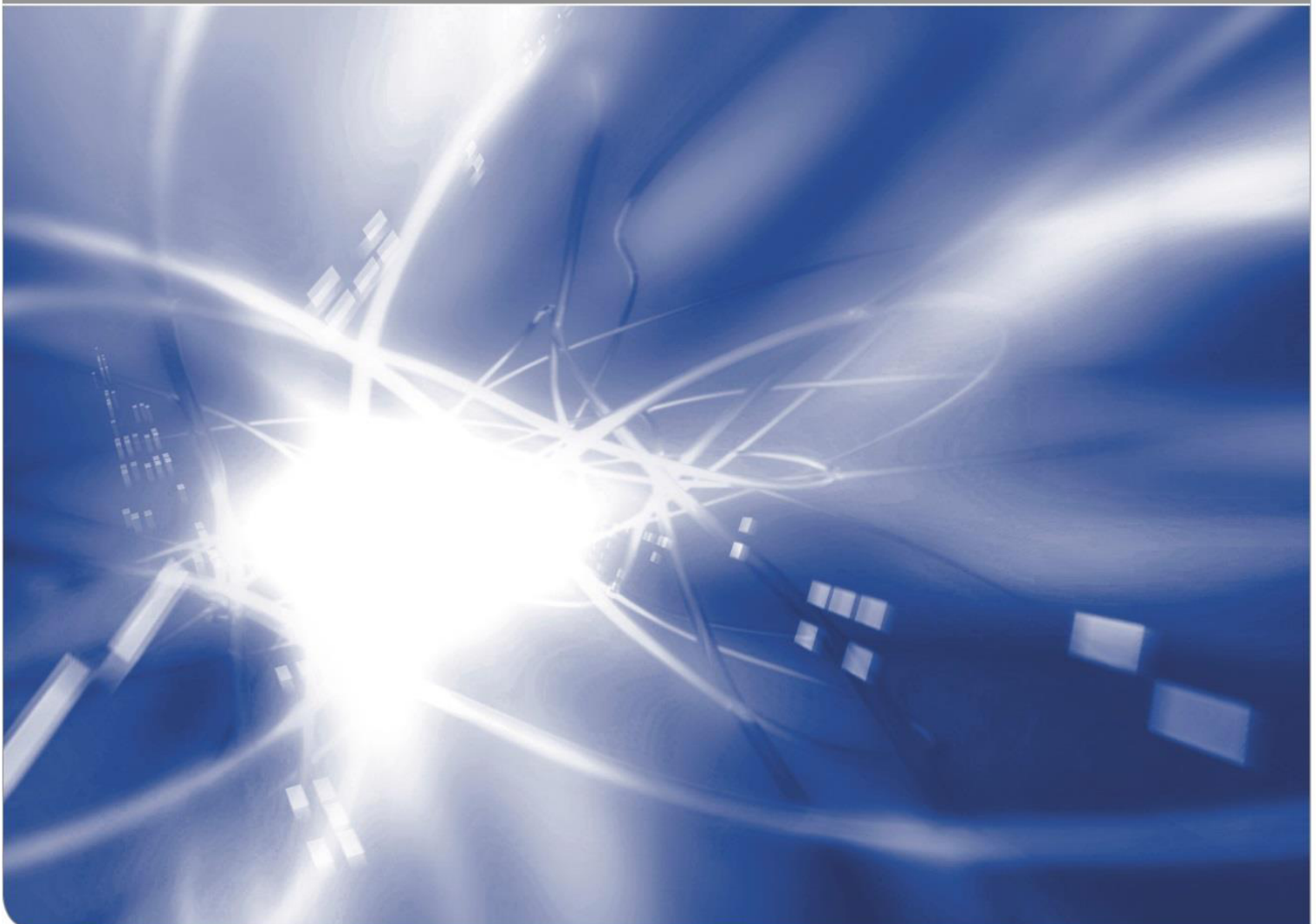


Handreichung Forschungsdatenmanagement in der Sportwissenschaft

von Claudia Albrecht¹, Klaus Bös¹, Lars Schlenker¹, Regine Tobias², Maike van Wasen², Nadin Weiß³, Alexander Woll¹

KIT SCIENTIFIC WORKING PAPERS 54



1 Institut für Sport und Sportwissenschaft

2 KIT Bibliothek

3 Zentrale Einrichtung für Informationsverarbeitung und Kommunikation –
ZEIK, Universität Potsdam

Diese Arbeit ist im Rahmen des DFG Projekts MO|REdata: eResearch Infrastruktur für
Motorikforschungsdaten, DFG Förderlinie wissenschaftliche Literaturversorgungs- und
Informationssysteme entstanden

Institut für Sport und Sportwissenschaft (IfSS)
Engler-Bunte-Ring 15, Gb.40.40
76131 Karlsruhe
<https://www.sport.kit.edu/>

Impressum

Karlsruher Institut für Technologie (KIT)
www.kit.edu



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung –
Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz (CC BY-SA 3.0 DE):
<http://creativecommons.org/licenses/by-sa/3.0/de/>

2016

ISSN: 2194-1629 DOI: 10.5445/IR/1000061538

INHALT

Abbildungs- und Tabellenverzeichnis	II
1 Warum ist das Thema wichtig für mich als Sportwissenschaftler?	3
2 Welche Art von Forschungsdaten fallen in der Sportwissenschaft an?	5
2.1 Das Projekt MO RE data	6
2.2 Was sind Metadaten?	6
3 Wer erfasst sportwissenschaftliche Forschungsdaten?	8
3.1 Welche sportwissenschaftlichen Forschungsdatenbanken gibt es?	8
3.2 Wo gibt es sportwissenschaftliche Forschungsdaten?	8
4 Wie sollten die Daten archiviert und zugänglich gemacht werden?	11
4.1 Wie findet man die Daten?	12
5 Welche Daten sollen archiviert werden - alle oder nur eine Auswahl?	13
6 Gibt es Qualitätskriterien für Forschungsdaten und Qualitätskriterien für digitale Forschungsdatenrepositorien?	14
6.1 Qualitätskriterien für Forschungsdaten in MO RE data	14
6.2 Qualitätskriterien für Forschungsdatenrepositorien	14
7 Wie kann ich Forschungsdaten zitierbar machen? Wie zitiere ich Forschungsdaten von anderen?	16
7.1 Das System des Digital Object Identifiers	16
7.2 Innovative Verwendung von DOIs in MO RE data	17
8 Wie verwalte ich meine anfallenden Daten?	18
8.1 Was ist ein Datenmanagementplan?	18
8.2 Und wieviel kostet das alles?	18
9 Vorsicht im Umgang: Nutzungsbedingungen, Datenschutz und Lizenzen für sportwissenschaftliche Forschungsdaten	20
10 Wo finde ich weitere Informationen?	21
11 Wie geht es weiter im FD-Wesen und was ist der Ausblick für das Projekt MO RE data?	23
Quellenverzeichnis	III

