungsdatenstrategien in jeder Forschungseinrichtung – innerhalb des Rahmens der künftigen nationalen Forschungsdatenstrategie (siehe auch *Handlungsempfehlung 1.1 Entwicklung und Umsetzung einer nationalen Forschungsdatenstrategie*)
- Etablierung klarer Zielsetzungen, Maßnahmen und Werkzeuge, um WissenschaftlerInnen bei der Umsetzung zu unterstützen

Wirkung:

- > Anerkennung von datengetriebener Forschung als strategisches Asset
- > Aufbau von Kompetenzen im Bereich Data Science und Data Stewardship in Forschungsinstitutionen
- > Klarheit über Umgang mit Daten bei WissenschaftlerInnen
- > Die (Wieder-)Nutzung von Daten steigt

Maßnahme:

Im Einklang mit der nationalen Forschungsdatenstrategie (siehe *Handlungsempfehlung 1.1: Entwicklung und Umsetzung einer nationalen Forschungsdatenstrategie*) sollen **Forschungseinrichtungen**, also z.B. Universitäten und außeruniversitäre Institutionen, **spezifische Strategien für den Umgang mit Daten entwickeln**, um die konkrete und individuell zielführende Umsetzung an der jeweiligen Institution zu

definieren. Um die Umsetzung und Weiterentwicklung dieser Strategien voranzutreiben, sollen an größeren Institutionen zudem **neue Führungspositionen für den Umgang mit Daten** und der Umsetzung der nationalen Forschungsdatenstrategie eingesetzt werden, wie es an einigen Einrichtungen in Österreich bereits der Fall ist (z.B. Chief Data Officers oder Vizerektoren-Positionen⁸⁸).

Klar definierte Data Ownership und Verpflichtungen geben WissenschaftlerInnen Sicherheit

Unter diese institutionelle Datenstrategie und die entsprechenden Aufgaben fällt, dass **Data Ownership und Verantwortlichkeiten für den Umgang mit Daten gut verständlich in den Arbeitsverträgen** geregelt werden sollen. Auch soll die **Verpflichtung zur Erstellung von FAIR Data**⁵⁰

(durch Vorgaben zu notwendigen Metadaten und Nutzung von sinnvollen Standards auf institutioneller Ebene) und zur **Anwendung der national standardisierten Data Managementpläne** verankert werden.

Institutionen müssen ihre WissenschaftlerInnen darin unterstützen, den richtigen Umgang mit Daten zu leben

Zum anderen sollen die **notigen Ressourcen, z.B. nötige Arbeitszeit für die Pflege und das Zur-Verfügung-Stellen von Daten langfristig in Infrastrukturen, Organisation und Human Resources** (z.B. Hiring Policies) berücksichtigt werden (siehe *Handlungsfeld 4: Datenkompetenz in Forschungseinrichtungen* sowie *Handlungsfeld 5: Ausbildung und Karriere-*

modelle). Auch in der **Aus- und Weiterbildung** soll diese Kompetenz verstärkt angeboten werden. Insbesondere in Bezug auf Human Resources sollte daher die DORA Deklaration³⁹ unterzeichnet werden. Auch sollen Institutionen ihre WissenschaftlerInnen dabei unterstützen, **nationale und internationale Daten-infrastrukturen nutzen zu können**.

Daten der Institutionen sollen genutzt werden, um strategische Entscheidungen treffen zu können

Um die Ziele dieser institutionellen Forschungsdatenstrategien zu erreichen, soll auch an den Institutionen selbst **datengetriebene Forschung genutzt werden, um die Lehre und Forschung der Hochschulen besser zu gestalten und dabei**

Ressourcen zu schonen. Die hierfür nötigen Daten sind oftmals bereits vorhanden, und können für einen erhöhten Wissensgewinn genutzt werden.

⁸⁸ Beispielsweise an der TU Wien: <https://www.tuwien.at/tu-wien/organisation/universitaetsleitung/rektorat/vizerektor-digitalisierung-und-infrastruktur/>

Handlungsempfehlung 4.2:

Anstellung (und Förderung) von Data Science ExpertInnen an allen Forschungseinrichtungen

Zielsetzungen:

- Erhöhung der Data Science-Kompetenz bei WissenschaftlerInnen
- Aufbau von Data Science-Services an den Forschungseinrichtungen

Wirkung:

- > Ermöglicht datengetriebene Wissenschaft in allen Disziplinen (auch weniger daten-affinen)
- > Stellt die richtige Nutzung von neuen Methoden, z.B. AI, sicher (bspw. in Bezug auf ausreichende Datengrundlagen, kein Ziehen vorschneller Schlüsse oder generell nicht-ziel-führender Einsatz)
- > Hebt die Wettbewerbsfähigkeit der Forschungseinrichtung

Maßnahme:

