

# An Open Repository for Research Data in Acoustics (OPERA)

Stefan Weinzierl<sup>1</sup>, Harald Joachim Kern<sup>2</sup>

<sup>1</sup> *FG Audiokommunikation, Technische Universität Berlin, Email: stefan.weinzierl@tu-berlin.de*

<sup>2</sup> *Universitätsbibliothek, Technische Universität Berlin, Email: kern@tu-berlin.de*

## Einleitung

Eine wichtige Voraussetzung für die Nachvollziehbarkeit und damit letztlich für die Objektivität wissenschaftlicher Erkenntnisse ist die Zugänglichkeit primärer Forschungsdaten. Damit können in der Akustik Messwerte oder Ergebnisse von Hörversuchen gemeint sein, aber auch die für die Versuche verwendeten Stimuli, die Methoden der statistischen Auswertung, numerische Simulationsalgorithmen und wissenschaftliche Software im Allgemeinen. Die Bedeutung dieser Daten ist in jüngster Zeit verstärkt ins Blickfeld geraten, auch bei den wichtigen Förderorganisationen der Wissenschaft. So hat die Allianz der deutschen Wissenschaftsorganisationen bereits 2010 Grundsätze zum Umgang mit Forschungsdaten [1] formuliert. Die EU-Kommission hat im Jahr 2013 im Kontext des Rahmenprogramms Horizon 2020 Richtlinien zum freien Zugang zu wissenschaftlichen Veröffentlichungen und Forschungsdaten [2] und Richtlinien zum Datenmanagement [3] veröffentlicht. Auch die Deutsche Forschungsgemeinschaft hat im Jahr 2015 eine Leitlinie zum Umgang mit Forschungsdaten verabschiedet, die nicht nur Konzepte zum Umgang mit Forschungsdaten bereits bei der Antragstellung einfordert, sondern darüber hinaus die wissenschaftlichen Fachgemeinschaften in die Pflicht nimmt, "ihren Umgang mit Forschungsdaten zu reflektieren und angemessene Regularien zur disziplinspezifischen Nutzung und ggf. offenen Bereitstellung von Forschungsdaten zu entwickeln". [4]

Das OPERA Projekt möchte durch den Aufbau einer technischen Infrastruktur und durch die Anregung einer breiteren Diskussion zum Thema Open Science einen Beitrag leisten, diesem Auftrag nachzukommen.

## Repositorien für Forschungsdaten

Die Veröffentlichung primärer Forschungsdaten erfordert idealerweise ein Forschungsdaten-Repositorium, d.h. eine dezidierte technische Plattform, die eine dauerhafte Zugänglichkeit der Daten und deren bibliographische Identifizierung sicherstellt, sowie ein Rechtemanagement anbietet, das abgestufte Optionen des Zugangs zu den veröffentlichten Daten ermöglicht. Andere Varianten der Datenveröffentlichung sind archivalisch und bibliographisch unbefriedigend, oder entsprechen der Idee eines freien Zugangs zu den erhobenen Forschungsdaten zur unzureichend.

So bieten einzelne Zeitschriften wie das Journal of the Acoustical Society of America (JASA) die Möglichkeit, Daten über das sogenannte "Supplemental material electronic depository" des American Institute of Physics ab-

zugeben. Allerdings ist diese Option eher als erweiterter Anhang zu einer bestimmten Publikation gedacht, etwa für Tabellen, deren Umfang den Rahmen des Artikels sprengen würden. Sie hat nicht den Charakter einer eigenständigen Datenpublikation, die häufig ja Grundlage für ganz verschiedene Publikationen ist, sie ist bibliographisch nicht eindeutig referenzierbar, sie erlaubt nur einen sehr überschaubaren Datenumfang (<10 MB), und sie ist über wissenschaftliche Suchmaschinen nicht auffindbar. Noch weniger befriedigend ist der Verweis auf Internetquellen wie institutionelle oder private Webseiten. Hier ist weder der dauerhafte Zugang gesichert, noch ist gewährleistet, dass die Daten auch nach Jahren noch in unveränderter Form vorliegen, was aber Voraussetzung für die Zitierbarkeit ist. Aus diesem Grund empfiehlt auch die DFG ausdrücklich die Benutzung von institutionell abgesicherten Repositorien für die Sicherung und Bereitstellung von Forschungsdaten [5].