Abbildungsverzeichnis

Abb. 1: Modell eines Forschungsdatenzklus.....	5
Abb. 2: MO RE data Logo.....	6
Abb. 3: Logo re3data.org.....	11
Abb. 4: Übersicht der Qualitätskriterien in MO RE data.....	14

Tabellenverzeichnis

Tab. 1: Bekannte Datenbanken und Portale mit Daten zur motorischen Leistungsfähigkeit.....	10
---	----

1 Warum ist das Thema wichtig für mich als Sportwissenschaftler?

Forschen dient nicht dem Selbstzweck. Forschung sollte den Anspruch haben, die eigenen Forschungsergebnisse anderen verfügbar zu machen. Traditionell wurden Forschungsergebnisse in Schriftform, als Zeitschriftenartikel, in gedruckten Büchern oder später in Online-Publikationen verbreitet. Die während des Forschungsprozesses erhobenen Daten wurden in der Regel nicht publiziert. Das heißt, die Daten konnten nicht von Gutachtern¹ überprüft werden. Heute können die während des Forschungsprozesses erhobenen Daten nicht nur archiviert, sondern auch Dritten frei, Open Access, zur Verfügung gestellt werden. Open Access bezeichnet „den unbeschränkten und kostenlosen Zugang zu wissenschaftlicher Information“ im Internet (Open-access.net, 2016). Die Grundsätze des Open Access in der Wissenschaft wurden schon 2003 in der Berliner Erklärung² aufgegriffen, die inzwischen von allen wichtigen deutschen Forschungsinstitutionen wie der DFG, der MPG und der Fraunhofer-Gesellschaft unterzeichnet wurde. Die Langzeitarchivierung und die Zurverfügungstellung von Forschungsdaten ist Teil guter wissenschaftlicher Praxis und wird immer mehr von Forschungsförderern eingefordert. Denn nur so kann die Grundlage begutachtet werden, auf Basis derer Wissenschaftler Aussagen treffen und nur so kann die Qualität der Forschung neutral und umfassend begutachtet werden.

Die Langzeitarchivierung und die Zurverfügungstellung von Daten eröffnen Forschern weitere Perspektiven um an Wissen zu gelangen, das man für die eigene Forschung benötigt. Veröffentlicht ein Forscher den gesamten Datenbestand, der sich während des Forschungsprozesses angesammelt hat, können andere Forscher darauf aufbauen, anderen Fragestellungen nachgehen und weitere Schlüsse daraus ziehen, ohne notwendigerweise selbst weitere Daten erheben zu müssen. Nicht zuletzt kann der Interpretationsansatz der Forschungsergebnisse untersucht werden und eine qualitative Einschätzung gegeben werden.

¹ Die maskuline Form schließt im Folgenden aufgrund der besseren Lesbarkeit auch die feminine Form ein.

² Siehe <http://openaccess.mpg.de/Berliner-Erklärung>

Gutes Forschungsdatenmanagement kann weiterhin die wissenschaftliche Reputation erhöhen, denn Veröffentlichungen bei denen zugehörige Forschungsdaten veröffentlicht wurden, werden häufiger zitiert (Piwowar & Vision, 2013). Es ist ebenfalls interessant zu bemerken, dass die „Entwicklung bibliometrischer Verfahren, die die Leistung der Daten-Produktion ähnlich den Impact-Faktoren bei Textpublikationen messbar machen“ zu einer höheren Relevanz von Datenzitationen beiträgt (ZBW, GESIS & RatSWD, 2014). Darüber hinaus verlangen immer mehr Institutionen, welche die Forschung finanzieren, oder kommerzielle Verlage ohnehin, dass Wissenschaftler ihre zur Publikation gehörigen Forschungsdaten mit veröffentlichen. Beispielsweise wird auch im EU-Programm „Horizon 2020“ in einer Pilotphase erwartet, dass geförderte Projekte einen Datenmanagementplan anlegen und Ihre Forschungsdaten als Open Access zur Verfügung stellen (EUROPEAN COMMISSION, 2016).

Forschungseinrichtungen, Zuwendungsgeber, wissenschaftspolitische Gremien oder Verlage formulieren „data policies“ als Richtlinien zum Umgang mit wissenschaftlichen Daten.³ Die Richtlinien enthalten Aussagen zur sicheren Speicherung und nachhaltigen Archivierung der Daten, zur Veröffentlichung und Nachnutzung und zur persistenten Identifizierung und Zitierbarkeit der Daten. Ziel ist allgemein die Transparenz der Forschung, insbesondere Nachvollziehbarkeit und Überprüfbarkeit (Bibliothek.kit.edu, 2014).

Insofern bleibt den Wissenschaftlern kaum eine andere Wahl, als sich mit dem Thema Forschungsdaten auseinanderzusetzen. Die vorliegende Broschüre soll die Wissenschaftler in diesem Prozess unterstützen. Der Schwerpunkt liegt auf Informationen zu Forschungsdaten im Bereich der sportwissenschaftlichen Motorikforschung. Neben allgemeinen Informationen zu Forschungsdaten wird im Speziellen auf das Projekt MO|RE data eingegangen und erläutert, wie in dem Projekt ein Forschungsdatenrepository (FD-Repository) aufgebaut wird.

³ Vergleiche zum Beispiel die „data policies“ von Nature (Nature.com, 2016), PLOS (Journals.plos.org, 2016) oder BMJ (Bmjopen.bmj.com, 2016).

2 Welche Art von Forschungsdaten fallen in der Sportwissenschaft an?

Unter dem Begriff Forschungsdaten versteht man „Daten, die im Zuge wissenschaftlicher Vorhaben z. B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen.“ (Allianzinitiative.de, 2014).

Der Großteil an Forschungsdaten in der Sportwissenschaft wird durch Messungen erhoben, z. B. bei Eignungsprüfungen, Tests zu motorischen Fähigkeiten (vgl. MoMo, DMT), Vergleichstest, Schulprüfungen etc. Es wird beispielsweise gemessen, wie weit ein Medizinball geworfen, wie schnell eine bestimmte Strecke laufend zurückgelegt oder wie viel Wiederholungen einer Bewegungsaufgabe (z. B. Sit-ups, Liegestütz etc.) in definierter Zeit absolviert wurden. Daneben werden Forschungsdaten in der Sportwissenschaft auch im Rahmen von qualitativen Befragungen (Interviews, Umfragen) erhoben.