An allen Forschungseinrichtungen sollen Data Scientists zum Einsatz kommen. Diese Data Scientists sollen je **nach Bedarf in den Forschungsgruppen** direkt integriert werden, um beispielsweise als dauerhafte Schnittstelle zwischen AI und ForscherInnen zu agieren (z.B. BioinformatikerInnen), oder als **beratende Data Scientists institutionsübergreifend** zur Verfügung stehen und als Servicestelle für alle Fragen zur

Anwendung von datengetriebener Forschung dienen. Als Vorbild hierfür können beispielsweise die Scientific IT Services der ETH Zürich dienen⁸⁹. Für diese Rolle sollen einerseits neue Stellenangebote geschaffen werden, aber auch ExpertInnen explizit von den Institutionen ausgebildet werden (siehe auch *Handlungsfeld 5: Ausbildung und Karrieremodelle*)

Data Scientists heben die Kompetenzen und sind das Bindeglied zu datengetriebener Forschung

Die Angebote der Data Science Teams an einer Forschungseinrichtung sollen möglichst **nah an den Bedürfnissen und Eigenheiten der Disziplinen ausgerichtet werden**. Hierfür müssen Personen ausgebildet und an diesen Positionen eingesetzt werden, die ausreichende Kenntnisse zu Data Science haben, aber auch fachliches, disziplinbezogenes Fachwissen aufweisen, um anschlussfähig zu sein.

Vor allem die übergreifend eingesetzten Data Science Serviceeinheiten benötigen ausreichend kulturelle Sen-

sitivität, um nicht nur fachlich ideal unterstützen zu können, sondern auch bestmöglich mit den WissenschaftlerInnen in den sehr unterschiedlichen Disziplinen zusammenarbeiten zu können und als Bindeglieder wirken zu können. Entsprechend dem Einsatzgebiet müssen **neue Rollen, Aufgaben und Verantwortlichkeiten an den Instituten spezifisch definiert** und entsprechende **Teams zusammengesetzt** werden.

Nationales Data Science Netzwerk in der Wissenschaft ermöglicht besseres Angebot auch an weniger gut ausgestatteten Institutionen

Zudem sollen **institutionsübergreifende Netzwerke** gebildet werden, die zwischen den an einzelnen Instituten angesiedelten Data Scientists einen Austausch ermöglichen, aber auch einen **Kommunikationskanal in die Forschungsteams bilden**, um dort direkt flexibel und zeitnah Bedarfe zu er-

mitteln und das Angebot entsprechend anpassen zu können. Diese **Netzwerke sollen auch dazu dienen, im Bedarfsfall kleineren Instituten** mit begrenztem Zugriff auf Data Science Expertise Unterstützung zu bieten.

⁸⁹ Data Science Support, Scientific IT Services, ETH Zürich. <https://sis.id.ethz.ch/datascience/>

Handlungsempfehlung 4.3:

Einsatz von Data Stewards und/oder Datamanagementteams an allen Forschungseinrichtungen

Zielsetzungen:

- Erhöhung der Datenmanagement-Kompetenz an allen Forschungseinrichtungen
- Schaffung von Datenmanagement-Unterstützung für WissenschaftlerInnen

Wirkung:

- > Erhöhung der Datenqualität und damit Nutzbarkeit von Daten
- > Sicherstellung von Datenmanagement mit FAIRen Daten⁵¹
- > Migration historischer Daten und Erweiterung der zugänglichen Daten

Maßnahme:

Um die kontinuierliche Datenpflege und -haltung an den Forschungseinrichtungen zu gewährleisten, sollen an den **Forschungseinrichtungen eigene Stellen für Data Stewards**

eingerrichtet werden, welche mit den **entsprechenden Ressourcen und Verantwortlichkeiten** ausgestattet sein sollen.

Data Stewardship als feste Kompetenz an den Forschungseinrichtungen

Diese Data Stewards sollen die WissenschaftlerInnen als erste Anlaufstelle beispielweise bei der **Datenpflege und der Nutzung von Repositorien unterstützen** und so die Datenqualität und -sicherung erhöhen (bezüglich technischer Standards, digitaler Formate, Algorithmen, Anforderungen an die Produktion und Verwaltung von Daten). Je nach Struktur, Größe und Tätigkeitsbereich der Forschungsinstitution sollen

Data Stewards **in einzelnen Forschungsgruppen oder institutionsübergreifend angesiedelt** werden. Hierfür sollen neue Stellen und Berufsbilder geschaffen, Ausbildungen angepasst (siehe *Handlungsempfehlung 5.2: Aus- und Weiterbildung von Data Scientists und Data Stewards spezifisch für die Wissenschaft*), aber auch schon bestehende Berufsbilder, wie z.B. das der BibliothekarInnen weiterentwickelt werden.

