

SONDERDRUCK  
aus

2 | 2024

**VBio**

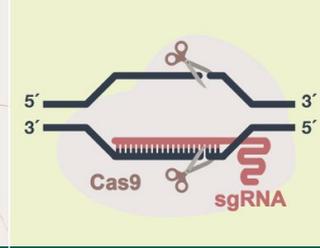
Verband | Biologie, Biowissenschaften  
& Biomedizin in Deutschland



**ÖKOLOGIE**  
Umwelt-DNA aus der  
Vergangenheit



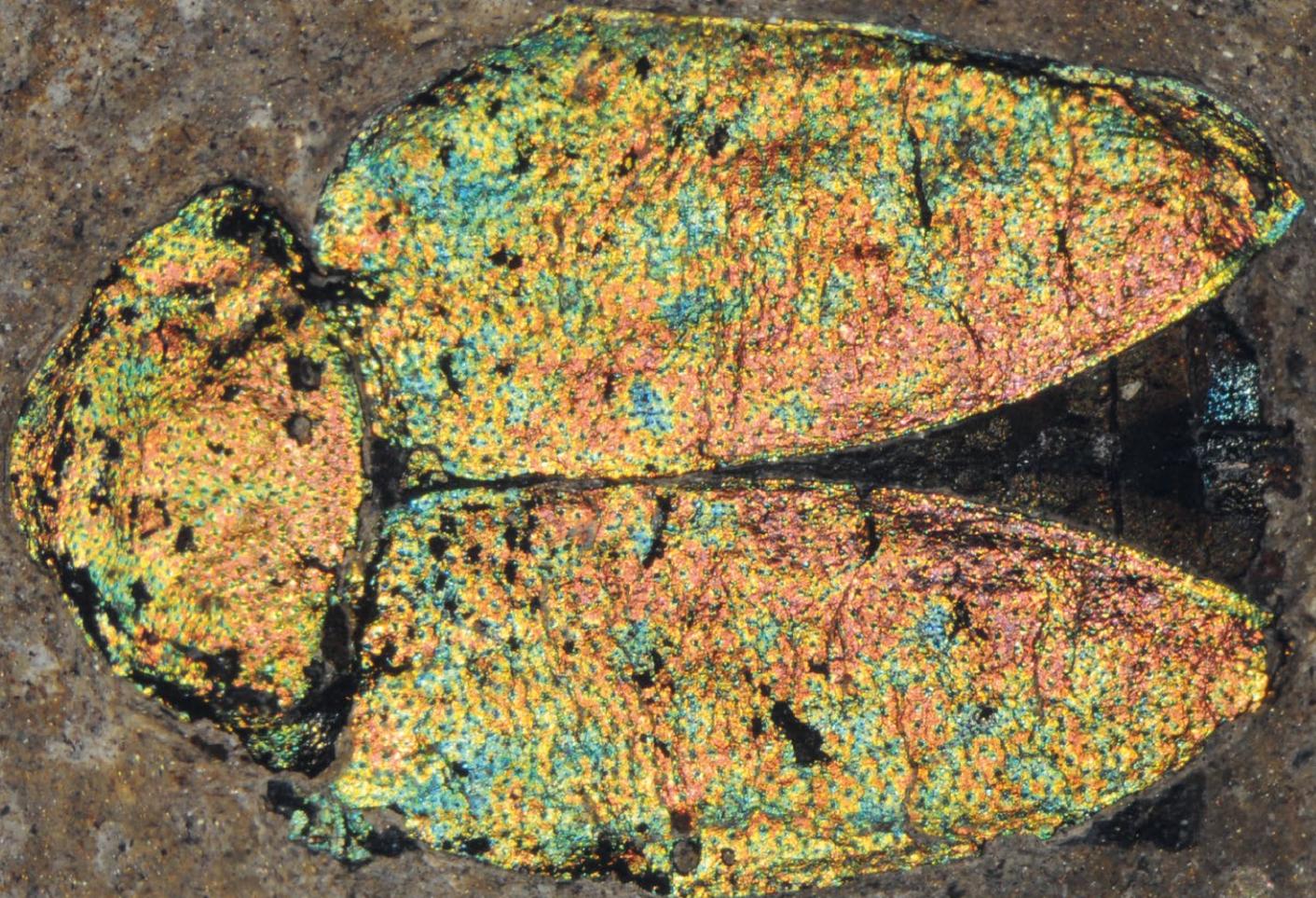
**ALGENFORSCHUNG**  
Nathanael Pringsheims  
sexuelle Revolution



**PFLANZEN  
ZÜCHTUNG**  
Innovationen durch  
Genom-Editierung

# BIOLOGIE

IN UNSERER ZEIT



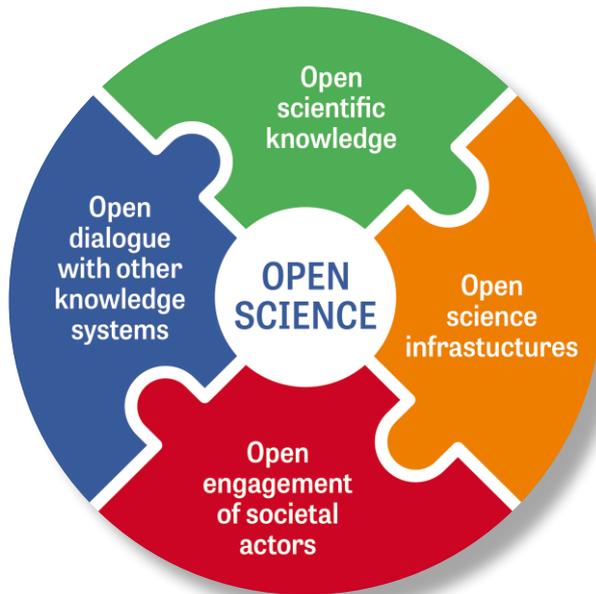
**Fossile Insekten  
aus der Grube Messel**

## Von Open Access zu Open Science

# Offene Daten, offene Wissenschaft?

KERSTIN ELBING | ERWIN BECK | BARBARA EBERT | KONRAD FÖRSTNER

**Open Science hat viele Facetten: Sie umfasst Offenheit in Hinblick auf wissenschaftliche Erkenntnisse und wissenschaftliche Infrastrukturen, aber auch bezüglich des Engagements gesellschaftlicher Akteure und Offenheit für andere Wissenschaftssysteme (z. B. traditionelles Wissen).** Abb.: P. Kneib/ZB Med nach [13].



