Empfehlungen für Forschungsdaten, Tools und Metadaten in der DARIAH-DE Infrastruktur

Inhalt

- 1 Grundsätzliches
- 2 Dateiformate für Langzeitarchivierung UND Nachnutzung
  - 2.1 Kriterien für die Langzeitarchvierbarkeit
  - 2.2 Kriterien für die Nachnutzbarkeit
  - 2.3 Empfohlene Dateiformate
- 3 Metadatenstandards
  - 3.1 Kriterien für die Eignung von Metadatenstandards
  - 3.2 Administrative, deskriptive Metadatenstandards
  - 3.3 Fachwissenschaftliche Metadatenstandards (Content)
- 4 Tools und Verfahren für die digitalen Geisteswissenschaften
- 5 Empfohlene Lizenzen
  - 5.1 Lizenzen für Content
  - 5.2 Lizenzen für Code
  - 5.3 Lizenzen für Dokumentation

Grundsätzliches

Dieses Seite soll als Empfehlung für interessierte GeisteswissenschaftlerInnen gelten.

Um Kollegen und kommenden Generationen die Beschäftigung mit Ihrer Forschung zu erleichtern, gelten folgende Grundsätze

- Dokumentieren Sie ihre Arbeit
- Verwenden Sie gängige Metadatenstandards
- Verwenden Sie zur Ablage und Übergabe an externe Repositorien Standard-Dateiformate, die auch außerhalb ihrer Community in Gebrauch sind
- Verwenden Sie nach Möglichkeit Normdaten und kontrollierte Vokabulare zur inhaltlichen Anreicherung.

Wir bieten hier eine Übersicht für empfohlene Dateiformate, Metadatenstandards, Lizenzen und Software zur Verwendung in den Geisteswissenschaften und Empfehlungen zu deren langfristigen Erhaltung und Bereitstellung – ohne Anspruch auf Vollständigkeit.

Eines der wichtigsten Kriterien bei der Wahl eines Dateiformats für die eigene Forschung sollte neben der fachlichen Eignung deren **Nachnutzbarkeit** und **Archivierbarkeit** sein. Nicht immer handelt es sich bei beiden um das Gleiche: So stellt die Archivierbarkeit an ein Dateiformat andere Kriterien (gut dokumentierterter, offener Formatstandard, hohe Akzeptanz, viele Metadaten) als deren Nachnutzbarkeit (v.a. Akzeptanz und Verbreitungsgrad der verarbeitenden Software in der Community, **Editierbarkeit der Inhalte**).

Sollte es sich bei den desiderierten Forschungsdaten um solche handeln, welche die Digitalisierung noch vor sich haben, ist unbedingt folgende Empfehlung der DFG zu beachten: http://www.dfg.de/formulare/12_151/12_151_de.pdf

Ein Kriterienkatalog zur besonderen Eignung von Dateiformaten zur Langzeitarchivierung findet sich zum Beispiel bei der Library of Congress: http://www.digitalpreservation.gov/formats/sustain/sustain.shtml. Demzufolge gelten folgende Kriterien als besonders relevant für die Langzeitarchivierung:

Dateiformate für Langzeitarchivierung UND Nachnutzung

Kriterien für die Langzeitarchvierbarkeit

- **Publikation**: der Grad der Publikation bezieht sich auf die Dokumentation einer Dateiformatspezifikation. Je besser ein Dateiformat dokumentiert ist, je genauer definiert ist, welche Software dieses lesen und validieren kann und mit welchen Mitteln, desto eher ist es zur Langzeitarchivierung geeignet.
- **Akzeptanz**: Je breiter ein Dateiformat akzeptiert und verwendet wird – sowohl von Anwendern als auch von Software, desto höher ist die Wahrscheinlichkeit, dass es auch in Zukunft unterstützt wird.
- **Transparenz**: Maßstab der direkten Interpretierbarkeit (durch Menschen). Also inwiefern kann ich eine Datei auch einfach im Texteditor öffnen und daraus Informationen gewinneen?. Das ist im Fall von XML sogar sehr ergiebig, im Falle von JPEG und anderen **komrimierten Dateiformaten** sehr ungünstig. Daher sollte auf die Verwendung von **komprimierten Dateiformaten** möglichst **verzichtet werden**.
- **Kompression**: Wenn unbedingt ein Dateiformat mit eingebetteter Kompression verwendet werden muss (JPEG2000), dann sollte hier eine Kompressionsvariante gewählt werden, die als **verlustfrei** gilt.
- **Selbstdokumentation**: Maß an enthaltenen zusätzlichen Informationen in Form von enthaltenen (deskriptiven, administrativen, strukturierenden) Metadaten, die bei der Interpretation und im Austausch von Archivsystemen helfen
- **Externe Abhängigkeiten**: beschreiben das Maße der Abhängigkeit von bestimmter Hard- oder Software, externen (womöglich patentierten und unfreien) Spezifikationen. Je höher die externe Abhängigkeit, desto schwieriger ist die Langzeitarchivierung von Dateien solchen Typs, da zu deren Interpretation womöglich weitere Komponenten notwendig sind. Diese Kategorie ist speziell für dynamische Multimedia-Inhalte und -Anwendungen relevant.