Repositorien hatten ursprünglich häufig einen bibliothekarischen Fokus und wurden entwickelt, um elektronische Publikationen zu sichern und nachzuweisen. Mit der Nutzung als Werkzeug zur Ablage von Forschungsdaten haben sich die Anforderungen an den Funktionsumfang von Repositorien jedoch deutlich erweitert, auch wenn die elektronische Veröffentlichung klassischer Publikationen als Sonderfall immer noch möglich ist. Die zentralen Funktionen von Repositorien für Forschungsdaten sind

- das Angebot der Langzeitarchivierung unterschiedlichster Datensätze mit hoher Datensicherheit,
- die Vergabe eines Persistent Identifier zur Identifikation, Referenzierung und als Zugang zu den Daten, sowie
- den Austausch von Metadaten zur Anbindung an wissenschaftliche Suchmaschinen, um die Auffindbarkeit der Daten zu gewährleisten.

Insbesondere die Vergabe von einem Persistent Identifier (PID) ist ein zentraler Bestandteil jedes Langzeitarchivierungskonzeptes für digitale Inhalte. Im Gegensatz zu Adressierungen wie dem Uniform Resource Locator (URL), die auf den *Standort* eines Datensatzes im Netz verweisen, müssen PID auf das digitale Objekt als Informationseinheit selbst verweisen, auch wenn sich dessen Standort verändert, oder wenn es an mehreren Speicherorten gleichzeitig hinterlegt ist. Beispiele für PID sind das Handle-System [6] oder der Digital Object Identifier (DOI) [7]. Die Struktur eines DOI ist im Standard ISO 26324 beschrieben [8], die Vergabe von DOIs wird seit 1998 durch die International DOI Foundation (IDF) als Dachorganisa-

tion koordiniert und durch eine Reihe von Agenturen wie DataCite als weltweite DOI-Registrierungsagentur für Forschungsdaten organisiert [9]. DOI-Adressen haben die Form “10.ORGANISATION/ID”. So verweist die Nummer “10.14279/depositonce-32” auf das Dokument “depositonce-32” bei der Organisation 14279 (der TU Berlin). Gibt man die URL <https://dx.doi.org/10.14279/depositonce-32> in den Browser ein, löst der DOI-Resolver der IDF die DOI auf und leitet den Webbrowser auf eine Seite um, die Zugang zum Datensatz bietet.

Die Metadaten eines Forschungsdatensatzes sollten in der Regel frei verfügbar und recherchierbar sein. Die Forschungsdaten selbst sind jedoch für eine unmittelbare und vollständige Freigabe häufig nicht geeignet. Es können z.B. Rechte von Verlagen oder Persönlichkeitsrechte zu beachten sein, andere Daten sollen den Wissenschaftler/innen vor einer vollständigen Veröffentlichung zunächst als Basis für weitere Forschungen und Veröffentlichungen dienen. Eine wesentliche Anforderung an eine Forschungsdateninfrastruktur ist daher ein Werkzeug zur Rechteverwaltung, das einen abgestuften und zeitlich – etwa über Embargo-Fristen – konfigurierbaren Zugriff ermöglicht.

Existierende Systeme sind z.T. als disziplinübergreifende Repositorien angelegt, etwa der von der Europäischen Kommission finanzierte Dienst Zenodo [10]. Andere sind disziplinspezifisch, wie Dryad [11] für medizinische oder PANGAEA [12] für geowissenschaftliche Daten. Auch Bibliotheken haben das Management und die Archivierung von Forschungsdaten als Zukunftsaufgabe identifiziert [13] und z.T. eigene, institutionelle Repositorien eingerichtet, wie das System DepositOnce der TU Berlin [14].

## Das OPERA Projekt

Im Gegensatz zu anderen Disziplinen wie der Medizin gibt es in der Akustik bisher weder eine etablierte Kultur im Hinblick auf die Veröffentlichung primärer Forschungsdaten noch eine technische Infrastruktur, die auf spezifische Bedarfe des Fachs, etwa die Vielzahl in der Akustik gebräuchlicher Datenformate zugeschnitten wäre. Aus diesem Grund hat das Fachgebiet Audiokommunikation als Koordinator der Forschergruppe “Simulation und Evaluation akustischer Umgebungen” (SEACEN, FOR 1557) in Kooperation mit der Universitätsbibliothek der TU Berlin Mittel für den Aufbau eines solchen Repositoriums eingeworben, das in Zusammenarbeit mit der akustischen Community und der Forschergruppe SEACEN als Pilotpartner entwickelt werden soll. Als Open Repository for Research Data in Acoustics (OPERA) soll es die langfristige Sicherung und Verfügbarmachung von Forschungsdaten aus der Akustik ermöglichen [15]. Wie zahlreiche existierende Repositorien für wissenschaftliche Daten (Dryad, DepositOnce) soll es auf der Open-Source-Software DSpace [16] realisiert und in die IT-Infrastruktur der TU Berlin integriert werden. Das TUB-Rechenzentrum tubIT sichert seine Systeme durch mehrere Firewalls und eine DMZ (Demilitari-