Mit der „UNESCO-Empfehlung für offene Wissenschaft“ haben 193 Staaten im Herbst 2021 gemeinsame wissenschaftspolitische Leitlinien formuliert, die von Publikationen, Daten und Lehrmaterialien bis hin zur Öffnung von Forschungsprozessen für gesellschaftliche Akteure reichen. Doch wie beeinflusst das Konzept von Open Science Ausbildung, Studium, Lehre und Forschung in der Biologie? Wird nun alles anders? Wie viel Aufwand bringt es für die Einzelnen und für das Wissenschaftssystem? Wo liegen die Chancen, wo die Grenzen des Konzeptes? Vier Autor/-innen mit sehr unterschiedlichem Hintergrund haben sich auf die Suche nach Antworten auf diese Fragen begeben. Gemeinsam beleuchten sie aus unterschiedlichen Blickwinkeln, welchen Effekt die wachsende Bewegung hin zu offener Wissenschaft auf das Alltagsgeschäft von Biolog/-innen hat.

Wissenschaft hat schon immer vom Austausch von Material und Daten aller Art gelebt. Je nach Forschungsgebiet unterscheiden sich die Herangehensweisen, die sich besonders in den letzten Jahrzehnten erheblich gewandelt haben. Und der Paradigmenwechsel hin zu einer breit verstandenen *Open Science* wird die wissenschaftliche Praxis weiter wandeln. Im Folgenden diskutieren vier Expert/-innen *Open Science* aus den Blickwinkeln verschiedener biologischer Fachrichtungen:

Erwin Beck, Jahrgang 1937, ist emeritierter Professor für Pflanzenphysiologie, ehemaliger Vorsitzender der Senatskommission für Biodiversitätsforschung der Deutschen Forschungsgemeinschaft (DFG) sowie der Arbeitsgruppe *Access and Benefit Sharing* und langjähriger Kurator von „Biologie in unserer Zeit“ (Mitgliederzeitschrift des Verbandes Biologie, Biowissenschaften & Biomedizin in Deutschland – VBIO e. V.).

Barbara Ebert, Jahrgang 1970, ist promovierte Biologin und Wissenschaftsmanagerin, die sich als Geschäftsführerin der Gesellschaft für Biologische Daten e. V. für den Aufbau von Forschungs- und Dateninfrastrukturen engagiert.

Kerstin Elbing, Jahrgang 1967, ist promovierte Biologin und Ressortleiterin „Wissenschaft und Gesellschaft“ im VBIO. Sie hat sich unter anderem intensiv mit den Themen „Biodiversitätskonvention“, „Nagoya-Protokoll“ und „digitale Sequenzinformationen“ beschäftigt.

Konrad U. Förstner, Jahrgang 1980, ist promovierter Bioinformatiker, Professor für *Data and Information Literacy* an der Technischen Hochschule Köln und leitet gleichzeitig bei ZB MED – Informationszentrum Lebenswissenschaften den Programmbereich *Data Science and Services*.

**Ebert:** *Open Science* ist ein vielseitiger, fast schillernder Begriff. Als ich Ende der 1990er Jahre in Göttingen promovierte, gab es schon die Erwartung oder sogar Pflicht, molekulare Sequenzdaten in der Europäischen Nukleotiddatenbank ENA zu veröffentlichen (zu den Nukleotiddatenbanken siehe Kasten „*Bermuda Principles*: Grundlage für offene Daten in der Molekulargenetik“). Das hat mich damals sehr beeindruckt und schien mir konsequent; wir hatten ja selbst diese riesige Genomdatenbank genutzt, um die Phylogenie unseres Kaliumkanalgens zu untersuchen. Aber noch immer sind offene Daten dieser



## BERMUDA PRINCIPLES: GRUNDLAGE FÜR OFFENE DATEN IN DER MOLEKULARGENETIK

Im Rennen um die Sequenzierung des menschlichen Genoms wurden 1996 die sog. Bermuda Principles formuliert: Die Konkurrenten Celera Genomics (geleitet von Craig Venter) und die akademischen Forschungsgruppen aus dem Humangenomprojekt einigten sich darauf, alle Sequenzdaten der Akteure aus dem Rennen um die Sequenzierung des menschlichen Genoms zeitnah in einer der anerkannten Genomdatenbanken als sog. Public Good öffentlich zugänglich zu hinterlegen. So konnte die befürchtete Privatisierung der Genomdaten abgewendet werden. In der Folge wurden die von der Human Genome Organisation formulierten Prinzipien als allgemeine Grundlage in der Molekulargenetik akzeptiert und große Dateninfrastrukturen entstanden. Die International Nucleotide Sequence Database Collaboration (INSDC) mit den drei Genomdatenzentren in den USA (National Center for Biotechnology Information, NCBI), Europa (European Molecular Biology Laboratory-European Bioinformatics Institute, EMBL-EBI) und Japan (DNA Data Bank of Japan, DDBJ) wurde 1987 gegründet. Sie verzeichnet heute Millionen von Sequenzdaten und ist eine unverzichtbare Säule moderner Forschung (Abbildung 1). So war die schnelle Identifikation des neuartigen SARS-Coronavirus 2003 nur durch die dort hinterlegten Vergleichsdaten möglich [2].

## BIOLOGISCHE DATEN

Biologische Daten werden in verschiedene Typen unterteilt:

Typ 1: Biodiversitäts- und Vorkommensdaten aus dem klassischen Sammlungs- und Biodiversitätsforschungsbereich, d. h. digitale Objekte mit Taxonomie(n), Georeferenzen, z. B. Ort, Datum und häufig referenzierte Ressourcen als Multimedia-Objekte.

Typ 1a: Sammlungsdaten (mit Bezug zum physischen Objekt)

Typ 1b: Beobachtungsdaten (ohne Bezug zu einem physischen Objekt)

Open Data Resource: Global Biodiversity Information Facility (GBIF) mit 2,3 Milliarden Datensätzen weltweit.

Typ 2: Taxondata

Es handelt sich um taxonbezogene Daten (z. B. in einem Katalog, einer Checkliste oder einer sogenannten Roten Liste).

Open Data Resource: keine koordinierte Infrastruktur; Daten oftmals nur aus Publikationen extrahierbar. Hoher Bedarf. Neu: API (Programmierschnittstelle) zum Checklist Editor Rote-Liste-Institut in Deutschland (mit Mitteln der Nationalen Forschungsdateninfrastruktur (NFDI) finanziert)

Typ 3: Umweltbiologische Daten und ökologische Daten

Dabei handelt es sich um Daten aus umweltbiologischen und ökologischen Studien (z. B. funktionelle und phylogenetische Merkmalsdaten).

Open Data Resource: PANGAEA Data Publisher for Earth and Environmental Sciences

Typ 4: Nicht-molekulare Analysedaten

Hierbei handelt es sich um nicht-molekulare Daten aus Messungen und Laboranalysen (Datensätze und/oder Datenpakete) in ihrem ursprünglichen Dateiformat (z. B. Massenspektren).

Open Data Resource: Keine konsolidierten Datenkorpora. Datensätze werden vereinzelt in thematischen und institutionellen Repositorien veröffentlicht, Zugang meist über die verknüpfte Literatur.

Typ 5: Molekulare Daten

Dies sind Daten, die sich u. a. auf Nukleotide, Proteine und Metabolite beziehen.

Open Data Resources: INSDC-Datenbanken für Nukleotiddaten sowie „Uniprot“ für Proteine.