Handlungsempfehlung 4.4:

Angebot eines Supports für Datenrecht an jeder Forschungseinrichtung

Zielsetzungen:

- Schaffung leicht zugänglicher Daten-Rechtsberatung für alle WissenschaftlerInnen
- Abbau von rechtlichen Unsicherheiten beim Datenumgang
- Erhöhung der Bereitschaft, vorhandene Daten zugänglich zu machen

Wirkung:

- > Erhöhte Rechtssicherheit im Umgang mit Daten unter den WissenschaftlerInnen
- > Steigerung der (Wieder-)Nutzung von Datensätzen aus unterschiedlichen Quellen
- > Zunahme der Daten-Kooperationen innerhalb der Wissenschaft als auch zwischen Wissenschaft und Unternehmen

Maßnahme:

Es sollen an den **Forschungseinrichtungen Rechtsberatungen eingerichtet** werden, welche **konkrete Unterstützung und AnsprechpartnerInnen zu den rechtlichen Rahmenbedingungen von datengetriebener Wissenschaft bieten**. Diese sollen als leicht zugängliche Anlaufstelle dienen, wenn im Rahmen von datengetriebener Forschung rechtliche Fragen

und Unsicherheiten, etwa zur Datennutzung, Datenspeicherung, Weiternutzung aber auch Lizenzierung, auftauchen. Vor allem in Kooperationen, in denen Zugänge zu Datenbeständen eine Rolle spielen, ist die Lösung dieser Rechtsfragen von großer Bedeutung.

Rechtliche Unsicherheiten und Problemstellungen unterscheiden sich stark und verursachen Stillstand

Da diese Fragestellungen je nach Institution sehr unterschiedlich ausfallen können, sollen die **Rechtsberatungen an den Forschungseinrichtungen** selbst eingerichtet werden, sich jedoch **untereinander national stark vernetzen**, um sich gegenseitig zu unterstützen und Lösungen auszutauschen. Dazu ist es notwendig konkrete Handlungsanweisungen (z.B. in Form von Handbüchern oder Guidelines) zu entwickeln (sowohl auf nationaler Ebene für den gesamten österreichischen Forschungsbetrieb durch das BMBWF als auch auf Ebene von einzelnen Forschungseinrichtungen). Diese Information soll so klar wie möglich an die WissenschaftlerInnen kommuniziert

werden. Unterstützend soll hier die Schaffung eines adäquaten Rechtsrahmens, die standardisierten nationalen Lizenzmodelle und die Vertragsbaukästen wirken (siehe auch *Handlungsempfehlung 3.4: Etablieren von Vertragsbausteinen für eine sichere gemeinsame Datennutzung zwischen Wissenschaft und Wirtschaft* sowie *Handlungsempfehlung 3.5: Aktive Mitgestaltung des europäischen Rechtsrahmens für die Nutzung von Daten durch Österreich*), welche im Idealfall die Notwendigkeit solcher Rechtsberatung an den Institutionen verringern, bzw. überflüssig machen.

5.5 Handlungsfeld 5:

Ausbildung und Karrieremodelle

Um langfristig das Wissen um datengetriebene Forschung zu stärken, braucht es entsprechende Ausbildungen als auch Karrieremodelle an den Forschungseinrichtungen. Aktuell ist der professionelle Umgang mit Daten kaum in den Ausbildungen der WissenschaftlerInnen verankert und findet in den verschiedenen Forschungsdisziplinen in unterschiedlicher Ausprägung statt.

Zudem sind die aktuellen Karrieremodelle wenig attraktiv für hochausgebildete Fachkräfte im Bereich Data Science. Data Scientists finden wesentlich lukrativere Anstellungsoptionen in der Privatwirtschaft als in Forschungseinrichtungen.

Handlungsempfehlung 5.1

Verankerung von Datenkompetenzen in allen Studiengängen

Zielsetzungen:

- Aufbau von Grundwissen zu datengetriebener Forschung und Forschungsdaten bei StudienabgängerInnen in allen Disziplinen
- Aufbau der Datenkompetenz in der Lehre

Wirkung:

- > Höheres Bewusstsein hinsichtlich des Werts von Forschungsdaten und Datenqualität
- > Professionelle Zugänge zu datengetriebener Forschung in der Breite
- > Höhere Bereitschaft, Daten zu teilen
- > Beurteilungskompetenz bezüglich Konditionen (z.B. Publikationen) und Services externer AnbieterInnen wird gestärkt

Maßnahme:

Es soll eine flächendeckende **Grundkenntnisvermittlung zu datengetriebener Forschung** und dem richtigen Umgang mit

Daten in allen Disziplinen bereits in der Ausbildung (curricular und extra-curricular) **erfolgen**.