Übergeordnet wird auch vom Forschungsdatenzyklus gesprochen. Dabei gibt es aber „keinen allgemeingültigen einheitlichen Begriff eines Forschungsdatenzyklus“, und die Ausprägungen verschiedener Darstellungen eines Forschungsdatenzyklus werden „von der Fachdisziplin geprägt (...), aus der ein Ansatz stammt.“ (DARIAH-EU, 2016). Für die Sportwissenschaft existiert bislang kein spezifisches Modell eines Forschungsdatenzyklus. Daher wird hier nur ein grobes, relativ allgemein gehaltenes Modell (Abb. 1) vorgestellt.



Zu Beginn des Forschungsdatenzyklus steht in diesem Modell die Erhebung der Daten, z. B. in Form von Experimenten. Die Daten, oder nur eine Auswahl davon, werden dann zumeist in eine Infrastruktur zur Speicherung übernommen. Um die Daten langfristig verfügbar zu machen werden Erhaltungsmaßnahmen durchgeführt. Als letzter Punkt im Forschungsdatenzyklus steht dann der Zugriff auf die Daten, entweder durch die Forscher selbst, andere Forscher oder die interessierte Öffentlichkeit.

2.1 Das Projekt MO|RE data

Das DFG-geförderte Projekt MO|RE data baut bis 2016 eine allgemein zugängliche eResearch-Infrastruktur für sportwissenschaftliche Motorikforschungsdaten auf. Im Mittelpunkt stehen Daten von ausgewählten normierten Testaufgaben mit großem Ver-



Abb. 2: MO|RE data Logo

breitungsgrad. Weiterhin generiert MO|RE data qualitativ hochwertige Normdaten und veröffentlicht begleitende Materialien zu sportmotorischen Tests. Im Projekt wird zwischen Rohdaten und aggregierten Daten unterschieden. Bei den aggregierten Daten handelt es sich um Zusammenfassungen von Rohdaten, die z. B. in Zeitschriftenartikeln veröffentlicht wurden (Sport.kit.edu, 2016).

2.2 Was sind Metadaten?

Metadaten sind Daten über Daten. Neben den eigentlichen Forschungsdaten werden in einer Datenbank bzw. einem FD-Repository auch Metadaten erfasst. Diese können z. B. sein: Autor, Erfassungsdatum, Informationen zur Lizenz für die Weiterverwendung der Daten, Versionshistorie der Daten usw. Die strukturierte Verwaltung von Metadaten ist wichtig, denn sie ermöglicht das Auffinden der „echten“ Daten durch Suchmaschinen im Internet:

„Metadata is key to ensuring that resources will survive and continue to be accessible into the future. (...) An important reason for creating descriptive metadata is to facilitate discovery of relevant information. In addition to resource discovery, metadata can help organize electronic resources, facilitate interoperability and legacy resource integration, provide digital identification, and support archiving and preservation. (...) Describing a resource with metadata allows it to be understood by both humans and machines in ways that promote interoperability“. (NISO, 2004, S. 1f)

Für Metadaten gibt es verschiedene Formatstandards, z. B. Dublin Core (DC), MARC21, METS, MODS etc. Die Verwendung von Standards zur Beschreibung der Metadaten wird zum Zwecke der Interoperabilität und langfristigen Verfügbarkeit dringend empfohlen.

Im Projekt MO|RE data wurde aus diesem Grund ein Metadatenmodell entwickelt, das sich an Vorgaben von DataCite⁴ und RADAR⁵ orientiert. Im Metadatenschema wird zwischen Metadaten zu übergeordneten Informationsaspekten (z. B. Erfassungsdatum, Dateigröße, Sprache), Metadaten zu einer Studie (z.B. Titel der Studie, Studienmethode) und Metadaten zu einem einzelnen Test (z. B. Datum der Testung) unterschieden.

⁴ DataCite (DataCite.org, 2016) ist „ein internationales Konsortium, das sich zum Ziel gemacht hat, einen einfachen Zugang zu wissenschaftlichen Forschungsdaten zu ermöglichen, die Akzeptanz von Forschungsdaten als legitime und zitierfähige Beiträge der wissenschaftlichen Forschung zu erhöhen und die Archivierung von Forschungsdaten zu unterstützen, um Transparenz und nachhaltige Nutzung für die Forschung zu gewährleisten.“ (Wikipedia.org, 2016a).

⁵ Das Projekt RADAR - Research Data Repositorium - des FIZ Karlsruhe (FIZ Karlsruhe, 2016) hat zum Ziel, eine generische Forschungsdateninfrastruktur aufzubauen und zu etablieren.

3 Wer erfasst sportwissenschaftliche Forschungsdaten?

In Deutschland werden sportwissenschaftliche Forschungsdaten auf kommunaler, Landes- und Bundesebene in verschiedenen Initiativen, Projekten und Studien erfasst. Die Daten werden von Wissenschaftlern, Trainern, Lehrkräften, aber auch in groß angelegten Studien (z. B. KommSport in NRW) erfasst. Aktuell gibt es in Deutschland in der Sportwissenschaft verschiedene Datenbanken mit unterschiedlichen Zielsetzungen.

3.1 Welche sportwissenschaftlichen Forschungsdatenbanken gibt es?

In der Sportwissenschaft finden sich übergeordnet 3 Kategorien von Datenbanken:

1. Sportliteratur-Datenbanken (auch Sammlungen von Kongressbeiträgen und sportwissenschaftliche Fachinformation im Internet)
2. Datenbanken für Trainer und Leistungssportler zur Trainings- und Wettkampfdokumentation
3. Datenbanken mit Forschungsdaten (Forschungsdaten-Repository kurz: FD)

Eine Übersicht zu Datenbanken (1.) mit sport- und trainingswissenschaftlichen Informationen findet sich auf der Internetseite <http://www.iat.uni-leipzig.de/service/datenbanken> des Instituts für Angewandte Trainingswissenschaft (IAT Leipzig, 2016).

Seit einigen Jahren werden Datenbanken für Trainer und Leistungssportler (2.) entwickelt, deren Ziel es ist, Trainern und Leistungssportlern die Trainings und Wettkampfdokumentation zu erleichtern und wichtige Prozessabläufe im Wettkampfsport zu automatisieren. Hierbei stehen die individuellen, personenbezogenen Daten des Leistungssportlers im Vordergrund (Lsb-nrw-dalid.de, 2016).

3.2 Wo gibt es sportwissenschaftliche Forschungsdaten?

Öffentlich zugängliche Datenbanken zur Ablage von Forschungsdatensätzen (3.) zur motorischen Leistungsfähigkeit sind den Autoren in Deutschland bislang nicht bekannt. Daten zur motorischen Leistungsfähigkeit werden zwar in einer Vielzahl von Projekten seit Jahrzehnten national und international erhoben (z. B. Motorik-Modul-Studie; WIAD-Studie; Emotikon-Studie, Berlin hat Talent; Check-Düsseldorf, KommSport, Gesunde Kinder in gesunden Kommunen“ (gkgk), etc.), diese Datensätze sind aber nicht

Teil eines FD-Repositoriums, sondern liegen auf lokalen Datenbanken ohne öffentlichen Zugang. Die Daten werden nach manueller Erfassung üblicherweise in eine Excel-Tabelle oder in SPSS eingetragen. Das heißt, bislang werden sportwissenschaftliche Forschungsdaten in Deutschland nicht zentral gespeichert und aufbereitet.