Zusammengestellt nach [4], German Federation for the Curation of Biological Data, Major Types of Biological Data.

## Was sind biologische Forschungsdaten und wem gehören sie?

Biowissenschaftliche Forschungsdaten sind vielfältig und umfassen Messwerte aller Art, Beobachtungs- oder Sequenzdaten von Proteinen, DNS und RNS. Auch Informationen zur Methodik der Datengewinnung oder zur Georeferenzierung sind wichtig (siehe Kasten „Biologische Daten“). Derartige Daten werden tausendfach erhoben. Aber wem gehören diese Forschungsdaten eigentlich? Wer ist überhaupt befugt, Daten zu teilen und zu nutzen und wo gibt es Einschränkungen? Leider sind „biologische Daten“ im juristischen Sinne nicht eindeutig definiert.

**Förstner:** Bei öffentlich finanzierter Forschung sollte für die anfallenden Daten gelten: „Open per Default – standardmäßig offen“. Die Forschungsdaten sollten damit Teil der Allmende werden. Natürlich gibt es Einschränkungen, zum Beispiel, wenn es sich um personenbezogene Daten handelt und ein öffentlicher Zugang mit dem Datenschutz kollidieren würde.

**Ebert:** Viele der biologischen Daten, mit denen gearbeitet wird, sind reine Mess- oder Beobachtungsdaten ohne die „schöpferische Höhe“, die das Urheberrecht für einen Urheberschutz fordert. Bei einer Veröffentlichung sind das also „freie“ Daten. Trotzdem geht es immer darum: Wem können sie zugeordnet werden? Wer ist befugt, eine Entscheidung über die Weitergabe und Veröffentlichung der erhobenen Daten zu treffen? Das kann der Auftraggeber sein, eine eigenständig forschende Person oder auch ein Verbund von Projektpartnern. In der Praxis sind diese Fragen oftmals nicht sauber geklärt, was das Datenteilen nicht einfacher macht.

**Elbing:** In manchen Fällen liegen Verfügungsrechte aber auch außerhalb der Wissenschaft. Das Übereinkommen über die biologische Vielfalt (*Convention on Biological Diversity*, CBD) legt seit 1992 fest, dass „genetische Ressourcen“ Eigentum des jeweiligen Herkunftslandes sind. Diesem steht es frei, die Erforschung ihres „Eigentums“ zuzulassen oder die Daten mit Einschränkungen zu versehen, die beispielsweise auch die Datenweitergabe oder gar das Publizieren der Forschungsergebnisse betreffen können. Das wird durch Verträge geregelt. Ähnliches gilt für die Nutzung traditionellen Wissens, das mit der genetischen Ressource assoziiert ist. Im schlimmsten Fall kann das quasi einem Verbot der Publikation oder Weitergabe von Daten gleichkommen, was in der Praxis jedoch selten vorkommt. In Zeiten von *Access and Benefit Sharing* (ABS – Zugang und gerechter Vorteilsausgleich) ist „Offenheit“ immer auch verbunden mit der Verpflichtung, den entstandenen „Vorteil“ zu teilen. Im Bereich der Grundlagenforschung kann es dabei nicht um das große Geld gehen, sondern vielmehr um das Gewähren nicht-monetärer Vorteile, wie sie sich etwa aus wissenschaftlichen Kooperationen, gemeinsamen Publikationen oder Trainings (zum Beispiel in Hinblick auf Massenspektroskopie oder visuelle Daten für die Fernerkundung) ergeben.

**Beck:** Daten, die nicht mit öffentlichen Mitteln gewonnen wurden, gehören natürlich dem Auftraggeber, dem Datenerzeuger oder Eigentümer des Materials (s. o.), wenn dies bei der Genehmigung des Projekts so vereinbart wurde. Eine heikle Frage ergibt sich bei der Auftragsforschung, z. B. bei Forschungen, die im Auftrag der Staatsverwaltung (Ministerien) und mit deren Finanzmitteln durchgeführt werden. Nicht immer waren staatliche Dienststellen bereit, die Daten und Erkenntnisse zeitnah zu publizieren. Im Umweltbereich aber müssen die Daten (wenn sie nicht einem anderen besonderen Geheimhaltungsgrund unterliegen) nach dem Umweltinformationsgesetz auf Anfrage veröffentlicht werden. Bei der DFG und anderen Forschungsförderern gilt der Grundsatz, dass mit Mitteln der öffentlichen Hand gewonnene Erkenntnisse auch der Öffentlichkeit zur Verfügung gestellt werden müssen. Embargo-Regelungen sind aber weitgehend akzeptiert, z. B. können die für ein Patent notwendigen Erhebungen oder die im Rahmen einer wissenschaftlichen Qualifikation (z. B. einer Dissertation) erhobenen Daten mit einer Schonfrist versehen werden, deren Dauer vor Projektbeginn vereinbart werden muss. Eine aktuelle Darstellung zum Eigentumsrecht an Daten findet sich bei Frisch [3].

### Open Data in der täglichen Praxis

*Open Data* ermöglicht die Nachnutzung von Daten und damit auch die Generierung neuer Erkenntnisse – etwa, wenn es um historische oder globale Aspekte geht. Soweit die Theorie. In der Praxis ist *Open Data* hingegen mit vielerlei Hürden verbunden. Deren Überwindung erfordert ebenso einen gewissen Aufwand wie die Auflösung von Zielkonflikten in der Datennutzung.

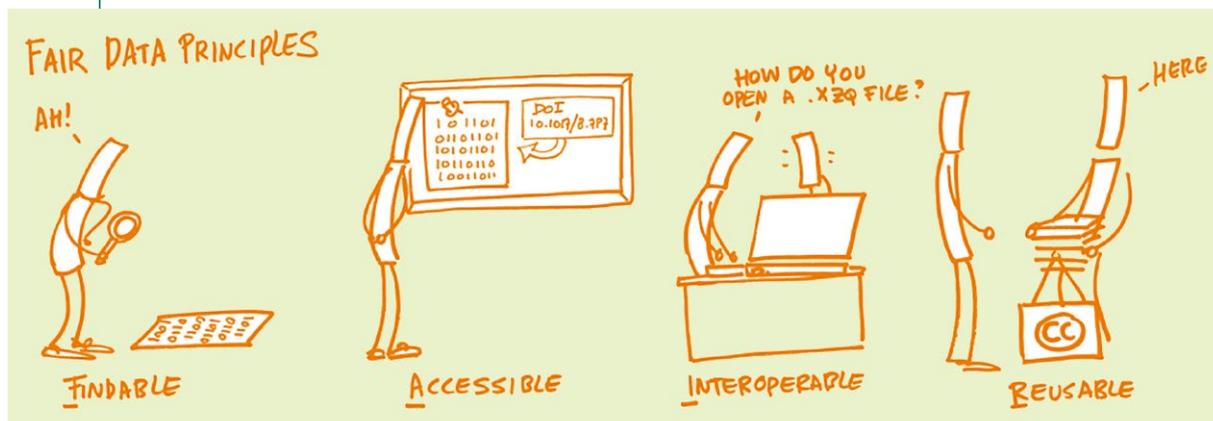
**Beck:** Noch in meiner Postdoc-Zeit war Geheimhaltung von Forschungsergebnissen ein wichtiges Thema, zum einen – wie auch heute – wegen möglicher Patente. Zum anderen war das noch die Zeit (Mitte der 1950er bis Anfang der 70er Jahre) der „großen“ Entdeckungen, die bis heute mit den Namen von Wissenschaftler/-innen verbunden sind wie z. B. der Krebs- oder der Calvin-Zyklus. Hier war das