Eine breite, grundlegende Wissensbasis ist nötig

Diese Grundkenntnisse sollen folgende Punkte beinhalten:

- **Grundlagen des Programmierens**
- **Computational Thinking**
- **Grundlagen im Umgang mit Datenbanken**
- **Erstellung von Daten gemäß Standards** (siehe auch *Handlungsempfehlung 1.1: Entwicklung und Umsetzung einer nationalen Forschungsdatenstrategie*)
- **Data Analytics**
- **Forschungsdatenmanagement**
- **Data Ethics**
- **Datenrecht** (inkl. Lizenzen)

Disziplinenübergreifend und praxisnah lässt sich datengetriebene Forschung bestmöglich verankern

Im Bestfall soll die Vermittlung dieser Kenntnisse in interdisziplinären Kursen erfolgen, in diesen können die TeilnehmerInnen **gemeinsam Data Science und zugleich interdisziplinäre Arbeit** erfahren. Als besonders hilfreich wird hier zudem die Vermittlung von praxisnahen Erfahrungen

erachtet, beispielsweise in Data Science Bootcamps wie es die Universität Innsbruck bereits durchführt⁹⁰, in denen interdisziplinär und gemeinsam mit der Wirtschaft gearbeitet werden soll und Use Cases geschaffen werden.

Handlungsempfehlung 5.2:

Aus- und Weiterbildung von Data Scientists und Data Stewards spezifisch für die Wissenschaft

Zielsetzungen:

- Aufbau von Expertise in Data Science und Datamanagement in und für die Wissenschaft
- Abdeckung des Fachkräftebedarfs im Datenbereich – Vorbeugung von Fachkräftemangel

Wirkung:

- > Genügend Fachkräfte für den spezifischen Bedarf von Forschungseinrichtungen
- > Entwicklungschancen für WissenschaftlerInnen im Datenbereich anbieten

Maßnahme:

An den **Hochschulen, aber auch in anderen Einrichtungen sollen verstärkt Aus- und Weiterbildungen im Bereich der Data Science und Data Stewardship** für die Wissenschaft angeboten und kommuniziert werden. Hierfür sollen **neue**

Aus- und Weiterbildungen für WissenschaftlerInnen als auch QuereinsteigerInnen entwickelt werden, um den wachsenden Bedarf an Fachkräften zu decken.

Datengetriebene Forschung entwickelt sich schnell, Wissensvermittlung muss schritthalten

Diese Ausbildungen sollen **regelmäßig evaluiert werden und flexibel an die realen Bedarfe und Veränderungen angepasst** werden. Neue Weiterbildungsformate sollen schnell implementiert und möglichst bald in einer **experimentellen Herangehensweise** umgesetzt werden (aus-

probieren, lernen, verbessern), auch um neue Erkenntnisse zur Ausbildung in Data Science und Stewardship zu gewinnen. Teil dieser Aus- und Weiterbildungen sollen neben Data Science und Data Stewardship auch Vertiefungen zu interdisziplinärer Arbeit und AI sein.

Aus- und Weiterbildungen sollen dem Fachkräftemangel entgegenwirken

Um einem Fachkräftemangel entgegenzuwirken, sollen WissenschaftlerInnen unterschiedlichen Alters und Erfahrungsgrades **aus unterschiedlichen Disziplinen**, die somit über Disziplinwissen verfügen, als auch **QuereinsteigerInnen** mit unterschiedlichem Vorwissen in Data Science bzw. Stewardship aus- und weitergebildet werden. Dabei wird es in einigen Bereichen der datengetriebenen Forschung nicht nötig sein, ein volles Studium zu absolvieren, häufig kann auch der flexible Erwerb von **Zusatzqualifikationen, etwa in Form von**

Erweiterungsstudiengängen ausreichend sein. Zudem sollten die Zugangsbeschränkungen möglichst niedrig sein und ein offenes Angebot geschaffen werden, welches durch Förderinstrumente finanziert wird. Um das Interesse an diesen Angeboten zu verstärken, sollen neue Kommunikationsformate genutzt werden, die auch abseits der akademischen Welt Anklang finden. Auch Job Visiting bei bereits tätigen Fachkräften soll im Rahmen der Ausbildungen möglich sein, um reale Praxiserfahrung zu erlangen.

Handlungsempfehlung 5.3:

Entwicklung von attraktiven unbefristeten Karrieremodellen und Aufgabenbereichen in Data Science, Data Stewardship und Datenrecht

Zielsetzungen:

- Langfristige Integration von Data Scientists, Data Stewards und DatenrechtsexpertInnen in Forschungsinstitutionen

Wirkung:

- > Hebung des Kompetenzniveaus und der Wettbewerbsfähigkeit der Forschungseinrichtungen
- > Bindung von Fachkräften an die Forschungsinstitutionen, um nicht nur auf die Akquise von Fachkräften am (ausgedünnten) Markt angewiesen zu sein

Maßnahme:

Um Data Scientists, Data Stewards und DatenrechtsexpertInnen in den Forschungsinstitutionen einsetzen und halten zu können (wie in *Handlungsfeld 4: Datenkompetenz in*

Forschungseinrichtungen beschrieben) beschrieben), braucht man Möglichkeiten, ihnen **attraktive Karrieremöglichkeiten an Forschungseinrichtungen** bieten zu können.

