

eSciDoc workshop at NIMS

Governmental policies and projects of digital archiving of research outputs in EU, Germany, MPG

Tsukuba, 13. October 2011

Publications
– at least their preprints –
are (secondary) research data

Introduction

What is Research Output ?

- Publications
- Research data

What are relevant Policies?

- Funding rules
- Recommendations, guidelines, mandates of employers
- Laws

What is included in Digital Archiving?

- Data storage
- plus
- Metadata creation and preservation
 - Research process/setting documentation
 - Ensuring self-containment and well-definition of data
 - Ensuring navigability (methods for some-sense-making on archive level, e.g. collection topics, catalog hierarchies, keywords, navigation structures, thumbnails)
 - Curation over the Lifecycle (incl. sustainable funding and legal aspects)

Elements of Digital Archiving of Research Output

Elements	EU	Germany	MPG
Key policy maker & shaping players	<ul style="list-style-type: none"> ▪ EU commission (commissioneer Nelli Kroes) ▪ EUROHORCs / ESF 	<ul style="list-style-type: none"> ▪ DFG ▪ Allianz ▪ DINI ▪ BMBF ▪ Legislator 	<ul style="list-style-type: none"> ▪ MPG President ▪ slnfo / MPDL Beirat ▪ BAR
Policies & key statements	<ul style="list-style-type: none"> ▪ EU commission's Digitale Agenda ▪ WP7 OA/Repository Mandate <p>upcoming:</p> <ul style="list-style-type: none"> ▪ Europe 2020 ▪ WP8 starting 2014 	<ul style="list-style-type: none"> ▪ DFG: Empfehlungen zur Sicherung guter wissenschaftlicher Praxis, 1998 <p>and</p> <ul style="list-style-type: none"> ▪ DFG: Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten, 2009 ▪ Pflichtexemplare / DNBG ▪ Berlin Declaration, 2003 ▪ Allianz Grundsätze zum Umgang mit Forschungsdaten ▪ DINI: Positionspapier Forschungsdaten, 2009 ▪ OECD: Good Laboratory Practice 	<ul style="list-style-type: none"> ▪ MPG statement of Good Scientific Practice
Projects	<ul style="list-style-type: none"> ▪ OpenAire/DRIVER ▪ Parse.Insight ▪ Europeana ▪ Study Cost of European Cultural Heritage Digitization 	<ul style="list-style-type: none"> ▪ BMBF project „eSciDoc“ ▪ Various local and content specific initiatives, e.g. PANGAEA 	<ul style="list-style-type: none"> ▪ MPG institutional repository (permanent funding thru MPDL) ▪ Software/Communities eSciDoc, PubMan ▪ Various local and content specific initiatives, e.g. The Language Archive Nijmegen

EU Policy

EU funding increasingly requires Open Access dissemination in repositories

- Chapter 2.5.2. of the [current] Digital Agenda for Europe – Driving ICT innovation by exploiting the single market – refers to effectively managed knowledge transfer activities and states that **publicly funded research should be widely disseminated through Open Access publication of scientific data and papers.**
- Europe 2020 Flagship Initiative Innovation Union has a similar Open Access clause: the Commission will promote Open Access to the results of publicly funded research; and **it will aim to make Open Access to publications the general principle for projects funded by the EU** research Framework Programmes.
- "The European Commission is conducting a pilot initiative on open access to peer reviewed research articles in its Seventh Research Framework Programme (FP7 [, total budget EUR 50bn over years 2007-2013]). In this pilot, open access to articles resulting from research funded in areas participating in the pilot should be provided within a specified time period. This pilot will run until the end of FP7. This initiative covers approximately 20% of the FP7 budget and applies to the following areas: Energy; Environment; Health; Information and Communication Technologies (Challenge 2: Cognitive Systems, Interaction, Robotics); Research Infrastructures (e-Infrastructures); Science in Society; Socio-economic Sciences and Humanities.
Grant agreements in these areas signed after the beginning of the open access pilot **will contain a special clause requiring beneficiaries**
 1. **to deposit articles** resulting from FP7 projects **into an institutional or subject-based repository;**
 2. **to make their best efforts to ensure open access** to these articles within six months (Energy, Environment, Health, Information and Communication Technologies, Research Infrastructures) or twelve months (Science in Society, Socio-economic Sciences and Humanities). "
- Currently the future, Eighth Research Framework programme is being prepared. It is generally expected that much of the programme, **potentially all areas will require Open Access publication in repositories beginning 2014.**
- The EU-funded Projects OpenAire/DRIVER support participants of WP7 to comply with their OA dissemination requirements
- The EU-funded Project Parse.Insight investigates infrastructure requirements to comply in the future
- EU supports dissemination of Knowledge and Heritage also in other areas, e.g. Europeana and Study Cost of European Cultural Heritage Digitization.

313 signing organizations of the Berlin Declaration

Berliner Declaration, 2003

- „Our mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society. [...] Establishing open access as a worthwhile procedure ideally requires the active commitment of each and every individual producer of scientific knowledge and holder of cultural heritage. Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.“

ESF

ESF mandates 5-10 years of archiving of all primary and secondary research data

All primary and secondary data should be stored in a secure and accessible form.

- Original scientific or scholarly research data should be documented and archived for a substantial period (at least 5 years, and preferably 10 years).
- Research data should be placed at the disposal of colleagues who want to replicate the study or elaborate on its findings.