Konkurrenzdenken der Grund für die Geheimhaltung. Es gab Anweisungen der Chefs, was auf einer Tagung gesagt werden durfte und was (noch) nicht. Im Zuge der heute in der Biologie dominierenden Viel-Autoren-Publikationen hat sich die Einstellung verändert: Man freut sich, wenn man eigene Daten für eine Publikation zur Verfügung stellen kann und in die Autorenliste aufgenommen wird. Nicht selten handelt es sich dabei um Daten, die im Zuge einer anderen Fragestellung erhoben wurden. Ohne *Open Science/Open Data* wäre die Suche nach potenziellen „Datenlieferanten“ eher ein Zufallsgeschäft. Das zeigt auch, wie wichtig eine umsichtige Verschlagwortung und Datenkuration (Entwicklung allgemein nutzbarer Datenformate) ist.

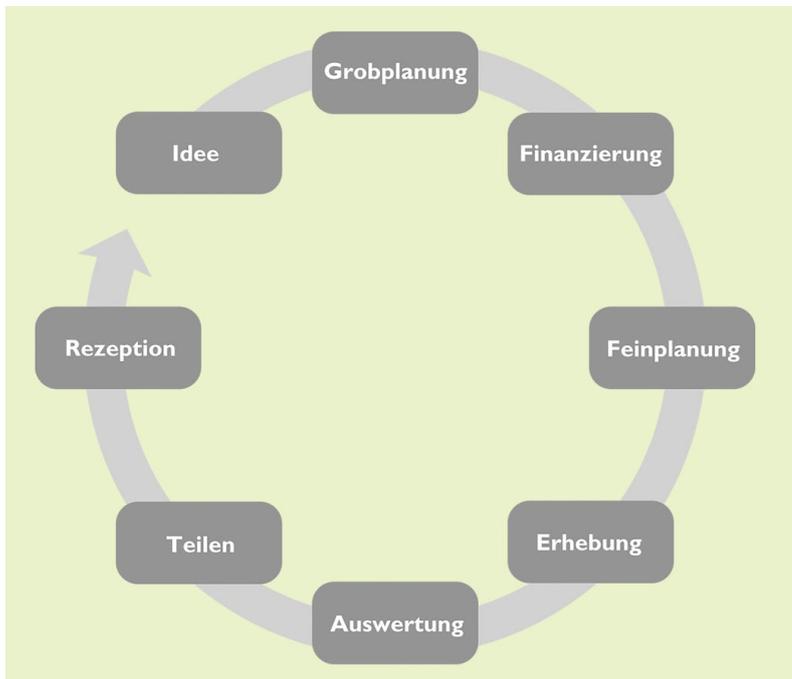
**Förstner:** Datennutzung muss entsprechend der FAIR-Prinzipien erfolgen [5] (Abbildung 2). FAIR steht dabei für *Findable* (Auffindbar), *Accessible* (Zugänglich), *Interoperable* (Interoperabel), *Reusable* (Wiederverwendbar). In der praktischen Anwendung der *Open Science*-Idee und der FAIR-Prinzipien gibt es zahlreiche Hürden. Häufig fehlt das Bewusstsein, dass die eigenen Daten von anderen in einem anderen Kontext nachgenutzt werden könnten und damit ein Mehrwert generiert werden kann. Zudem ist der Umgang mit großen Datenmengen – z. B. sogenannten Omics-Daten – nicht einfach. Daten müssen in Standardformaten vorliegen, was nicht immer der Fall ist. Es müssen entsprechende Datenbanken/Repositoryen verfügbar sein, die es ermöglichen, die Daten durch Metadaten präzise zu beschreiben und somit auch auffindbar zu machen. Die Weitergabe der eigenen Forschungsdaten nach den FAIR-Prinzipien ist häufig mit einem hohen Aufwand für die Forschenden verbunden, ohne dass ein unmittelbarer Nutzen dabei bereits erkennbar ist. Im Rahmen der Nationalen Forschungsdateninfrastrukturinitiative versuchen wir, diesem Problem sowohl durch gezieltes Training als auch durch die Entwicklung von Standards und die Bereitstellung von Speicher- und Analyseinfrastruktur Rechnung zu tragen.

**Ebert:** Im Netzwerk der Biodiversitätsakteure sind Daten unheimlich wichtig, sowohl im praktischen Natur-

ABB. 2 | DIE FAIR-PRINZIPIEN



Daten FAIR nutzen – das heißt, die Daten sind auffindbar, zugänglich, interoperabel und nachnutzbar. Abb. aus [7].



**ABB. 3 Open Science muss sich auf den gesamten Forschungskreislauf beziehen.** Abb.: Oliver Tacke, copyrighted free use.

schutz als auch in der Forschung. Alle hätten gern mehr davon zur Verfügung. Im Einzelfall gestaltet sich dies aber durchaus schwierig – nicht nur wegen ungeklärter Verfügungsrechte, sondern weil es oft an Ressourcen fehlt. Es gibt auch Zielkonflikte, zum Beispiel in Gutachten für Bauvorhaben. Landschaftsplaner haben oft viele Daten, die (noch) nicht in die bestehenden Atlanten und Datenkorpora integriert sind. Trotz vieler Fortschritte ist es noch ein längerer Weg zu offenen oder wenigstens auffindbaren Daten im Sinne der FAIR-Prinzipien [6].

### Verändert Open Science das Publizieren?

*Open Access* hat das Publikationswesen stark verändert. Herausgeber und Gutachter können ihrerseits *Open Science* voranbringen. Zum Beispiel werden *Data availability statements* abgegeben bzw. die Daten in Anhängen zu den Publikationen ebenfalls offengelegt.

**Förstner:** Momentan sehen wir noch einen historisch bedingten Bruch. Eigentlich gehören Texte, Daten und Code/Modelle als Einheit zusammen. Allerdings sind Artikel heute pseudodigitale Abbilder einer Druckversion, wie sie seit Jahrhunderten existiert. Das ist sehr schade, denn hier wird viel Potenzial verschenkt. Wenn zum Beispiel der Code, der zur Erstellung von Grafiken genutzt wurde, offen und dokumentiert wäre, so könnten Dritte die Abbildung mit anderer Parametersetzung neu gestalten, was neue Erkenntnisse ermöglichen könnte. Zudem sind die Artikel nicht semantisch aufbereitet. Wenn aber zum Beispiel ein Gen in einem Artikel als solches maschinenlesbar gekennzeichnet wäre, könnte von hier auf Datenbanken mit mehr Informationen verwiesen werden. Umgekehrt

könnten Datenbanken den Artikel mit dem Datensatz verknüpfen. Dies läuft momentan noch manuell oder durch fehleranfälliges Text-Mining. Die Möglichkeiten, die neuere, semantische Technologien bieten, sind gewaltig – wir könnten unser Wissen ganz anders organisieren, erkunden und nachnutzen. Leider sind hier sowohl die wissenschaftliche Community als auch die kommerziellen Verlage wenig bereit zu Experimenten.