Auch international lassen sich kaum Initiativen finden, die sportwissenschaftliche Daten in FD-Repositorien öffentlich zugänglich machen wollen. Tab. 1 gibt einen Überblick über bekannte nationale und internationale Datenbanken mit Daten zur motorischen Leistungsfähigkeit. Erste Ansätze finden sich für die Speicherung von Accelerometerdaten (International Children's Accelerometry Database (ICAD)) und Gesundheitsdaten z. B. im Youth data Archive der Stanford University.

Tab. 1: Bekannte Datenbanken und Portale mit Daten zur motorischen Leistungsfähigkeit

Name der Datenbank	Quelle	Institution/ Verantwortlicher	Ziel und Inhalte	Zugang
SpoDat	Referenzpublikation Beck & Bös (1995) Normwerte motorischer Leistungsfähigkeit	Beck, Bös , KIT Karlsruhe IfSS	Sammlung von Daten zur motorischen Leistungsfähigkeit	Kein öffentlicher Zugang
GKGK	https://www.gkgk-online.eu/index.php	Willibald Gebhardt Institut Naul, Strauß, Utesch	Daten zur motorischen Leistungsfähigkeit können mit online-Software gespeichert und ausgewertet werden.	Zugang zum Auswertungstool für Projektbeteiligte. Keine Ausgabe von Datensätzen
DMT 6-18	http://www.deutscher-motorik-test.de	Karlsruher Institut für Technologie: Bös, Woll, Schlenker	Daten zur motorischen Leistungsfähigkeit können mit online Software gespeichert und ausgewertet werden.	Zugang zum Auswertungstool mit Code für Software.. Keine Ausgabe von Datensätzen
International Children's Accelerometry Database (ICAD)	http://www.mrc-epid.cam.ac.uk/research/studies/icad/	University of Cambridge School of Clinical Medicine, Esther van Sluijs	Datenbank für Accelerometer Rohdatensätze welche der Wissenschaft zugänglich gemacht werden sollen	Datensätze können auf Anfrage mit Abgabe einer Vorhabensbeschreibung beantragt werden
FitnessGram	http://www.fitnessgram.net	The Cooper Institute	Daten zur motorischen Leistungsfähigkeit können mit fitness assessment, education and reporting tool“ gespeichert und individuell ausgewertet werden.	Zugang zum fitness assessment, education and reporting tool kostenpflichtig. Keine Ausgabe von Datensätzen
Youth Data Archive (YDA)	http://gardnercenter.stanford.edu/current_initiatives/youth_archive.html	Standford University	Sammlung von Daten zum Gesundheitsverhalten von Schulkindern	Nach Anfrage Ausgabe bestimmter Datensätze

4 Wie sollten die Daten archiviert und zugänglich gemacht werden?

Wissenschaftler suchen für die eigene Forschung nach Literatur in Bibliothekskatalogen oder Zeitschriftendatenbanken oder direkt in Internetsuchmaschinen. Um für die eigene Forschung bereits von anderen erhobene Daten zu finden, kommt es oftmals zur persönlichen Kontaktaufnahme innerhalb der wissenschaftlichen Community oder es gibt vereinzelt Datenbanken mit relevantem Bestand. Viele Quellen allerdings befinden sich nach wie vor im sogenannten Deep Web, d. h. sie werden mit Suchmaschinen nicht gefunden.

Während früher Forschungsdaten z. B. in Tabellen verzeichnet wurden, die nur über den Arbeitsplatz des Forschers zugänglich waren, gibt es heute die technischen Möglichkeiten, die Daten allgemein zugänglich zu machen, nämlich über FD-Repositoryen. Dabei gibt es verschiedene Arten von FD-Repositoryen, z. B. fachspezifische Repositoryen für Forschungsdaten nur aus einer Wissenschaftsdisziplin oder institutionelle Repositoryen für Forschungsdaten einer Institution unabhängig vom Fachgebiet. Der Vorteil an Repositoryen ist, dass die Daten strukturiert abgelegt werden und damit durchsucht werden können. Die Verwaltung von Forschungsdaten in Repositoryen ermöglicht neben der langfristig gesicherten Archivierung den Austausch der Daten und damit die Voraussetzung für wissenschaftlichen Diskurs (vgl. Büttner et al., 2011, S. 101).

Im Sinne des Open-Access-Gedanken sind die Inhalte von FD-Repositoryen zumeist weltweit frei verfügbar. Für die Speicherung von Forschungsdaten sind Repositoryen also der sinnvollere Weg und in verschiedenen Wissenschafts-Communities inzwischen auch gängiger Standard, insbesondere in den Lebenswissenschaften, wo sehr große Mengen an Daten anfallen (z. B. CERN, Klimaforschung). Große Forschungseinrichtungen betreiben immer häufiger eigene FD-Repositoryen, die jedoch in vielen Fällen nur Forschungsdaten archivieren, die von institutseigenen Wissenschaftlern erhoben wurden.

Wer als Sportwissenschaftler seine Forschungsdaten in einem Repositoryum archivieren möchte, kann, über die Plattform re3data (re3data.org,



Abb. 3: Logo re3data.org

2016) ein geeignetes Repository suchen. Re3data verzeichnet über 1300 Forschungsdatenrepositorien (Stand Oktober 2015).

4.1 Wie findet man die Daten?

In der Datenbank kann thematisch gesucht werden. Mit Stand Oktober 2015 lieferte eine Suche nach dem Stichwort "Sport" beispielsweise aber nur eine für die Sportwissenschaft relevantes FD-Repository, das "NCAA Student-Athlete Experiences Data Archive" (NCAA, 2016). Hier will das Projekt MO|RE data also eine Lücke schließen und mit dem im Aufbau begriffenen FD-Repository MO|RE data (<http://www.sport.kit.edu/more/>) Sportwissenschaftlern eine Möglichkeit bieten, ihre Forschungsdaten insbesondere aus dem Bereich der Motorikforschung zu archivieren.

Es empfiehlt sich, für die Langzeitarchivierung geeignete Formate zu verwenden. Die Library of Congress hat eine Richtlinie veröffentlicht, in der zwischen bevorzugten („preferred“) und akzeptablen („acceptable“) Formaten für die Langzeitarchivierung unterschieden wird (Library of Congress, 2016). Die bevorzugten Datenformate zeichnen sich dadurch aus, dass sie weit verbreitet und plattformunabhängig sind.