sierte Zone) gegen Angriffe von außen, um größtmögliche Datensicherheit zu erreichen.

## OPERA Features

Als Persistent Identifier wird OPERA den in den technisch-naturwissenschaftlichen Fächern gebräuchlichen Digital Object Identifier (DOI) verwenden. Hierdurch garantiert OPERA die Referenzierbarkeit der eingestellten Daten über eine Laufzeit von mindestens 10 Jahren. OPERA soll über eine Web-Schnittstelle verfügen, über die nach den Akustik-Daten gesucht und auf diese zugegriffen werden kann. Um eine möglichst weite Verbreitung zu gewährleisten, werden die Metadaten über das Standardprotokoll OAI-PMH (Open Access Initiative – Protocol for Metadata Harvesting) bereitgestellt, so dass sie über andere Suchsysteme und Nachweisinstrumente recherchiert werden können, wie z.B. Google Scholar oder Web of Science. OPERA soll möglichst mit allen in der Akustik verwendeten Datentypen - wie Bild- und Audiodaten, Messwerten, Ergebnissen psychologischer Experimente, statistischen Analysen und auch Software - umgehen und die entsprechenden Dateiformate verarbeiten können. Zudem sollen Nutzerinnen und Nutzer über die Benutzeroberfläche nach spezifischen Daten- und Medienformaten recherchieren, bestimmte Stellen etwa in Audio-Dateien gezielt ansteuern und die entsprechenden Segmente vorhören können. Weiterhin wird ein Zugriffsrechte-Management eingerichtet, das einen differenzierten Zugriff auf die Forschungsergebnisse erlaubt. Auch wenn die Grundintention von Repositorien für Forschungsdaten die *Veröffentlichung* und der freie Zugang zu den Daten ist, wird es in OPERA hierfür keinen Automatismus geben. Die Wissenschaftlerinnen und Wissenschaftler können entscheiden, ob sie den Zugriff auf ihre Forschungsergebnisse oder Teile davon frei zugänglich machen oder diese erst auf eine E-Mail-Anfrage hin zur Verfügung stellen. Darüber hinaus können Embargofristen eingestellt werden, wie dies auch bei elektronischen Veröffentlichungen von Verlagen häufig praktiziert wird. Hierzu wird eine Funktion implementiert, welche einen zeitlich begrenzten Zugriffsschutz erst nach Ablauf einer definierten Frist aufhebt. Lediglich die in OPERA eingestellten Metadaten, welche die Auffindbarkeit der Publikation sicherstellen, sind uneingeschränkt zugänglich.

Ein Anwendungsbeispiel für die Verknüpfung einer Datenpublikation und einer Textpublikation zeigt Abb. 1. Im Text der Zeitschriftenpublikation wird auf die Daten verwiesen, welche die Grundlage der im Paper veröffentlichten Analysen bilden. Auf die über eine DOI referenzierten Daten kann mit dem Browser direkt zugegriffen werden. Sie bilden eine eigenständige, bibliographisch vollwertige Publikation mit eigenem Autorenteam – mit dem für die Wissenschaftler angenehmen Nebeneffekt, dass die Datenpublikationen auch in Zitationsdatenbanken und bei einer wissenschaftlichen Leistungserfassung in der Regel als vollwertige Nachweise anerkannt werden.

#### 4.4. Original Data

The measured and simulated impulse responses for all listener positions (Figure 3), along with the 3D models used for simulation (in DXF and SKP format) are available as an electronic publication [22]. This also contains an anechoic recording of the prologue of *Edipo tiranno* [23] and convolutions with stereo impulse responses recorded in the three theatres with AB stereo miking (1 m microphone basis), suitable for loudspeaker reproduction.