**Ebert:** Wann und wie Ergebnisse publiziert werden, fällt immer noch unter die Wissenschaftsfreiheit. Man muss aber schon konstatieren, dass der Druck in Richtung *Open Access* und Datenpublikationen zunimmt, auch durch entsprechende Förderklauseln. Das hat nicht nur positive Effekte: Die Gebühren für eine Publikation in Fachzeitschriften sind in den letzten Jahren fast unerschwinglich geworden; es entwickelt sich ein „Markt“ für digitale Publikationen und zum Wirtschaftsmodell der Verlage gehören nun auch die Forschungsdaten. Das führt dazu, dass immer mehr „digitale Objekte“ publiziert werden wie z. B. Artikel oder Tabellen und Grafiken in diversen Repositorien. Man hat das Gefühl, das Forschungskarussell dreht sich immer schneller. Die Qualitätssicherung wird immer schwieriger und eine finanzielle Entlastung der Bibliotheksetats ist durch *Open Access* auch nicht wirklich in Sicht.

**Beck:** Wissenschaftliche Publikationen sollen kurz und knackig sein. Wenn aber Unmengen von Daten in die Publikation einfließen, ist dieses Postulat nicht zu erfüllen. Deshalb sind die meisten Journale dazu übergegangen, der eigentlichen Publikation noch eine Rubrik *Supplementary* oder *Supporting Information* (SI) anzuhängen, die vom Seitenumfang nicht beschränkt wird und viele der Originaldaten enthält. Sie ist über das Internet zugänglich. In der Originalpublikation wird dann an entsprechender Stelle auf die SI hingewiesen. Oftmals ist diese SI länger als die Publikation. Das Zitat der Publikation umfasst auch die angehängte SI-Datei. Diese Dateien sind mittlerweile auch zunehmend in maschinenlesbarer Form, was im Sinne der FAIR-Prinzipien sehr zu begrüßen ist.

### Wissenschaftliche Literatur und Daten sind aber nicht alles – was gehört noch zu Open Science?

**Förstner:** Bei der Öffnung der Wissenschaft muss man alle Teile des Forschungskreislaufs bedenken (Abbildung 3). *Open Access* betrachtet nur das finale Produkt, die Publikation. Neben den Daten gibt es auf dem Weg dorthin aber zahlreiche Komponenten, die ebenfalls offen sein sollten. Das fängt eigentlich schon bei der Idee an. Aber auch bewilligte Drittmittelanträge, die genutzte Software, der Workflow der Datenanalyse (mit allen seinen Parametern) und die Abschlussberichte sollten offen verfügbar sein. Auch der Weg vom Manuskript zur Publikation kann und wird weiter geöffnet. So wird das Ablegen von Manuskripten vor dem Peer-Review auf Preprint-Servern immer weiter zur gängigen Praxis. Es gibt auch Journale, die den Peer-

Review offen gestalten und das Feedback der Fachkolleg/innen mit den Artikeln veröffentlichen.

**Ebert:** *Citizen Science* ist für mich noch wirkliches Neuland; ich habe aber in den letzten drei Jahren einige Projekte kennenlernen dürfen, in denen die Wissenschaft zusammen mit Freiwilligen Daten sammelt und auswertet – zum Beispiel die „Herbonauten“ am Botanischen Garten in Berlin oder das Projekt „*Ocean Sampling Day*“, bei dem auch viele Schulklassen mitmachen.

**Elbing:** Ja, gerade im Bereich der Biodiversitätsforschung gelingt es oft erst durch *Citizen-Science*-Ansätze ausreichend Proben oder Beobachtungsdaten zu generieren. Der Ansatz hat aber auch Grenzen – gerade, wenn bestimmte Erfahrungen oder Fertigkeiten erforderlich sind. Qualitätskontrolle ist daher insbesondere bei *Citizen-Science*-Projekten notwendig. Bei Laboruntersuchungen können auch Sicherheitsvorgaben die Beteiligungsmöglichkeiten von Bürgerwissenschaftler/-innen einschränken. Sicher braucht es mehr Phantasie und Mut, *Citizen Science* weiterzudenken. Ich denke, wir brauchen da auch ein gewisses Erwartungsmanagement bezüglich der Chancen und der Grenzen von *Citizen-Science*-Aktivitäten.

**Beck:** *Citizen Science* ist sicher eine Herausforderung. Ein anderes Problem sehe ich in der sog. grauen Literatur, das sind Publikationen, die nicht in einem anerkannten wissenschaftlichen Journal mit Review-System erscheinen. Die meisten von ihnen haben keine DOI, sind aber für eine lokale oder regionale Community durchaus wertvoll und basieren oft auf Daten z. B. von *Citizen-Science*-Projekten – aber: „Kein Impact-Faktor – keine zählbare Publikation“. Vielfach sind die Autor/-innen solcher Aufsätze nicht mit dem modernen Datenwesen vertraut. Hier sehe ich noch Entwicklungsmöglichkeiten.

**Elbing:** *Last but not least:* Zu *Open Science* gehört eigentlich auch *Open Innovation* (Abbildung 4). Gemeint ist die Öffnung des Innovationsprozesses für die Welt jenseits der eigenen Organisation mit dem Ziel, das Innovationspotenzial zu vergrößern und den Innovationsprozess zu beschleunigen. Das hört sich in der Theorie gut an und war 2015 auch Bestandteil des „3-O-Konzept“ (*Open Sci-*

*ence, Open Innovation, Open to the World*), das der damalige EU-Kommissar Carlos Moedas als *Vision for Europe* propagiert hatte [8]. Leider sieht es so aus, dass *Open Innovation* noch nicht mal im wissenschaftsnahen Bereich funktioniert – geschweige denn im kommerziellen Bereich. Insgesamt scheint mir die EU mittlerweile mit *Open Science* deutlich vorsichtiger geworden zu sein. Die FAIR-Prinzipien, bei denen Auffindbarkeit und transparente Zugangsregelungen im Vordergrund stehen, scheinen wesentlich konsensfähiger.