German National Library (DNB)

DNB mandates any media works including online media works to be delivered to them for archiving

§ 14 Ablieferungspflicht DNBG

„Die Ablieferungspflichtigen haben Medienwerke in körperlicher Form [...] in zweifacher Ausfertigung [...] abzuliefern. [...] Die Ablieferungspflichtigen haben Medienwerke in unkörperlicher Form [...] in einfacher Ausfertigung [...] abzuliefern. [...]“

§ 15 Ablieferungspflichtige DNBG

„Ablieferungspflichtig ist, wer berechtigt ist, das Medienwerk zu verbreiten oder öffentlich zugänglich zu machen und den Sitz, eine Betriebsstätte oder den Hauptwohnsitz in Deutschland hat.“

§ 7 Netzpublikationen PFIIV

„Netzpublikationen sind in marktüblicher Ausführung und in mit marktüblichen Hilfsmitteln benutzbarem Zustand abzuliefern.“

DFG

DFG mandates to archive research data for 10 years with repository grade quality

Empfehlungen zur Sicherung guter wissenschaftlicher Praxis, 1998

- „Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden.“
- „Auf die Aufzeichnungen später zurückgreifen zu können, ist schon aus Gründen der Arbeitsökonomie in einer Gruppe ein zwingendes Gebot. Noch wichtiger wird dies, wenn veröffentlichte Resultate von anderen angezweifelt werden.“

Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten, 2009

„Die nun vorliegenden Empfehlungen richten sich sowohl an die Wissenschaftler als auch die Informationsspezialisten. Es soll damit das Bewusstsein für die Dringlichkeit dieses Problems geweckt werden. Darüber hinaus werden die Ziele benannt, die bei der Umsetzung berücksichtigt werden sollten und deren Einhaltung von grundlegender Bedeutung ist. Schließlich sollen die Wissenschaftler dazu angeregt werden, dieser Aufgabe auf dem Wege der Selbstverpflichtung nachzukommen und für eine sachgemäße Umsetzung im Rahmen der gegebenen Strukturen in eigener Initiative Sorge zutragen.“

1. Forschungsprimärdaten sind Daten, die im Verlauf von Quellenforschungen, Experimenten, Messungen, Erhebungen oder Umfragen entstanden sind. Sie stellen die Grundlagen für die wissenschaftlichen Publikationen dar.
2. Es ist ein fachspezifisches Organisationskonzept festzulegen, das die nachhaltige Speicherung der Daten regelt.
3. Die Speicherung der Forschungsdaten erfolgt im Rahmen definierter Standards
4. Die Daten sind persönlich gekennzeichnet und werden unter dem Namen des Forschers abgelegt.
5. Jeder Wissenschaftler stellt seine Forschungsprimärdaten nach Möglichkeit überregional und frei zur Verfügung.
6. Die Daten werden durch Metadaten beschrieben.
7. Die Daten sind qualitätsgeprüft. „

Helmholtz Society

Helmholtz mandates a minimum of 10 years „documentation and holding available“ for research data

Helmholtz Society, 2007

- „Primärdaten, die Grundlage einer wissenschaftlichen Veröffentlichung sind, werden dokumentiert und *in der Verantwortung der Abteilung oder Arbeitsgruppe* für mindestens zehn Jahre verfügbar gehalten.
- Archivierung von Forschungsdaten [dient] zum Zwecke der **Beweissicherung** und der **Nachnutzung** und es geht zunehmend auch um den **Zugang zu Forschungsdaten** durch Dritte.“

MPG

MPG mandates 10 years of archiving for research data, publication and – as far as possible – Open Access

MPG Guidelines of Good Scientific Practice, 2000 & 2009

- „Allgemeine Regeln für die wissenschaftliche Praxis:
 - genaue Beachtung disziplinspezifischer Regeln für die Gewinnung, Auswahl und Bearbeitung von Daten,
 - zuverlässige Sicherung und Aufbewahrung der Primärdaten für 10 Jahre;
 - eindeutige und nachvollziehbare Dokumentation der angewandten Verfahren (z. B. Laborbuch) und aller wichtigen Ergebnisse“
- „Regeln für die Veröffentlichung von Ergebnissen:
 - prinzipielle Veröffentlichung der Forschungsergebnisse (Prinzip der Öffentlichkeit der Forschung), [...]
 - nach Möglichkeit freie Verfügbarmachung der mit öffentlichen Mitteln erzielten Forschungsergebnisse.“

Various local and content specific initiatives (selected examples Germany)

PANGAEA (AWI und MARUM)
BW-eLabs und LiLa (u.a. U Stuttgart)
D-GRID Initiative
DataCite (u.a. TIB, ZB Med, ZBW)
Earth System Science Data (u.a. AWI)
Economics und NEEO – Networked of European Economists Online (u.a. ZBW)
ICFA Study Group on Data Preservation and Long Term Analysis in HEP (u.a. DESY)
KoLaWiss (U Göttingen)
Konzeptstudie "Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie," (TIB)
PlanktonNet (u.a. AWI)
Scientific Drilling Database - SDDDB (GFZ)
Publication and Citation of Scientific Primary Data - STD-DOI (WDCs und TIB)
WDC-Climate (MPI für Meteorologie und Deutsches Klimarechenzentrum)
WDC-RSAT (Deutsches Zentrum für Luft- und Raumfahrt)
WDC-MARE (AWI und MARUM)
Wibaklidama (FH Potsdam)
World Atlas of Language Structures (MPDL)
...

General use,
not content specific:
eSciDoc

„Summary

- Mandate for Open Access publishing by the EU and MPS is there (currently partially by EU, best effort by MPG)
- Mandate for approx. 10 years of enriched Research Data Archiving is there (often not funded and practiced, yet)
- eSciDoc is a way to address both, for big data a registry is needed (registry is being built)

ENDE