5 Welche Daten sollen archiviert werden - alle oder nur eine Auswahl?

Welche der Daten, die im Laufe eines Forschungsprojekts anfallen, archiviert werden sollten, ist den Wissenschaftlern selbst überlassen. Die Befürworter einer "open notebook science" plädieren dafür, alle Forschungsdaten sofort öffentlich zugänglich zu machen, sobald sie anfallen (Wikipedia.org, 2016c). Zumindest ist es aber empfehlenswert und gute wissenschaftliche Praxis, wenn alle diejenigen Daten veröffentlicht und archiviert werden, auf Basis derer die Forscher Ihre Schlussfolgerungen gezogen und z. B. in wissenschaftlichen Zeitschriftenartikeln veröffentlicht haben.

Ob die in einem Forschungsprojekt erhobenen Daten über das Projekt hinaus relevant sind, um später anderen Forschern die Untersuchung weiterer Fragestellungen zu ermöglichen, lässt sich im Laufe des Projektes nicht immer sofort beantworten. Das soll aber kein Hinderungsgrund sein, die Daten nicht dennoch langfristig und zitierfähig zu archivieren.

Ein Argument gegen die Archivierung aller im Projekt erhobenen (qualitätsgeprüften) Daten ist jedoch sicherlich eine mögliche eingeschränkte Infrastruktur: bietet die vorhandene Archivinfrastruktur zum Beispiel nicht ausreichend Speicherplatz für alle Daten, kann natürlich eine Auswahl getroffen werden. Ebenfalls könnten urheberrechtliche Einschränkungen ein Hindernisgrund für die Veröffentlichung und Langzeitarchivierung der Daten sein.

6 Gibt es Qualitätskriterien für Forschungsdaten und Qualitätskriterien für digitale Forschungsdatenrepositorien?

6.1 Qualitätskriterien für Forschungsdaten in MO|RE data

Damit Wissenschaftler beurteilen können, ob die in einem Repository vorhandenen Forschungsdaten den Ansprüchen guter Wissenschaftspraxis genügen, können gewisse Kriterien abgefragt werden. In MO|RE data wird die Datenqualität nach folgenden Kriterien bewertet (vgl. Albrecht, 2015):

- Q1: Durchführungsqualität (Studienqualität)
- Q2: Datenqualität (Plausibilitätskontrolle)
- Q3: Zugriffsqualität (Dokumentationsqualität)
- Q4: Expertenrating durch Editorial Board

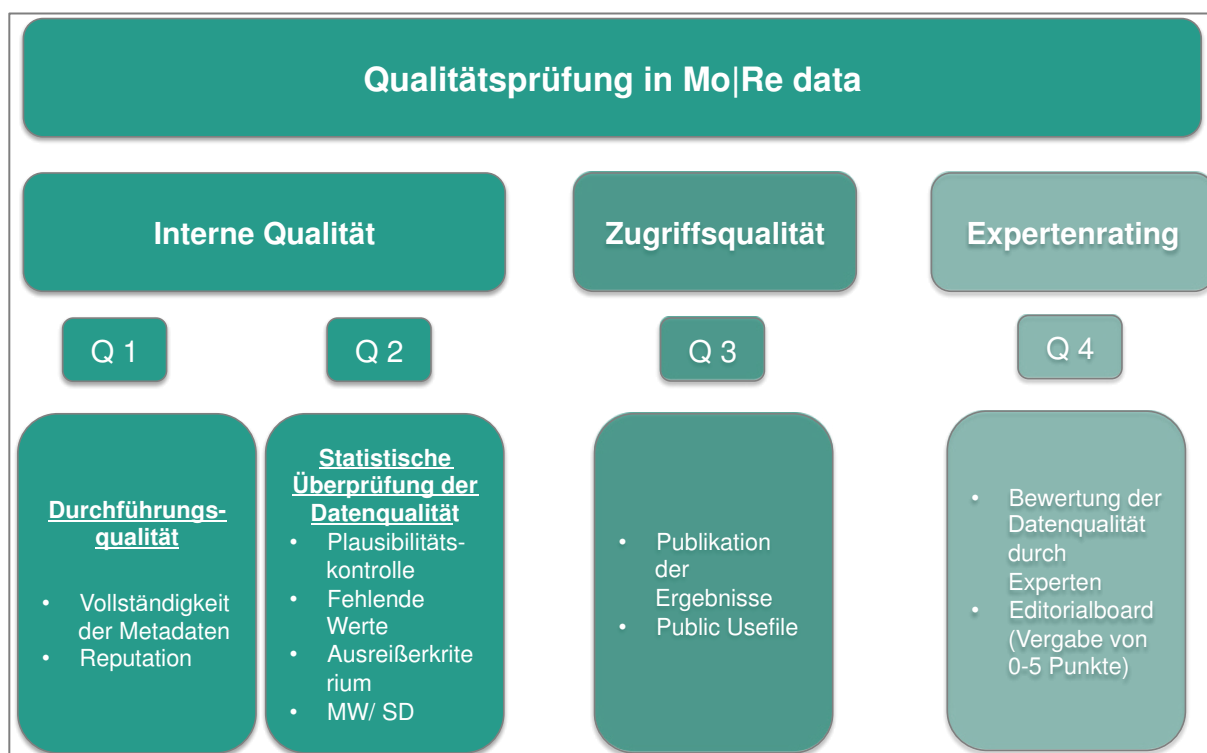


Abb. 4: Übersicht der Qualitätskriterien in MO|RE data.

6.2 Qualitätskriterien für Forschungsdatenrepositorien

Es gibt eine Reihe von Kriterien, welche die Qualität eines Forschungsdatenrepositoriums kennzeichnen. Beim Aufbau eines FD-Repositorys sollte angestrebt werden, diesen Kriterien zu entsprechen. Damit Wissenschaftler beurteilen können, ob ein gewisses FD-Repository vertrauenswürdig ist, sollte z. B. auf „Gütesiegel“ geachtet

werden. Auf der Plattform re3data (Re3data.org, 2016) kann speziell nach Repositorien gesucht werden, die ein Gütesiegel haben. Besonders bekannt sind folgende Siegel:

- Data Seal of Approval, mit 16 Richtlinien, die ein Repository als vertrauenswürdig kennzeichnen (Datasealofapproval.org, 2016).
- Trustworthy Repositories Audit and Certification (TRAC), das 2012 in den ISO Standard 16363 überging, hier bewerten externe Auditoren das Repository auf einer Skala von 1 bis 5 (ISO 16363:2012, 2012).
- In Deutschland wurde außerdem der ISO Standard 31644 von der DINI-nestor-AG entwickelt. Er legt 34 Kriterien zur Bewertung an und es werden Punkte vergeben je nach Erfüllungsgrad eines Kriteriums. Die Erfüllungsgrade sind: „noch offen“, „konzipiert“, „spezifiziert“ und „implementiert“.