- **Patenteinschränkungen**: Die gebührenfreie rechtliche Nutzungseinschränkung von Formatspezifikationen kann negative Folgen bei der Verbreitung dieser haben. So schließt eine solche Einschränkung bspw. die Nutzung durch die Open-Source Community aus. Weiterhin können bei gebührenfinanzierten Formatlizenzen unübersehbare Kosten zukommen, weswegen solche Dateiformate von einer Langzeitarchivierung tendentiell ausgeschlossen sind.
- **Technische Schutzmechanismen**, wie Passwortschutz oder Verschlüsselung erschweren den Zugriff auf Dateien und können u.U, diesen vollständig unterbinden. Ohne Zugriff, kann der Inhalt einer Datei aber nicht ausgelesen und anderen zur Verfügung gestellt werden. Auch die Langzeitarchivierung mit der Option auf regelmäßige Formatüberprüfung und -Migration ist somit nicht möglich.

Kriterien für die Nachnutzbarkeit

Die folgende Liste sollte als Ergänzung und Korrektiv zur obigen Liste von Auswahlkriterien zur Langzeitarchivierbarkeit von Dateiformaten bewertet werden und erhebt keinen Anspruch auf Vollständigkeit

- **Editierbarkeit:** Eine Wandlung eines Word Dokuments nach PDF/A zur Archivierung gilt als ungemein sinnvoll, da PDF/A als offen, gut dokumentert und speziell zur Archivierung entwickelt gilt. Nichtsdestotrotz möchte man Forschungsdaten auch aktiv nutzbar machen. Ein einmal erstelltes PDF ist aber nicht mehr editierbar und somit sehr schlecht als Forschungsdatum nachnutzbar. Hier eignen sich andere Dateiformate besser (RTF, ASCII offen Office Formate).
- **Geeignete Publikationsart:** Für jede Art von Inhalt gelten ganz eigene Kriterien, welches der beste Weg ist, diesen zu publizieren, d.h. für andere zugänglich zu machen. Nicht immer ist eine Bibliographie in einem Tabellenkalkulationsprogramm gut aufgehoben, aber auch die direkte Verwaltung von Bibliographien in eigens dafür bereitgestellten Tools (Zotero, BibTex, Word) bringt Tücken mit sich:
    - Was, wenn Zotero in 5 Jahren nicht mehr weiter entwickelt werden kann und die Server abgeschaltet werden?
    - Was wenn die neue Word-Version die Bibliographie eines älteren Office-Dokuments nicht mehr interpretieren kann?
    - Was, wenn auch CSV keine optimale Kodierung einer Bibliographie verspricht?

  Hier sollte sicher gestellt werden, dass man **die eigene Entscheidung für ein Format möglichst bewusst und unter Einbeziehung aller bekannten Variablen** trifft. Die beste Entscheidung gibt es häufig nicht. Im Zweifel hilft es, eine Datei im Quellfomat **UND** in einem weit verbreiteten Exportformat zugänglich zu machen.
- Gute **Trennung von Form und Inhalt, Dokumentation und Administration**: Nur wenn für die Nachnutzenden ersichtlich ist, in welchem Teil der Daten die Inhalte (bspw. Bilddaten, Rohtexte) stecken und in welchem Teil ergänzende Informationen und die alles übergreifende Struktur versteckt sind und wie sie zu verstehen sind, können die bereitgestellten Daten entsprechend nachgenutzt werden und die Beantwortung einer Forschungsfrage nachvollzogen werden.