### Was tut sich in Deutschland und Europa?

**Förstner:** Das Thema *Open Science* ist mittlerweile in der Wissenschaftspolitik angekommen. Es gibt Länder, die sehr ambitionierte Ziele verfolgen wie die Niederlande [9]. In Deutschland sind wir noch nicht ganz so weit – der Koalitionsvertrag der Ampel-Regierung adressiert aber zumindest das Thema Forschungsdaten und erklärt das Ziel, *Open Access* zum Standard zu machen [10]. Die Deutsche Forschungsgemeinschaft (DFG) und andere Wissenschaftsorganisationen haben *Open Science* als wichtigen Teil der Forschungskultur anerkannt. Trotz der vielen Dokumente und Empfehlung zu dem Thema ist *Open Science* in der wissenschaftlichen Praxis bei weitem noch nicht erreicht. Klares Problem sind die etablierten Bewertungsverfahren und die Anreizstruktur im kompetitiven Wissenschaftssystem.

**Beck:** Forschungsprojekte, die durch die DFG gefördert werden, müssen zum Abschluss einen Bericht nach einem bestimmten Format abliefern. Das gilt wohl auch für andere Forschungsförderer. Diese Berichte werden von Gutachtern gelesen und kommentiert. Meistens wird die Annahme des Berichts empfohlen. Neuerdings hat die DFG beschlossen, diese Berichte als Publikationen zu betrachten. Denn nicht immer werden die erhofften Resultate und Erkenntnisse erzielt; aber gerade daraus können wiederum andere Antragsteller lernen und deshalb will die DFG diese Berichte publik machen, was ich sehr begrüße.

**Elbing:** Die biowissenschaftlichen Fachgesellschaften entdecken den Themenkomplex „*Open Science*“ erst nach und nach. Das Ganze ist eher punktuell und oft auf den Aspekt „*Open Access*“ fokussiert – etwa bei der Weiterentwicklung des eigenen Fachjournals. Im Vergleich zu anderen Wissenschaften gibt es ja schon viele biowissenschaftlichen Datenbanken, die auch ihre Tradition haben. Da engagieren sich aber – Beispiel *Global Biodiversity Information Facility* (GBIF) – eher Institutionen und weniger die Fachgesellschaften. Ich denke, dass die Diskussion um digitale Sequenzinformationen bzw. eine Regulierung des Zugangs zu den INSDC-Datenbanken ein wichtiger Anstoß auch für die Fachgesellschaften ist, sich intensiver als bisher mit *Open Data* und *Open Science* zu beschäftigen. Wenn man etwas über den Tellerrand schaut, hat man den Eindruck, dass es – wie bei der Deutschen Gesellschaft für Psychologie – wohl erst einer krisenhaften Zuspitzung (im Fall der Psychologie der Replikationskrise) bedarf, die das Grundverständnis des Faches betrifft. „Die eine“ *Open-Sci-*



**ABB. 4** Das Prinzip „Offenheit“ reicht weit über den Wissenschaftsbereich hinaus. Abb. Pexels auf Pixabay CC0.



**ABB. 5 Der Blick in die Kristallkugel: Wie entwickelt sich Open Science?**  
Abb.: Jürgen auf Pixabay CC0.

### OPEN SCIENCE RECOMMENDATION DER UNESCO

Mitte November 2021 hat die Mitgliederversammlung der UNESCO die Empfehlungen zu Open Science (OS) angenommen. Sie enthalten eine internationale vereinbarte Definition des Begriffs „Open Science“. In einem mehrjährigen Konsultationsprozess wurden zahlreiche regionale und nationale Open-Science-Initiativen, daten- und ordnungspolitische Konzepte zusammengeführt.

Unter Open Science versteht die UNESCO das Bestreben

- mehrsprachliches Wissen für alle verfügbar und wiederverwendbar zu machen (offenes Wissen),
- die wissenschaftliche Zusammenarbeit und den Informationsaustausch zum Nutzen von Wissenschaft und Gesellschaft zu verbessern,
- die Prozesse des wissenschaftlichen Erkenntnisgewinns, der Evaluierung und Kommunikation für die gesellschaftlichen Akteure außerhalb der Wissenschaft zu öffnen.

Die UNESCO-Empfehlung bezieht sich auf alle Arten und Disziplinen der Wissenschaft und fußt auf den Pfeilern:

- offenes Wissen
- offene Infrastrukturen für die Wissenschaft
- offene wissenschaftliche Kommunikation
- offenes Engagement gesellschaftlicher Akteure
- offener Dialog mit allen anderen Wissenssystemen

Die Empfehlung schlägt sieben konkrete Maßnahmen für die Etablierung von Open Science (OS) vor.

1. Förderung eines gemeinsamen Verständnisses von OS, inklusive ihrer Vorteile und Herausforderungen
2. Entwicklung eines günstigen politischen Umfelds für OS
3. Investitionen in offene Wissenschaftsinfrastrukturen und -dienste
4. Investitionen in Bildung, digitale Kompetenz und Kapazitätsaufbau für OS
5. Förderung einer Kultur der OS und Ausrichtung der Anreize dafür
6. Förderung innovativer Ansätze für OS in den verschiedenen Phasen des wissenschaftlichen Prozesses
7. Förderung der internationalen und Multi-Stakeholder-Zusammenarbeit mit dem Ziel, digitale, technologische und Wissenslücken zu schließen

Das Papier soll helfen, digitale, technologische und Wissenslücken weltweit zu schließen und ist völkerrechtlich bindend [13].

ence-Strategie der Biowissenschaften sehe ich angesichts der Vielfalt der Fächer, ihrer spezifischen Ansätze und Organisationsstrukturen so schnell nicht. Das ist bedauerlich, denn den Fachgesellschaften kommt auch im Kontext der UNESCO-Empfehlung (Kasten „Open Science Recommendation der UNESCO“) eine gewisse Vorreiterrolle zu.

**Ebert:** Die Wissenschaft kann sehr profitieren von den Open-Data-Bemühungen in angrenzenden Sektoren, vor allem der öffentlichen Hand. Der *Open Data Act* der Europäischen Union [10] und das Forschungsdatengesetz [12] sind sehr spannende Entwicklungen, was die Datenverfügbarkeit in der Forschung anbelangt.

### Ein Blick in die Zukunft: Wo stehen wir in 10–20 Jahren?

**Förstner:** Ich hoffe, dass wir Forschung in den nächsten Dekaden mit viel weniger Reibung und Brüchen, dafür mit viel mehr Offenheit in allen Dimensionen durchführen werden (Abbildung 5). Experimente und andere Datenakquise werden programmiert und automatisch von Maschinen durchgeführt, was eine hohe Skalierbarkeit und Reproduzierbarkeit ermöglicht. Die dabei anfallenden Daten werden mit einem reichhaltigen Satz an biologischen und technischen Metadaten direkt in öffentlichen Datenbanken abgelegt. Forschende besitzen ein hohes Maß an *Data Literacy* und können diese Daten effizient zusammenführen, um ihre jeweiligen Fragestellungen beantworten zu können. Auf Grund der Offenheit können wir KI-Systeme trainieren, die den Forschungsprozess unterstützen. Das dabei generierte Wissen ist offen für alle Menschen verfügbar und durch smarte Systeme leicht auffindbar. Somit werden auch die Grenzen zur Gesellschaft weiter abgebaut – Forschungsergebnisse fließen zum Beispiel viel schneller in die Behandlung von Krankheiten ein. Gleichzeitig können durch eine zunehmende Digitalisierung von Alltag und Medizin auch Daten aus der Gesellschaft in die Forschung fließen. Das Thema „offene Wissenschaft“ hat sich dann erledigt, denn es ist einfach nur noch gängige, gute wissenschaftliche Praxis.