7 Wie kann ich Forschungsdaten zitierbar machen?

Wie zitiere ich Forschungsdaten von anderen?

7.1 Das System des Digital Object Identifiers

Damit Forschungsdaten als Ergebnisse von Forschung überhaupt von Dritten in weiteren Publikationen eindeutig identifiziert werden können, müssen sie zitierbar sein. Elektronisch vorliegende Forschungsdaten sollten daher - ähnlich wie Textpublikationen mit einer ISBN oder ISSN - mit persistenten Identifikatoren versehen werden. Es gibt verschiedene Arten von persistenten Identifikatoren. In den letzten Jahren hat sich vor allem der Digital Object Identifier (DOI) durchgesetzt. Im Gegensatz zu URLs (Uniform Resource Locators) verweist ein DOI auf das Objekt an sich und nicht den Ort im Web, wo dieses abgelegt wurde, das heißt:

"es kann über den DOI jedem digitalen Objekt ein oder mehrere URLs zugeordnet werden. Wenn sich der Standort (also der URL) eines digitalen Objektes ändert (z. B. weil der Inhalte-Anbieter seinen Server umstrukturiert), so hat dies keinen Einfluss auf den DOI, es muss lediglich die Zuordnung in der DOI-Datenbank aktualisiert werden." (Wikipedia.org, 2016b).

Die Struktur von DOIs ist wie folgt: 10. Organisation/ID, wobei der Teil „Organisation“ aus Buchstaben und Zahlen bestehen kann. Der erste Teil wird als Präfix bezeichnet, der hintere Teil als Suffix. Das Präfix beginnt immer mit „10.“. Danach folgt die eindeutige Kennung des jeweiligen Datenzentrums. Die Organisationen, die DOIs vergeben, können selbst ein individuelles System für die Vergabe der Suffixe entwickeln. Ein Beispiel:

Paskus, Thomas. NCAA Division I Academic Progress Rate, 2003-2014.
ICPSR26801-v4. Ann Arbor, MI: Inter-university Consortium for Political
and Social Research [distributor], 2015-08-18.
<http://doi.org/10.3886/ICPSR26801.v4>

Bei Eingabe eines DOIs mit Voranstellen der Zeichenfolge <http://doi.org/> in den Suchschlitz eines Browsers, wird man zum Speicherort des über diesen spezifischen DOI identifizierten Objekts weitergeleitet. Über Plug-Ins für Browser können auch die DOIs selbst direkt aufgelöst werden (vgl. ZBW, GESIS & RatSWD, 2014), siehe auch Doi.org, 2016.

Es gibt noch keine einheitlichen Standards für das Zitieren eines Forschungsdatensatzes (vgl. ZBW, GESIS & RatSWD, 2014). Als Orientierung kann man sich an die Vorschläge von DataCite halten (DataCite.org, 2016). Es kann jedoch auch sein, dass man sich an Vorgaben halten muss, die zum Beispiel der Verlag vorgibt, bei dem man publizieren möchte.⁶

7.2 Innovative Verwendung von DOIs in MO|RE data

Auch im Projekt MO|RE data wird auf DOIs als Identifikatorensystem gesetzt. Für Datensätze aus MO|RE data lautet das Präfix 10.18152. Alle Datensätze, die in MO|RE data abgelegt werden, erhalten einen DOI, der über DataCite registriert ist. Nutzer des MO|RE data Portals können diese Datensätze dann über Eingabe des DOI in der Datenbank suchen.

Einen besonderen Mehrwert bietet MO|RE data für Wissenschaftler, indem für eine Auswahl des Datenbestands, den Nutzer zum Beispiel für eigene Publikationen zitieren möchten, ein separater DOI erzeugt wird. Damit können Ausschnitte aus einem sich verändernden Datenbestand persistent identifiziert werden, selbst wenn der Datenbestand sich später verändert.

⁶ Eine ausführliche Anleitung zum Zitieren von Forschungsdatensätzen geben Ball und Duke (2015).

8 Wie verwalte ich meine anfallenden Daten?

Gutes Forschungsdatenmanagement mag aufwändig sein, ist aber im Hinblick auf die Qualität der Forschung und Wissenschaft von hohem Nutzen.

8.1 Was ist ein Datenmanagementplan?

Es wird empfohlen, im Rahmen von Forschungsdatenmanagement auch einen Datenmanagementplan (DMP) zu erstellen und diesen stets aktuell zu halten.

„Ein Data Management Plan (DMP) bezeichnet ein Dokument zur Beschreibung des Lebenszyklus von Daten (insbesondere Forschungsdaten) von der Erhebung bis zur Archivierung, einschließlich aller Maßnahmen, die gewährleisten, dass die Daten verfügbar, nutzbar und nachvollziehbar (verständlich) bleiben. In Deutschland gibt es noch kein vorgeschriebenes standardisiertes Verfahren zur Erstellung eines Data Management Plans. In vielen Ländern sehen die Förderorganisationen inzwischen auch die in den geförderten Projekten gewonnenen Daten als wertvolle Ressource an. Sie verlangen daher von den Antragstellern eine Aussage darüber, wie diese mit den im beantragten Projekt gewonnenen Daten verfahren werden. Diese Aussage wird im Rahmen eines Data Management Plans getroffen.“ (Forschungsdaten.org, 2016)

Es gibt bereits einige Tools, welche die Forscher bei der Erstellung eines DMP unterstützen, z. B. aus dem US-amerikanischen Raum "DMPtool", siehe <https://dmp.cdlib.org/index.html>. In Großbritannien wurde "DMPonline" entwickelt, siehe <https://dmp-online.dcc.ac.uk/> und in Deutschland startete im November 2015 das Projekt "DMPwerkzeug". Die web-basierten Tools sind zumeist noch nicht generisch und nicht immer ohne Umstände auf die eigenen Bedürfnisse anpassbar.

Im Projekt MO|RE data wurde das Tool DMPonline begleitend zu Hilfe genommen, um den Datenmanagementplan zu erstellen. Der DMP von MO|RE data wird demnächst auf der Webseite <http://www.sport.kit.edu/more/> veröffentlicht.