Im Wettbewerb um Daten-ExpertInnen müssen Forschungseinrichtungen mit der Wirtschaft mithalten können

Deshalb sollen an den Forschungseinrichtungen **langfristige Anstellungsverhältnisse für Daten-SpezialistInnen** geschaffen werden, welche auch abseits von Professurstellen gute Karriereoptionen bieten. Auf diese Weise können Daten-SpezialistInnen unabhängig von Drittmitteln möglichst lang-

fristig ihre Expertise in die Wissenschaft einbringen. Idealerweise werden hierfür Personen rekrutiert, welche die Institutionen bereits kennen, beispielsweise StudienabgängerInnen.

Handlungsempfehlung 5.4:

Neue Modelle der Leistungsmessung, Bewertung und Rekrutierung von WissenschaftlerInnen

Zielsetzungen:

- Stärkung der Grundkompetenzen im Bereich der datengetriebenen Forschung an den Forschungsinstitutionen
- Aufbau sich ergänzender, interdisziplinärer Forschungsteams

Wirkung:

- > Ausreichend vorhandene, vielfältige Kompetenzen, die für datengetriebene Forschung nötig sind
- > Mehr Diversität (berufliche Hintergründe, Herkunft, Gender, ethnische Zugehörigkeit, etc.)
- > Klare Rollenverteilung und Nutzung von synergistischen Kompetenzen in Forschungsteams

Maßnahme:

Datengetriebene Forschung fordert neue Arbeitsweisen an den Institutionen und damit auch eine Vielzahl neuer Kompetenzen. Dafür wird auch ein zeitgemäßes Anreizsystem benötigt. Daher sollen **WissenschaftlerInnen anhand neuer**

Metriken bewertet werden und **entsprechend ihrer individuellen Fähigkeiten interdisziplinär in Teams zusammenarbeiten**, um einen Datenfluss über die Disziplingrenzen hinweg zu erleichtern.

Die bisherige Bewertung wissenschaftlicher Leistung ist zu eindimensional für datengetriebene Wissenschaft

Es sollen daher **neue Bewertungskriterien für ForscherInnen entwickelt** werden, welche nicht ausschließlich auf Publikationslisten basieren. Diese sollen **Kompetenzen berücksichtigen, welche beim datengetriebenen wissenschaftlichen Arbeiten** benötigt werden. Unter anderem sollen folgende Dimensionen mitberücksichtigen werden:

- der **langfristig geplante Umgang mit Daten**
- **Veröffentlichung von Forschungsergebnissen als hochqualitative, weiternutzbare Daten in Form von Open Access oder Open Data** (auch unabhängig von

Publikationen und auch im Fall von negativen Forschungsergebnissen)

- das **Schreiben und Veröffentlichen von Software-Codes und Algorithmen**
- sowie stark **interdisziplinäre Zusammenarbeit und Vermittlerrollen**.

Wichtig ist hierbei, dass nicht alle WissenschaftlerInnen in allen Bereichen exzellente Kompetenzen aufweisen müssen, jedoch ein gewisses Grundverständnis mit wenigen tiefergehenden Wissensgebieten haben und entsprechend dieser angestellt und befördert werden sollen.

Leistungsmessung verändern: Nicht die Kompetenzen des/der einzelnen ForscherIn zählen, sondern die des gesamten Teams

Zudem sollen diese **neuen Bewertungskriterien in die Planung der Teams** einfließen, damit sich die **Teammitglieder ideal ergänzen**. Es muss somit nicht jedes Teammitglied alle Kompetenzen mitbringen, sondern gemeinsam sollen alle Bereiche gut abgedeckt werden. Hierfür soll nach Möglichkeit Cluster-Hiring⁹¹ angewendet werden, das bedeutet, dass neue ForscherInnen nicht einzeln angestellt werden, sondern direkt vollständige, sich ergänzende Teams rekrutiert werden.

Entsprechend der individuellen Fähigkeiten innerhalb der Teams sollen auch die **Verantwortlichkeiten und Aufgaben aufgeteilt** werden und diese **Teilung klar vermittelt und abgegrenzt** werden. Somit können alle Teammitglieder entsprechend ihrer Kernaufgaben, aber gemeinsam auf einer sich ergänzenden Basis arbeiten.

⁹¹ Faculty Hiring: 4 Trends to follow (2019). <https://universitybusiness.com/4-faculty-hiring-trends-to-follow/>

5.6 Handlungsfeld 6:

Dateninfrastrukturen und europäische Kooperation

Aktuell sind die meisten Dateninfrastrukturen in Österreich auf institutioneller bzw. disziplinärer Ebene realisiert. Obwohl Österreich bereits an internationalen und europäischen Initiativen teilnimmt, sind die nationalen Infrastrukturen bisher nicht flächendeckend an diese anschlussfähig. Dadurch wird interdisziplinäre und länderübergreifende Nutzung und Veröffentlichung von Daten verhindert und die Zusammenarbeit erschwert.