- [22] C. Büttner, F. Schultz, S. Weinzierl: Room acoustical measurements and simulations of Italian renaissance theatres. <http://dx.doi.org/10.14279/depositonce-32>.

**Abbildung 1:** Beispiel für die Verknüpfung von Daten- und Textpublikationen [17]. Im Text der Zeitschriftenpublikation wird auf die Daten verwiesen, als Grundlage der im Paper veröffentlichten Analysen. Auf die über eine DOI referenzierten Daten kann mit dem Browser direkt zugegriffen werden.

## Zusammenfassung und Ausblick

Durch die Entwicklung eines Repositoriums für Forschungsdaten aus der Akustik (OPERA) soll es möglich sein, Forschungsergebnisse der in der DEGA organisierten akustischen Community langfristig zu archivieren, mit Metadaten und einem Persistent Identifier zu versehen und im Internet zur Recherche und Nachnutzung zur Verfügung zu stellen. Durch die Anpassung der Metadaten-Schemata und durch die Entwicklung geeigneter Web-Plugins soll eine präzise spezifizierte Suche und die Online-Überprüfung der für die Akustik relevanten Medienformate möglich sein. Als nächster Schritt der Entwicklung ist eine Ermittlung der Bedarfe und Anforderungen an ein solches System unter den Mitgliedern der DEGA geplant, bevor das System Anfang 2018 seinen Probetrieb aufnehmen soll. Mittelfristig soll dadurch ein aktiver Beitrag zu einem offenen Umgang mit Forschungsdaten aller Art in der Akustik geleistet werden.

## Literatur

- [1] Grundsätze zum Umgang mit Forschungsdaten der Allianz der deutschen Wissenschaftsorganisationen 2010, URL: <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/>, (letzter Zugriff: 26.03.2017)
- [2] Richtlinien zum freien Zugang zu wissenschaftlichen Veröffentlichungen und Forschungsdaten, URL: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf), (letzter Zugriff: 26.03.2017)
- [3] Richtlinien zum Datenmanagement zu wissenschaftlichen Veröffentlichungen und Forschungsdaten, URL: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf), (letzter Zugriff: 26.03.2017)
- [4] DFG Leitlinien zum Umgang mit Forschungsdaten 2015, URL: [http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien\\_forschungsdaten.pdf](http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf), (letzter Zugriff: 28.02.2017)
- [5] Leitfaden für die Antragstellung von Projektanträgen im Bereich Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS) 2016, URL: [http://www.dfg.de/formulare/12\\_01/12\\_01\\_de.pdf](http://www.dfg.de/formulare/12_01/12_01_de.pdf), (letzter Zugriff: 26.03.2017)
- [6] Sun, S., Lannom, L., & Boesch, B. (2003). Handle system overview (No. RFC 3650).
- [7] Paskin, N. (2010). Digital object identifier (DOI) system. *Encyclopedia of library and information sciences*, 3, 1586-1592.
- [8] ISO 26324:2012. Information and documentation - Digital object identifier system.
- [9] DataCite Homepage, URL: <https://www.datacite.org/>, (letzter Zugriff: 26.03.2017)
- [10] Zenodo Repository, URL: <https://zenodo.org/>, (letzter Zugriff: 26.03.2017)
- [11] Dryad Digital Repository, URL: <http://datadryad.org/>, (letzter Zugriff: 26.03.2017)
- [12] PANGAEA - Data Publisher for Earth & Environmental Science, URL: <https://www.pangaea.de/>, (letzter Zugriff: 26.03.2017)
- [13] Bertelmann, R., Pfeiffenberger, H. (2015): Forschungsdaten und Bibliotheken. In: Griebel, R., Schäffler, H., Söllner, K. (Eds.), *Praxishandbuch Bibliotheksmanagement*, Berlin: De Gruyter, p. 639-651
- [14] DepositOnce - Institutionelles Repositorium der TU-Berlin, URL: <https://depositonce.tu-berlin.de/>, (letzter Zugriff: 26.03.2017)
- [15] OPERA - Projektwebseite, URL: <http://opera.tu-berlin.de>, (letzter Zugriff: 26.03.2017)
- [16] DSpace Homepage - URL: <http://www.dspace.org/>, (letzter Zugriff: 26.03.2017)
- [17] Weinzierl, S., Sanvito, P., Schultz, F., Büttner, C. (2015). The Acoustics of Re-naissance Theatres in Italy, *Acta Acustica united with Acustica* 101, 632-641.