8.2 Und wieviel kostet das alles?

Beim Forschungsdatenmanagement fallen unweigerlich Kosten an, die es z. B. bei Projektbudgets einzuplanen gilt. Wobei es sich hier auch um Kosten handeln kann, die ohnehin Bestandteil des Forschungsprozesses sind (vgl. FD-Broschüre Geowissenschaften S. 12). So sollte dies beim Formulieren eines Projektantrags für eine Förderorganisation zum Beispiel in der Planung des benötigten Personalaufwands bedacht werden. Es muss außerdem bedacht werden, welche langfristigen Kosten für das Forschungsdatenmanagement anfallen könnten. Eine gute Übersicht zu anfallenden Kosten gibt es in der Broschüre „Einstieg ins Forschungsdatenmanagement in den Geo-

wissenschaften“. Dort werden die Kosten aufgeteilt in Personal, Material (z. B. Software), Dienstleistung (z. B. für Beantragung von DOIs) Overhead/Gemeinleistung (Broschüre S. 12) Die Vorteile eines guten Forschungsdatenmanagements rechtfertigen im Allgemeinen den finanziellen und personellen Aufwand.

9 Vorsicht im Umgang: Nutzungsbedingungen, Datenschutz und Lizenzen für sportwissenschaftliche Forschungsdaten

Forschungsdaten insbesondere in den Sportwissenschaften enthalten häufig personenbezogene Informationen: Name, Alter, Geschlecht, Gewicht etc. Die Beachtung von Bestimmungen zum Datenschutz ist daher unerlässlich. In MO|RE data werden daher die Daten bereits vor dem Ingest (Einspielen in die Datenbank) anonymisiert.

Damit andere Wissenschaftler die Forschungsdaten nachnutzen können, ist es wichtig anzugeben, unter welchen Bedingungen die erhobenen Daten weiterverwendet werden dürfen. Wenn man als Wissenschaftler Forschungsdaten z. B. im Zusammenhang mit einer Publikation eines Textes veröffentlicht, kann es sogar sein, dass finanzierende Institutionen oder Verlage Vorgaben zu den Lizenzbestimmungen machen. Die Lizenz sollte beispielsweise regeln, ob bei Weiterverwendung der Daten der Urheber erwähnt werden muss, ob man die Daten für einen kommerziellen Zweck verwenden darf u. ä. Nicht zu vergessen ist, dass die Information zur Lizenz für die Forschungsdaten für Nutzer leicht auffindbar sein sollte.

In der Wissenschaftswelt sind inzwischen Creative Commons-Lizenzen weit verbreitet. Ausführliche Informationen dazu findet man auf der CC-Webseite unter <http://de.creativecommons.org/>.

Andere Lizenzen speziell für Forschungsdaten sind z. B. die Lizenzen von Open Data Commons (opendatacommons.org) und die Open Government Licence (Open Government Licence, 2016).

Im Sinne des Open Access-Gedankens werden möglichst offene Lizenzen empfohlen. Creative Commons bietet dafür z. B. die Lizenz CC-0, CC-BY oder CC-BY-SA an (vgl. Ball, 2014). Im Projekt MO|RE data stimmen Datenlieferanten den Nutzungsbedingungen und damit der Vergabe einer CC-BY-SA Lizenz zum.

10 Wo finde ich weitere Informationen?

An immer mehr Forschungseinrichtungen werden Einheiten begründet, die sich mit Forschungsdatenmanagement speziell an der jeweiligen Einrichtung beschäftigen. Verschiedene Organisationen in Deutschland und international beschäftigen sich übergreifend mit dem Thema Forschungsdaten. Sie veröffentlichen Empfehlungen, Informationsmaterial und Handreichungen. Beispielhaft erwähnt werden sollen hier Folgende:

DataCite, <https://www.datacite.org/>

Das internationale Konsortium DataCite wurde 2009 gegründet und hat heute ca. 30 Mitgliedsinstitutionen. Die Organisation hat zum Ziel „einen einfachen Zugang zu wissenschaftlichen Forschungsdaten zu ermöglichen, die Akzeptanz von Forschungsdaten als legitime und zitierfähige Beiträge der wissenschaftlichen Forschung zu erhöhen und die Archivierung von Forschungsdaten zu unterstützen, um Transparenz und nachhaltige Nutzung für die Forschung zu gewährleisten.“ (Wikipedia.org, 2016a). DataCite vergibt zu diesem Zweck DOIs und die Webseite bietet einige Informationen insbesondere zu DOIs und verweist auf Veranstaltungen.

DINI AG Forschungsdaten, http://www.forschungsdaten.org/index.php/AG_Forschungsdaten

Die Arbeitsgruppe Forschungsdaten der Deutschen Initiative für Netzwerkinformationen e.V. hat sich mit dem nestor-Verbund (Kompetenznetzwerk Langzeitarchivierung in Deutschland) zusammengetan, um „das Forschungsdatenmanagement in Deutschland koordiniert voran [zu] bringen.“ In Kooperation mit weiteren Projektpartnern pflegt die AG u. a. das Wiki [forschungsdaten.org](http://www.forschungsdaten.org) (Forschungsdaten.org, 2016), das zahlreiche Informationen zu Forschungsdaten aufbereitet.

Helmholtz Gemeinschaft, <http://oa.helmholtz.de>

Die Helmholtz Gemeinschaft hat ein Open Science Koordinationsbüro gegründet. Es werden regelmäßig Webinare rund um Forschungsdaten veranstaltet, die nach Anmeldung für jeden offen sind. Außerdem verbreiten die Mitarbeiter relevante Informationen über einen Newsletter und eine Mailingliste und sie pflegen eine bereits sehr ausführliche Literaturliste.

OpenAIRE, <https://www.openaire.eu/>

Das Projekt OpenAIRE wurde im 7. Forschungsrahmenprogramm der EU (Laufzeit 2007 bis 2013) entwickelt. Es hatte zum Ziel, die Implementierung der EC und ERC-Open Access policies zu unterstützen. Über das im Projekt begründete OpenAIRE-Portal soll Forschungsoutput, der mit Hilfe von EU-Förderung erzeugt wurde, sichtbar gemacht werden. Das Folgeprojekt OpenAIREplus ist ein Projekt im Rahmen der Initiative Horizon2020 (Laufzeit 2014 bis 2020). Es zielt u. a. darauf ab, im OpenAIRE-Portal verlinkte Publikationen mit den dazugehörigen Forschungsdaten, Projektinformationen und Autorenangaben zu verknüpfen. Unter <https://www.openaire.eu/guides/> gibt es Hilfestellungen für Wissenschaftler, OpenAIRE veranstaltet außerdem Workshops sowie Webinare.

Open Knowledge Foundation, <https://okfn.org> und <https://okfn.de/>

Die Open Knowledge Foundation Deutschland ist „ein gemeinnütziger Verein, der sich für offenes Wissen, offene Daten, Transparenz und Beteiligung einsetzt.“ (<https://okfn.de/>). Die OKFN bietet in Deutschland verschiedene Trainings und Dienstleistungen im Umgang mit Daten. Die internationale OKFN hat ein umfangreiches „Open Data Handbook“ veröffentlicht, siehe <http://opendatahandbook.org/>.