Zurzeit müssen WissenschaftlerInnen in Österreich oftmals aufgrund eines Mangels an Alternativen auf kommerzielle An-

gebote z.B. für die Datenspeicherung oder Datenanalyseservices zurückgreifen. Hierdurch laufen sie Gefahr, die Kontrolle über die von ihnen erarbeiteten und in Österreich finanzierten Daten zu verlieren, was eine Abhängigkeit und potenziellen Schaden für den Wissenschaftsstandort darstellt. Dies behindert aktuell besonders die Arbeit mit sensiblen Daten. Es besteht ein großer Bedarf an europäischen Infrastruktur-lösungen, die als vertrauenswürdig wahrgenommen werden. Dieser wird durch die kommerziellen Angebote derzeit nicht gestillt.

Handlungsempfehlung 6.1

Verstärkte aktive Teilnahme Österreichs an internationalen Initiativen für Dateninfrastruktur

Zielsetzungen:

- Einnahme einer federführenden Rolle Österreichs bei der Ausgestaltung europäischer Datenkooperationen
- Aktive Mitgestaltung internationaler Dateninfrastrukturen
- Verstärkte Nutzung von europäischen Dateninfrastrukturen im Vergleich zu nationalen Lösungen

Wirkung:

- > Zugang zu großen Infrastrukturen und zukünftigen Projekten der internationalen Gemeinschaft
- > Gesteigerte Resilienz durch föderierte, dezentral verknüpfte Infrastrukturen

Maßnahme:

Der **österreichische Forschungsstandort soll die EOSC (European Open Science Cloud) aktiver mitgestalten**, wann immer möglich diese anstelle von nationalen Lösungen nutzen, und diese in weiterer Folge auch mitfinanzieren. Österreich soll

insgesamt eine **federführende und visionäre Rolle in der Ausgestaltung der EOSC und weiterer internationaler Initiativen** einnehmen.

Auf nationaler Ebene müssen internationale Initiativen gestützt und evaluiert werden

Hierfür sollen **kontinuierlich Bedarfe der Forschungseinrichtungen für Infrastrukturen, Dienstleistungen und Services erhoben werden** und diese in die Diskussionen und Arbeit auf internationaler Ebene einfließen. Wichtig sind hierbei eine **langfristige Finanzierung und Ausrichtung**. Insbe-

sondere die Mitfinanzierung von europäischen Initiativen soll außerhalb von zeitlich begrenzten Projekten realisiert werden. Weiters soll die Teilnahme und Nutzung von anderen europäischen Infrastrukturen, wie z.B. dem Pan-Europäischen Forschungsnetz GÉANT⁹², verstärkt werden.

Handlungsempfehlung 6.2

Ausbau österreichischer Dateninfrastrukturen und Services

Zielsetzungen:

- Aufbau nationaler Dateninfrastrukturen und Services zur Unterstützung des Forschungsstandortes und der Anschlussfähigkeit an europäische Dateninfrastrukturen
- Zugang zu Rechenkapazitäten, Speichermöglichkeiten, neuester Technik, aber auch Services und Unterstützung für alle österreichischen ForscherInnen

Wirkung:

- > Sicherstellung der Anschlussfähigkeit und Unabhängigkeit der österreichischen Forschungsdateninfrastruktur
- > Schaffung von Alternativen zu kommerziellen Anbietern, Verhinderung von Lock-In-Effekten
- > Optimierung von Ressourcennutzung und gestärkte Ausfallssicherheit

Maßnahme:

Abgestimmt auf die nationale Forschungsdatenstrategie (siehe Handlungsempfehlung 1.1: *Entwicklung und Umsetzung einer nationalen Forschungsdatenstrategie*), soll Österreich **eigene**

Dateninfrastrukturen dort aufbauen, wo die europäischen Angebote nicht ausreichen, oder wenn sie als resiliente Ausfallstrukturen dienen können.

National den Forschungsstandort stärken, langfristig die Möglichkeiten durch internationale Anbindung erhöhen

Diese national aufgebauten Infrastrukturen sollen so gestaltet sein, dass dabei **mehrere Forschungseinrichtungen zusammenarbeiten und die Infrastrukturen optimal an internationale Infrastrukturen angeschlossen** werden können (z.B. in die EOSC integriert werden können), bzw. zumindest weitgehende Interoperabilität gewährleistet ist. Auch sollen Standards, die auf europäischer Ebene entstehen (z.B. FAIR Prinzipien) oder in der

nationalen Forschungsdatenstrategie definiert werden, konsequent umgesetzt werden. Dazu ist die enge Zusammenarbeit mit entsprechenden Initiativen (z.B. GO-FAIR Implementation Networks⁹³) von großer Bedeutung. Wie auch in der nationalen Forschungsdatenstrategie soll gelten: **So international/kooperativ wie möglich, so lokal wie nötig.**

Klare Aufgabenteilung zwischen Infrastrukturen ermöglicht einen sparsamen Einsatz der Mittel

Um Ressourcen intelligent zu nutzen, soll national auf eine **klare Aufgabenverteilung und Spezialisierung zwischen den Infrastrukturen** für unterschiedliche Aufgabenstellungen geachtet werden. Dort wo dies Sinn macht, sollen **kleinere**

existierende Infrastrukturen in größere integriert werden. Eine gewisse Redundanz soll dabei berücksichtigt werden, um die notwendige Ausfallsicherheit und Resilienz zu gewährleisten.