**Elbing:** Die notwendigen Infrastrukturen werden sich ebenso rasant entwickeln, wie die schon genannte *Data Literacy* zur Grundvoraussetzung wird. Dabei wird es natürlich Unterschiede geben, denn die Ausgangspositionen unterscheiden sich im globalen Vergleich doch erheblich. Da kann man nur hoffen, dass die bereits existierenden Ungleichheiten nicht noch größer werden und alle Wissenschaftler/-innen teilhaben können. Aber selbst wenn man optimale Voraussetzungen für alle annimmt, hege ich doch Zweifel, ob Forschungsergebnisse dadurch wirklich in jedem Fall effizienter gewonnen werden und schneller in die Anwendung gelangen als bisher. Gerade, weil reichhaltige, gut gepflegte Datenbanken so ein riesiges Potenzial haben und von Maschinen automatisch durchforstet werden können, wird der (politische) Druck zunehmen, den Zugang zu Datenbanken zu regulieren. Wir erleben das ja gerade in Ansätzen beim Thema digitale

Sequenzinformationen (DSI). Da geht es allerdings „nur“ um einen Vorteilsausgleich und nicht um eine generelle Beschränkung des Zugangs (siehe Kasten „Digitale Sequenzinformation“). Restriktive Zugangsregelungen halte ich aber – zum Beispiel unter Gesichtspunkten von *dual use* – langfristig für nicht ausgeschlossen. Ich bin sehr gespannt, wie sich das entwickelt. Fest steht auf jeden Fall: Wissenschaftler/-innen werden völlig anders mit Forschungsdaten umgehen, als wir es heute tun.

**Ebert:** Ich hoffe, dass sie dies sehr selbstverständlich und mit Freude tun können, weil wertvolle Datenschätze gehoben sind und gut nutzbare Werkzeuge vorhanden sind. Heute ist Datenmanagement oftmals noch harte Arbeit.

**Beck:** In den achtziger Jahren enthielten die Autorenrichtlinien der renommierten Zeitschrift *Plant Physiology* die Vorgabe, maximal 30 Referenzen von Bedeutung für die Arbeit zu zitieren. Heute sind 80 oder 100 Zitate selbst für Originalarbeiten keine Seltenheit, ebenso wie Autorenlisten, die länger sind als die eigentliche Publikation. Ich befürchte, dass *Open Science* diesen Trend zur Masse noch befördert. Wir brauchen ein System der Wissensverarbeitung, in dem das Wissen und nicht die Person und ihre Karrierechancen im Vordergrund einer Publikation stehen.

### Zusammenfassung

*Open Science* bedeutet einen Paradigmenwechsel: In Datenrepositorien zugängliche Originaldaten anderer Autor/-innen werden wie selbst erhobene Forschungsdaten für die Bearbeitung von eigenen, meist übergreifenden Fragestellungen verwendet. Dabei handelt es sich nicht nur um Daten einer Publikation, sondern um alle brauchbaren Daten einer Studie, die sog. Primärdaten. Durch das Teilen von Daten entsteht ein potenzieller Mehrwert. Daraus ergibt sich die Pflicht, auch die eigenen Forschungsdaten offen zugänglich und für andere nutzbar abzulegen (Open Data).

Dies bietet – zumal beim Einsatz von Künstlicher Intelligenz – ein enormes wissenschaftliches Potenzial. Damit sich dieses auch entfalten kann, ist es notwendig, mit den Daten nach den FAIR-Prinzipien umzugehen – das heißt, das eigene Datenkonvolut muss auffindbar, zugänglich, interoperabel und nachnutzbar (englisch: findable, accessible, interoperable, and reusable = FAIR) sein.

Die FAIR-Prinzipien liegen sowohl dem EU Data Act als auch dem EU Data Governance Act als den beiden Säulen der europäischen Datenstrategie zugrunde, die in eigenen Gesetzen der Mitgliedsländer einen Niederschlag finden. Forschungsdaten aus den Biowissenschaften haben unterschiedliche Inhalte, etwa Biodiversitäts- und Vorkommensdaten, Taxondaten, umweltbiologische bzw. ökologische Daten, nicht-molekulare Analysedaten sowie molekulare, zumeist Sequenzdaten. Sie werden in unterschiedlichen Datenbanken (Open data resources) verwahrt und stehen allen Forschenden zur Verfügung.

Nicht für alle Daten einheitlich zu beantworten ist die Frage nach den Eigentumsrechten an dem deponierten Datenkonvolut. Dabei spielen Gesichtspunkte wie die Origina-

### DIGITALE SEQUENZINFORMATION

*Seit einiger Zeit wird im Rahmen der globalen Biodiversitätsabkommen kontrovers über den Status digitaler Sequenzinformationen (DSI) diskutiert. Soll der – bisher offene – Zugang zukünftig reguliert werden, um einen gerechten Vorteilsausgleich zu ermöglichen? Und wie wäre ein finanzieller oder auch nicht-finanzieller Ausgleich für die Nutzung von DSI völkerrechtlich abzusichern?*

*Vor dem Hintergrund postkolonialer Strukturen spielen in der Diskussion vor allem wirtschaftliche, politische und rechtliche Aspekte eine Rolle. Die möglichen Konsequenzen für die Wissenschaft sind erst in jüngerer Zeit in den Blick genommen worden. Überlegungen zu Modellen und Mechanismen zur möglichen Umsetzung stecken noch in den Anfängen. Eine endgültige Verständigung darüber, welche Daten unter „DSI“ fallen, steht ebenfalls noch aus. Vor dem Hintergrund, dass die reinen Sequenzdaten ohne Metadaten nur sehr beschränkt nutzbar sind, darf man gespannt sein, welche Daten einbezogen werden.*

*Der VBIO und die vier biowissenschaftlichen NFDI-Konsortien engagieren sich im „Bündnis der universitären und außeruniversitären Biodiversitätsforschung Deutschlands“. Dieses hat deutlich gemacht, dass jegliche DSI-Regelung den freien, ungehinderten Zugang zu DSI-Daten für die wissenschaftliche Forschung nicht einschränken darf. Nötig sind multilaterale und entkoppelte Lösungsansätze, da bilaterale Lösungsansätze bestehende Ungleichgewichte verstärken und Track & Trace erfordern. Rechtssichere Regelungen müssen universell und praxisorientiert für alle DSI-Daten anwendbar sein – sowohl für nicht-kommerzielle als auch für kommerzielle Nutzer. Bezahl- und/oder Registrierschranken würden den weltweiten wissenschaftlichen Datenaustausch, die Datenverarbeitung und damit den Erkenntnisgewinn schnell zusammenbrechen lassen.*

lität der Generierung, der Zweck der Nutzung, der Finanzierung des Forschungsvorhabens sowie der zeitlich definierte Schutz der Daten eine Rolle. Das große Potenzial des freien Datenzugangs ist allerdings nicht ohne Schattenseiten, gerade im Hinblick auf den Missbrauch von an sich nützlichen Biodaten (Dual Use of Concern). Gleichwohl ist „Offene Wissenschaft“ – unterstützt durch eine verbindliche Empfehlung der UNESCO – auf dem besten Weg vom Konzept hin zur gelebten guten wissenschaftlichen Praxis.