Die folgende Liste von empfohlenen Dateiformaten ist der Versuch eines Brückenschlags zwischen beiden Verwendungszwecken:

Die folgende Liste von Dateiformaten richtet sich insbesondere an Geisteswissenschaftler und enthält eine Reihe von Dateiformaten, die als für beide Zwecke (Langzeitarchivierbarkeit und Nachnutzung) geeignet gelten können. Für weitere Dateiformate und zum eigenen Nachschlagen empfiehlt sich insbesondere dieser von der Library of Congress zur Verfügung gestellte Service: http://www.digitalpreservation.gov/formats/fdd/browse_list.shtml. Ein Dateiformat, welches dort nicht gefunden wird, ist mit Vorsicht zu genießen oder aber so hoch fachspezifisch, dass es über Communitygrenzen hinaus nicht bekannt sein dürfte und entsprechend wenig unterstützt sein wird.

Empfohlene Dateiformate

Medientyp	Format	Empfohlen weil
Bild	TIFF baseline	Gute Akzeptanz, gute Publikationstiefe, keine Patenteinschränkung oder technische Schutzmechanismen. Der Begriff *Baseline* bezieht sich darauf, dass hier eine Untermenge von formatspezifischen Eigenschaften definiert ist, die von allen Computerprogrammen unterstützt werden müssen, damit diese eine TIFF-Datei lesen können.
Bild	PNG	Gute Akzeptanz, gute Publikationstiefe, keine Patenteinschränkung, weite Verbreitung. Im PNG Standard werden die Daten komprimiert (wenn auch lossless), was zu einer kleineren Dateigröße (optimal zu **Webanzeige**) aber auch zu einer Gefährdung in der Langzeitarchivierung führen kann, da im Falle von Kopierfehlern ganze Codezeilen nicht mehr gelesen werden könn(t)en und die resultierende Bilddatei somit nicht mehr dargestellt werden kann.
Vektorgrafik	SVG	Gute Publikationstiefe, keine Patenteinschränkung oder technische Schutzmechanismen, breite Akzeptanz, hohe Transparenz. Ist **XML basiert** und lässt sich somit auch gut konvertieren.
Text (statisch)	PDF/A	Gute Akzeptanz für den Zweck der **Langzeitarchivierung**, wurde eigens zur verbesserten **Archivierbarkeit** von PDF-Dokumenten erschaffen, sehr gute Publikationstiefe. Wird für den Zweck der **Nachnutzung von Textdokumenten nicht empfohlen**, wenn man nicht allein am "Druckbild" (digitale Inkunabel) Interesse hat. Das Gleiche gilt für die Archivierung von Strukturen (Fußnoten, Register, Paginierungen etc.). Als layoutbasiertes Format erlaubt es **keine eindeutig definierte Umwandlung in XML** oder flexible Visualisierungen oder Textaufbereitung; umgekehrt kann PDF oft leicht aus XML Formaten generiert werden.
Text	ASCII (TXT)	Sehr gute Akzeptanz von allen Betriebssystemen und den meisten Textprogrammen. Bietet allerdings keine Seitenbeschreibung oder Strukturauszeichnung von Text, ist also nicht mit Office oder DTP Dateien vergleichbar.
Text	RTF (RichText Format)	Gilt im Gegensatz zu Word-Dokumenten (Doc, Docx), als vollständig strukturiert, ist **transparent** und von viel Konvertierungssoftware interpretierbar (z.B. als Zwischenformat bei freier ePub-Konvertierungssoftware). Auch ermöglicht es die bessere Nachnutzbarkeit eines Inhalts verglichen mit Word. Der Verbreitungsgrad des Formats nimmt allerdings in der Tendenz ab.
Text / Code	XML	Sehr verbreitet unter Geisteswissenschaftlern. Ist vollständig strukturiert. Die Transparenz von XML ist mittelmäßig, d. h. zwar von Menschen lesbar, aber die Wohlgeformtheit und Validität eines XML Dokuments lässt sich besser maschinell prüfen. XML lässt sich recht gut in andere visuelle Präsentationsformate überführen (PDF, HTML), allerdings NUR, wenn die Wohlgeformt und Validität gegeben sind! Daher ist hierauf und diesbezüglich auch auf die Konformität mit einem bekannten und verbreiteten Profil / Schema besoders zu achten.
Text (Office)	ODT	Das beste Format aus der Familie der Office-Formate. ODT ist ein offener Standard mit hoher Transparenz. Sehr geeignet um die Nachnutzung der eigenen Inhalte zu fördern. Nur bedingt geeignet zur Langzeitarchivierung, da es häufig von gänger Identifizierungs-Software / Extraktionssoftware nicht erkannt wird.
Text (Publikation	TeX	Sehr gute Dokumentation, offener Standard, Verbreitungsgrad in manchen Disziplinen hoch (nicht unbedingt in den Geisteswissenschaften). Die Akzeptanz durch Software ist leider nicht groß, da es sich bei TeX Dokumenten selbst um