Von überall aus möglichst einfach Zugriff erhalten

Es ist wesentlich, dass künftig bei **Datenspeicherzentren auch entsprechende Rechenzentren angedockt** werden. Diese sollen wiederum für spezifische Aufgabenstellungen optimiert werden, sodass ein Spektrum unterschiedlicher spezialisierter Rechenzentren für WissenschaftlerInnen zur Verfügung steht. Insbesondere sehr datenintensive Aufgaben sollen an solchen kombinierten Rechen- und Speicherzentren durchgeführt werden können. Diese Infrastrukturen sollen **remote genutzt**

werden können und möglichst **unbürokratisch, offen und kostenfrei** für WissenschaftlerInnen **zugänglich** sein (z.B. durch „Single Sign-On“ bzw. „Free-at-Point-of-Use“). Für österreichische BürgerInnen, sowie in gewissem Maße für internationale WissenschaftlerInnen sollen die Datensets ebenso zugänglich sein. Wichtig ist eine möglichst einfache Usability, ausgerichtet an den Bedürfnissen der WissenschaftlerInnen sowie ein guter Überblick über alle Inhalte.

ForscherInnen aller Disziplinen und Wissensstände bestmöglich unterstützen

Diese Infrastrukturen und Services sollen zudem möglichst **interdisziplinär nutzbar sein**. Neben dem Betrieb soll auch entsprechender **Support und Services angeboten** werden, um die Angebote möglichst breit zu öffnen. Hierfür sollen an diesen Datenzentren besonders gut ausgebildete ExpertInnen eingesetzt werden, die Datenspeicher und Anwendungen ideal bedienen und begleiten können. Zusätzlich zu dieser Unterstützung sollen dort Programme entwickelt werden, um Daten automatisiert zu verarbeiten und somit den Human-

ressourcenbedarf langfristig zu reduzieren. Diese Infrastrukturen, Services und Unterstützungen sollen regelmäßig aktualisiert und an den Bedürfnissen der WissenschaftlerInnen und der technologischen Weiterentwicklung ausgerichtet werden, auch auf Grundlage der Empfehlungen der Kommission für Datenethik und Foresight zu datengetriebener Wissenschaft (siehe *Handlungsempfehlung 1.2: Aufbau einer Kommission für Datenethik und Foresight zu datengetriebener Wissenschaft*).

Infrastrukturen können nur die volle Wirkung entfalten, wenn sie langfristig angelegt werden

Die **Finanzierung solcher Infrastrukturen und Services soll langfristig gestaltet** werden. Insbesondere soll sichergestellt werden, dass Infrastrukturen zumindest nach einer initialen Etablierungsphase nicht im Rahmen von Projekten finanziert

werden, sondern langfristig. Dafür braucht es nachhaltige Finanzierungssysteme. Umgekehrt sollen die Prozesse für die Bezahlung von Services einfach, schnell und transparent gestaltet sein.

⁹³ Go FAIR Implementation Networks. <https://www.go-fair.org/implementation-networks/>

Handlungsempfehlung 6.3

Aufbau von Data Trusts, welche sensitive Daten hosten und deren sichere Nutzung ermöglichen

Zielsetzungen:

- Neutrale Stelle zur Speicherung und Nutzung von sensitiven Daten
- Überwachung von Datennutzung und Datenhosting

Wirkung:

- Ermöglichung der Nutzung von ansonsten nicht zugänglichen, sensitiven Daten
- Aufbau von Vertrauen, sodass auch Unternehmen relevante Daten hinterlegen und zur Verfügung stellen

Maßnahme:

Als Ergänzung zu den bereits beschriebenen Dateninfrastrukturen sollen, wie es unter anderem die deutsche Datenethikkommission empfiehlt⁹⁴, spezifische Stellen geschaffen werden, die besonders im Falle von sehr sensitiven Daten eingesetzt werden und so neue Datenzugänge, etwa für AI, ermöglichen. Die Nutzbarmachung solcher sensibler Daten passiert in diesen nicht mehr bilateral zwischen zwei Vertragsparteien, sondern über **Data Trusts** bzw. Datentreuhandstellen, die als Intermediäre agieren. Diese **neutralen, vertrauenswürdigen Stellen** hosten die Daten und **überwachen die Nutzung** (denkbar in Kombination mit *Handlungs-*

empfehlung 6.2: Ausbau österreichischer Dateninfrastrukturen und Services). Aufgabe dieser Stellen ist es insbesondere sicherzustellen, dass Daten stets ausschließlich auf die vertraglich festgelegte Art und Weise genutzt werden. Dadurch kann es beispielsweise Unternehmen ermöglicht werden, ausgewählten Forschungsgruppen Daten zu definierten, wissenschaftlichen Zwecken zur Verfügung zu stellen, ohne befürchten zu müssen, dass durch unlautere Nutzung Geschäftsgeheimnisse verraten werden. Um hier ausreichend Vertrauen aufzubauen, sollen diese Stellen unabhängig und neutral agieren.