### Summary

#### Open data, Open science?

*Open science means a paradigm shift: Deposited in a database, data from other researchers are used like data that have been collected by themselves for new research questions, usually of an overarching nature. This approach does not only apply to data which are selected for a publication but also to all trustworthy data, so-called primary data. Sharing data results in a considerable added value. Thus, researchers are obliged to deposit their own research data as “open data” in public databases. In particular in combination with artificial intelligence, this generates an enormous scientific potential. To enable its unfolding, it is nec-*

## LISTE AUSGEWÄHLTER ABKÜRZUNGEN

**ABS:** Access and Benefit Sharing  
**CBD:** Convention on Biological Diversity, *CBD*  
**DDBJ:** Data Bank of Japan  
**DFG:** Deutsche Forschungsgemeinschaft  
**DSI:** Digitale Sequenzinformationen  
**EMBL-EMBI:** European Molecular Biology Laboratory-European Bioinformatics-Institute  
**ENA:** Europäische Nukleotiddatenbank (European Nucleotide Archive)  
**EU:** Europäische Union  
**FAIR-Prinzipien:** Findable, Accessible, Interoperable, Reusable  
**INSDC:** International Nucleotide Sequence Database Collaboration  
**GBIF:** Global Biodiversity Information Facility  
**NFDI:** Nationale Forschungsdateninfrastruktur  
**NCBI:** National Center for Biotechnology Information  
**PANGAEA:** Data Publisher for Earth and Environmental Sciences  
**PDB:** Protein Data Bank  
**SI:** Supplementary oder Supporting Information  
**UNESCO:** United Nations Educational, Scientific and Cultural Organization

essary that those data researchers provide must meet the principles: findability, accessibility, interoperability, and reusability (FAIR).

These principles underlie the EU Data Act as well as the EU Data Governance Act, both of them representing the pillars of the European Data Strategy, and are reflected in the individual laws of the European member states. Research data in biosciences deal with different contents such as: biodiversity and occurrence data, taxon data, environmental and ecological data, non-molecular analysis data, and molecular – predominantly – sequence data. They are stored in domain-specific databases and are at the scientific community's free disposal.

A difficult matter is the right of data ownership, as this question cannot be answered easily. Several aspects must be considered, e.g. ingenuity/originality of the scientific project and its approach, the purpose of the use of the data, the funding of the project, as well as the protection of the data (e.g. for patents) for a defined period of time. Admittedly, the great potential of free access to data is not without dark sides, especially with regard to the misuse of bio-data that are actually useful (Dual Use of Concern). Nevertheless, Open science, supported by the binding UNESCO Recommendation, is well on the way of becoming a lived and good scientific practice.

## Schlagworte

Open Data, Open Access, Open Science, FAIR-Prinzipien, Forschungsdaten, Biologische Daten

## Literatur

- [1] H. M. Berman et al. (2012). The Protein Data Bank at 40: reflecting on the past to prepare for the future. *Structure* 20(3), 391–6, <https://doi.org/10.1016/j.str.2012.01.010>
- [2] J. Maxson et al. (2018). The Bermuda Triangle: The Pragmatics, Policies, and Principles for Data Sharing in the History of the Human Genome Project. *Journal of the history of biology* 51 (4), 693–805, <https://doi.org/10.1007/s10739-018-9538-7>
- [3] K. Frisch (2023). Fortschritt wider die Fairness? Ethische Betrachtungen zum Umgang mit Forschungsdaten. *Laborjournal, Essays*, 14.07.2023
- [4] German Federation for the Curation of Biological Data (Knowledge Base): Major Types of Biological Data. Last update October 2023, <https://t1p.de/gfemi>
- [5] M. Wilkinson et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018, <https://doi.org/10.1038/sdata.2016.18>
- [6] N. Enke et al. (2012). The user's view on biodiversity data sharing – Investigating facts of acceptance and requirements to realize a sustainable use of research data. *Ecological Informatics* 11, 25–33, <https://doi.org/10.1016/j.ecoinf.2012.03.004>
- [7] S. Bezjak et al. (2018). Open Science Training Handbook (1.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.1212496>  
CC0 1.0 Universal Public Domain Dedication.
- [8] <https://www.fosteropenscience.eu/content/open-innovation-open-science-open-world-vision-europe>
- [9] NPOS2030 Ambition Document, Version 0.91, 21 April 2022, <http://www.doi.org/10.5281/zenodo.7010402>
- [10] [https://www.spd.de/fileadmin/Dokumente/Koalitionsvertrag/Koalitionsvertrag\\_2021-2025.pdf](https://www.spd.de/fileadmin/Dokumente/Koalitionsvertrag/Koalitionsvertrag_2021-2025.pdf)
- [11] [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_23\\_3491](https://ec.europa.eu/commission/presscorner/detail/en/ip_23_3491)
- [12] <https://www.bmbf.de/bmbf/shareddocs/kurzmeldungen/de/2023/03/230306-forschungsdatengesetz.html>
- [13] UNESCO 2021. „UNESCO Recommendation on Open Science“. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>

## Verfasst von:



Kerstin Elbing



Erwin Beck



Barbara Ebert



Konrad U. Förstner

Die Lebensläufe unserer Autor/-innen finden Sie ausnahmsweise am Beginn dieses Artikels.

## Korrespondenz

Kerstin Elbing  
VBIO – Geschäftsstelle Berlin  
Luisenstraße 58/59  
10117 Berlin  
E-Mail: [elbing@vbio.de](mailto:elbing@vbio.de)



Verband | Biologie, Biowissenschaften  
& Biomedizin in Deutschland

**GEMEINSAM  
FÜR DIE**

**BIEWISSENSCHAFTEN**

### **Gute Gründe, dem VBIO beizutreten:**

- Werden Sie Teil des größten Netzwerks von Biowissenschaftlern in Deutschland.
- Unterstützen Sie uns, die Interessen der Biowissenschaften zu vertreten.
- Nutzen Sie Vorteile im Beruf.
- Bleiben Sie auf dem Laufenden – mit dem VBIO-Newsletter und dem Verbandsjournal „Biologie in unserer Zeit“.
- Treten Sie ein für die Zukunft der Biologie.



[www.vbio.de](http://www.vbio.de)

**Jetzt beitreten!**