plus Formeln)		Code handelt, welcher je nach LaTex Prozessor unterschiedlich interpretiert werden kann (Vgl. http://apsr.anu.edu.au/publications/LaTeX-preservation.pdf).
Spreadsheet / Tabellen	CSV	Comma Separated Values ist ein weit verbreitetes und **offenes Austauschformat für tabellarische Inhalte**. Es kann von den meisten Tabellenkalkulationsprogrammen sowohl gelesen als auch geschrieben werden. Allerdings eignet sich CSV weder zur Darstellung gestalterischer Eigenschaften (Farben, Fonts, etc) noch zur Darstellung komplexer tabellekalkulatorischen Formeln o.ä.
Video	MXF (Material Exchange Format)	Von der Library of Congress explizit empfohlener Dateiformatstandard zur Aufnahme von Audio- und Videostreams. Auch wenn das Format auch jedwede Art von weiteren Bitstreams aufnehmen kann, sollte es – laut Experten – als das digitale Equivalent zur Videokassette gesehen werden (Zitiert nach LOC). Es handelt sich um einen offenen, gut dokumentierten und gut archivierbaren Standard.
Audio	WAV, AIFF	Sowohl AIFF als auch WAV sind sogenannte Pulse-Code-Modulation Verfahren zur Codierung von Audiosignalen. Mithilfe solcher Verfahren werden klassischerweise analoge Audiodokumente digitalisiert. Die daraus resultierenden Dateiformate gelten aber auch als sehr handhabbare und gut dokumentierte Formatstandards zur Speicherung von Audiodaten. WAV wurde von einer Kooperation von IBM und Microsoft entwickelt. AIFF ist das Pendant von Apple. Beide gelten als zwar proprietär aber sehr weit verbreitet und gut dokumentiert. Auch werden beide Standards durch Drittanbieter Software unterstützt.
Noten	Music XML	XML-basierter Standard zur Codierung und Bearbeitung von musikalischen Noten. Es handelt sich um einen offenen, weit verbreiteten Standard der von vielen Programmen aus der Branche unterstützt wird.
Datenbanken	SQL Dump	Relationale Datenbanken im Web werden in der Regel in einer Datenbanksprache, MySQL oder PGSQL o.ä. abgelegt. Mithilfe eines "Dumps" können Auszüge oder der vollständige Inhalt einer Datenbank als Textdatei exportiert werden – was als erste ad-hoc Lösung zur Archivierung allemal Sinn macht. Ein solcher Dump als Datei gehorcht – je nach Datenbanksprache – einer unterschiedlichen Syntax und ist nicht notwendigerweise vollständig dokumentiert. Der große Vorteil eines solchen Exports liegt aber in der menschlichen Lesbarkeits (Transparenz) des Inhalts und darin, dass das Ergebnis auch in jedem Texteditor Informationen preisgibt. Auch lassen sich vollständige SQL Dumps in den gängigsten DB Sprachen gut als Datenbanken importieren.
Datenbanken	SIARD	SIARD ist ein auf XML basierendes Dateiformat, welches vom Schweizerischen Bundesarchiv offiziell zur Langzeitarchivierung entwickelt wurde und zusammen mit einem Softwarepaket – der SIARD-Suite – kostenlos genutzt werden kann. Es erlaubt sowohl den Im- als auch Export in verschiedene Datenbankformate und gehorcht ausschließlich offenen Standards.
3D-Daten	X3D, VRML	X3D ist ein auf XML basierendes 3D Format, welches von den meisten aktuellen Browsern unterstützt wird. X3D ist offen und Bestandteil von MPEG4. VRML gilt als generischeres 3D Dateiformat mit offener Spezifikation, welches von einer Vielzahl von 3D Software interpretiert werden kann und zusätzlich webkompatibel ist.
Software	als Quellcode (unkompiliert), & Dependency Information	Das Feld der Archivierung und Nachnutzung von Software ist relativ gering erforscht. Auch die Library of Congress kann bisher auf keine Empfehlungen verweisen. Grundsätzlich empfiehlt es sich im Falle von selbst geschriebener Software, diese zusammen mit allen technologischen Abhängigkeiten zu dokumentieren und den Code zusammen mit einem möglichst generischen Compilat (d.h. als ausführbares Programm möglichst Betriebssystem-unabhängig) abzulegen. Letzteres ist leider häufig nur allzu stark von der verwendeten Technologie und Sprache abhängig und daher nicht einfach umzusetzen.