Sicherer Zugang zu Daten und sicheres Teilen ist in vielen Anwendungsfällen dringend nötig

Die Data Trusts sollen verschiedenen StakeholderInnen dienen und damit in einem breiten Anwendungsfeld die Datennutzung erhöhen. So sollen einerseits **Vorbehalte in der Wirtschaft** gegenüber dem Teilen von Daten durch solche Infrastrukturen reduziert werden und Daten aus der Wirtschaft z.B. für die Wissenschaft zugänglich gemacht werden. Zudem kann die

Nutzung **sehr sensibler Daten aus der Wissenschaft** ermöglicht werden, welche sonst Datenschutzbeschränkungen unterliegen würden, wie auch von **Daten der öffentlichen Hand und von BürgerInnen**. Langfristig sind diese Datenzugänge insbesondere in der Nutzung und Entwicklung von AI essenziell.

Reines Datenspeichern ist nicht ausreichend, Überprüfung und Services sind nötig, um Sicherheit herzustellen

Data Trusts sind **Hochsicherheitsdatenzentren, die auch Services wie Datennutzung, ohne dass die Daten selbst geteilt werden (etwa lokale Berechnungen) und das Monitoring der Datennutzung anbieten** und an andere Infrastrukturen (institutionelle, nationale und internationale) anknüpfbar sind. Sie sollen ausgewiesene Personengruppen authentifizieren und für diese Zugriffe auf bestimmte sensitive Daten unter standardisierten Auflagen ermöglichen. Dies kann durch **neue Verschlüsselungsmethoden** (z.B. homomorphische

encryption), aber auch durch **Data Visiting oder Differential Privacy** geschehen. Idealerweise sollen die Daten nicht die Infrastruktur der Data Trusts verlassen, sondern beispielsweise in eigenen dafür eingerichteten Arbeitsräumen innerhalb der Data Trusts genutzt werden. Vorbild kann hier die Unabhängige Treuhandstelle der Universitätsmedizin Greifswald, Deutschland⁹⁵, sein. Nationale Data Trusts können mit europäischen Stellen, etwa aus dem Horizon 2020 Projekt TRUSTS - Trusted Secure Data Sharing Space⁹⁶ verknüpft werden.

⁹⁴ Arbeitsergebnisse der Datenethikkommission (2019). <https://www.bmi.bund.de/DE/themen/it-und-digitalpolitik/datenethikkommission/arbeitsergebnisse-der-dek/arbeitsergebnisse-der-dek-node.html>

⁹⁵ Unabhängige Treuhandstelle der Universitätsmedizin Greifswald. <https://www.ths-greifswald.de>

⁹⁶ TRUSTS - Trusted Secure Data Sharing Space. <https://www.trusts-data.eu/>

6. Resümee

Big Data, Algorithmen und AI revolutionieren derzeit das Wissenschaftssystem. Aus dieser Erhebung geht hervor, dass im österreichischen Forschungsbetrieb prinzipiell die Wichtigkeit des Themas datengetriebener Wissenschaft erkannt wurde. Dies zeigt sich besonders durch erste fortschrittliche Projekte und Initiativen, die in Österreich bereits gesetzt wurden und als optimale Ausgangs- und Anknüpfungspunkte für den zukünftigen Ausbau von datengetriebener Wissenschaft dienen können.

Dennoch besteht akuter Handlungsbedarf für die österreichische Wissenschaft, um sowohl kurz- als auch langfristig die enormen Potenziale von datengetriebener Forschung heben zu können. Hierbei ist ein strategisches, koordiniertes Vorgehen (sowohl auf nationaler als auch auf europäischer Ebene) erfolgsentscheidend. Durch Adressierung der in dieser Studie identifizierten Handlungsfelder kann sichergestellt werden, dass die österreichische Forschung weiterhin wettbewerbsfähig bleibt, und menschenzentrierte, datengetriebene Forschung einen zentralen Beitrag zur Lösung der großen aktuellen gesellschaftlichen Probleme innerhalb der nationalen Grenzen und darüber hinaus, leisten kann.

Abbildungsverzeichnis

Abbildung 1 Die sieben internationalen Trends datengetriebener Wissenschaft	10
Abbildung 2 Die Handlungsfelder und -empfehlungen für datengetriebene Wissenschaft	12
Abbildung 3 Prozess der Studienerstellung	26
Abbildung 4 Die sieben internationalen Trends datengetriebener Wissenschaft	30
Abbildung 5 Handlungsfelder und Empfehlungen für datengetriebene Wissenschaft	48