RDA, <https://rd-alliance.org/>

Die internationale Research Data Alliance wurde im August 2012 gegründet, finanziert von Institutionen aus den USA, der EU und Australien. Die Ziele der Allianz sind:

„The Research Data Alliance (RDA) builds the social and technical bridges that enable open sharing of data. The RDA vision is researchers and innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society“.
([Rd-alliance.org](http://rd-alliance.org), 2016)

Die Allianz hat bereits zahlreiche Empfehlungen zum Forschungsdatenmanagement veröffentlicht ([Rd-alliance.org](http://rd-alliance.org), 2016).

11 Wie geht es weiter im FD-Wesen und was ist der Ausblick für das Projekt MO|RE data?

Ein aktuelles Thema im Bereich Forschungsdaten ist es, Standards dazu zu verbreiten, wie z.B. in Textpublikationen auf den Ablageort von zugehörigen Forschungsdaten verwiesen werden soll. Außerdem beschäftigen sich die Forschungsinstitutionen weiterhin mit geeigneten Datenformaten und den besten Technologien für Infrastrukturen zur Langzeitarchivierung von Forschungsdaten. Nachdem in den letzten Jahren viele Forschungsdatenrepositorien aufgebaut wurden, geht es nun auch darum, die Repositorien miteinander zu vernetzen und Informationen zwischen den Repositorien und mit anderen Informationsdiensten auszutauschen. Nicht zuletzt ist nun auch die Bibliometrie in Bezug auf Forschungsdaten in den Fokus gerückt, d. h. die Erfassung von Zitationen von Forschungsdaten und deren Auswirkung (Impact) auf die Reputation eines Wissenschaftlers.

Das DFG-geförderte Projekt MO|RE data hat eine Erstförderung der Deutschen Forschungsgemeinschaft (DFG) für den Zeitraum von Juli 2014 bis Juni 2016. Eine Anschlussförderung wird angestrebt. Ziele für eine zweite Projektphase sind z. B. eine erweiterte Datenlieferantenbasis und eine Automatisierung der Qualitätsprüfung. Es wird insgesamt angestrebt, das Bewusstsein zum Thema Forschungsdaten in der sportwissenschaftlichen Community in Deutschland und international zu stärken.

Quellenverzeichnis

- Albrecht, C. (2015). *Bewertungskriterien für die Qualität von angelieferten Datensätzen in MOJRE data*. Unveröffentlichtes Manuskript: Karlsruher Institut für Technologie.
- Allianzinitiative.de (2014). *Forschungsdaten*. URL: <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten.html>.
- Ball, A. (2014). *How to License Research Data*. Zugriff unter http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_To_License_Research_Data.pdf.
- Ball, A. & Duke, M. (2015). *How to Cite Datasets and Link to Publications*. Edinburgh: Digital Curation Centre. Zugriff unter: <http://www.dcc.ac.uk/resources/how-guides/cite-datasets#sthash.135vUxch.dpuf>
- BMJ Open (2016). *Resources for data management and sharing*. URL: <http://bmjopen.bmj.com/site/about/resources/datamanagement.xhtml>.
- Büttner, S., Hobohm, H.-C., Müller, L. (Hrsg.) (2011). *Handbuch Forschungsdatenmanagement*. Bad Honnef: BOCK + HERCHEN.
- DARIAH-EU (2016). *Bestehende Konzepte*. URL: <https://de.dariah.eu/bestehende-konzepte>.
- DataCite.org (2016). *Cite your data*. URL: <https://www.datacite.org/services/cite-your-data.html>.
- Datasealofapproval.org (2016). URL: <http://datasealofapproval.org/en/>.
- DMPonline (2016). URL: <https://dmponline.dcc.ac.uk/>.
- DMPTool (2016). URL: <https://dmp.cdlib.org/index.html>.
- Doi.org (2016). URL: <http://doi.org/>.
- EUROPEAN COMMISSION (2016). *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Version 2.1, 15. Februar 2016. Zugriff unter http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
- FIZ Karlsruhe (2016). *Etablierung eines Research Data Repositoriums (RADAR)*. URL: <https://www.fiz-karlsruhe.de/forschung/projekte/radar-research-data-repository.html>.
- Forschungsdaten.org (2016). *Data Management Pläne*. URL: http://www.forschungsdaten.org/index.php/Data_Management_PI%C3%A4ne.
- IAT Leipzig (2016). *Datenbanken*. URL: <http://www.iat.uni-leipzig.de/service/datenbanken>.
- ISO 16363:2012 (2012). *Space data and information transfer systems – Audit and certification of trustworthy digital repositories*. URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510.
- Journals.plos.org (2016): *Data Availability*. URL: <http://journals.plos.org/plosone/s/data-availability>.
- KIT-Bibliothek (2014). *Forschungsdatenmanagement* URL: <https://www.bibliothek.kit.edu/cms/forschungsdatenmanagement.php>.

Library of Congress (2016). *Datasets/Databases*. URL: <http://www.loc.gov/preservation/resources/rfs/data.html>.

Lsb-nrw-dalid.de (2016). *DaLiD* URL: <https://www.lsb-nrw-dalid.de>.

MO|RE data (2016). URL: <http://www.sport.kit.edu/more/>.

National Information Standards Organization (2004). *Understanding Metadata*. Bethesda (USA): NISO Press.

Nature.com (2016): URL: <http://www.nature.com/index.html>.

NCAA (2016). *Student-Athlete Experiences Data Archive*. URL: <http://www.icpsr.umich.edu/icpsrweb/NCAA/index.jsp>.

Open-access.net (2016): URL: <https://open-access.net>.

Open Government Licence (2016). URL: <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>.

Re3data.org (2016). URL: <http://www.re3data.org>.

Rd-alliance.org (2016). URL: <https://rd-alliance.org/about.html>.

Piwowar, H., Vision, T. (2013). *Data reuse and the open data citation advantage*. *PeerJ* 1:e175; DOI: 10.7717/peerj.175.

Wikipedia.org (2016a). *DataCite* URL: <https://de.wikipedia.org/wiki/DataCite>.

Wikipedia.org (2016b). *Digital Object Identifier*. URL: https://de.wikipedia.org/wiki/Digital_Object_Identifier.

Wikipedia.org (2016c). *Open notebook science*. URL: https://en.wikipedia.org/wiki/Open_notebook_science.

ZBW, GESIS & RatSWD (2014). *Auffinden, Zitieren, Dokumentieren. Forschungsdaten in den Sozial- und Wirtschaftswissenschaften*. DOI: 10.4232/10.fisuzida2014.1.

KIT Scientific Working Papers
ISSN 2194-1629

www.kit.edu