Quellen u.a.:

- http://fclaweb.fcla.edu/uploads/Lydia%20Motyka/FDA_documentation/recFormats.pdf
- http://www.data-archive.ac.uk/media/2894/managingsharing.pdf
- http://www.loc.gov/preservation/resources/rfs/TOC.html

Metadatenstandards

Kriterien für die Eignung von Metadatenstandards

Analog zu den Kriterien für die Langzeitarchivierungsfähigkeit und Nachnutzbarkeit von Dateiformaten soll hier eine Liste von Kriterien zur Nutzbarkeit von Metadatenstandards erfolgen. Grundsätzlich gelten dabei ähnliche Kriterien: Sowohl die **Verbreitung** eines Standards als auch der **Grad der Spezifikation / Dokumentation** sind ausschlaggebende Faktoren.

Die folgende Liste enthält alle Überlegungen, welche bei der Wahl eines Metadatenstandards eine Rolle spielen sollen:

- Grundsätzlich sollten Metadaten aus **kontrollierten Vokabularen bzw. Schemata oder Ontologien** zur Vermeidung von Ambiguitäten abgeleitet werden.

- **Verbreitung** in der jeweiligen Fachdisziplin: Wird der entsprechende Metadatenstandard von bekannten Institutionen oder Fachvertretern verwendet und propagiert?
- **Internationalisierung** – handelt es sich um ein nur in der Bundesrepublik eingesetzten Standard mit deutscher Terminologie oder um einen englisch-sprachigen und damit internationaleren Standard?
- Kann der entsprechende Metadatenstandard so verwendet werden, dass die **Objektgranularität** angemessen abgebildet werden kann?
- ...

Bei den folgenden Standards handelt es sich zum einen um Standards des kulturelles Erbes (Lido, Mets, EAD) (Dazu bspw: http://www.langzeitarchivierung.de/Subsites/nestor/DE/Standardisierung/Metadaten.html) zum zweiten sind dies aber auch Fachspezifische Empfehlungen für Daten und Metadaten

Die folgenden Listen geben beobachtete Metadatenstandards sowohl der einzelnen Fachdisziplinen als auch allgemeiner der Gedächnisinstitutionen wieder.

Administrative, deskriptive Metadatenstandards

Die folgende Tabelle führt die **gängigsten** in Bibliotheken / Archiven / Museen verbreiteten Metdadatenstandards auf

Herkunft	Bezeichnung	Link zur Spezifikation / zum Schema
Alle + WWW	DublinCore (DC)	http://dublincore.org/schemas/
Museen	LIDO	http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html
Museen, Kunstegeschichte, Archäologie	CIDOC CRM	http://www.cidoc-crm.org/rdfs/cidoc_crm_v5.0.4_official_release.rdfs
Institutionen des kulturellen Erbes -> Mapping zu Europeana (Archäologie)	CARARE	http://www.carare.eu/swe/Media/Files/CARARE-V2.0.1-XSD
Alle	EDM	EDM ist das Datenmodell der Europeana, welches unterschiedliche Metadatenschemata kombiniert und anreichert, so dass eine Objekt- und Eventbasiertes Perspektive zu Objekten des kulturellen Erbes abgebildet werden kann. http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf
Bibliotheken	METS / MODS	METS: http://www.loc.gov/standards/mets/mets.xsd MODS: http://www.loc.gov/standards/mods/v3/mods-3-5.xsd
Archive	EAD	http://www.loc.gov/ead/ead.xsd
Museen	FRBR	http://vocab.org/frbr/core.html
Kennzeichnung von Provinienz, Langzeitarchivierung	PREMIS	http://www.loc.gov/standards/premis/schemas.html
Kennzeichnung von Provinienz	W3C Prov	http://www.w3.org/TR/prov-overview/
Bilder (Technische Bildeigenschaften, Scans)	NISO	http://www.niso.org/schemas/iso25964/iso25964-1_v1.4.xsd

Fachwissenschaftliche Metadatenstandards (Content)

Objekt- und Medientyp	Disziplin	Standard	Spezifikation/Schema
Text - Noten	Musikwissenschaft	MEI	http://music-encoding.org/documentation/guidelines2013
Text - Handschriften	Kodikologie	Manuscriptum XML (MXML)	http://www.manuscripta-mediaevalia.de/hs/handbuch.pdf
Text - Charters	Geschichtswissenschaft	CEI (inkl. TEI-P4)	
Text	Editionswissenschaft, Judaistik, Geschichtswissenschaften, Papyrologie, Epigraphik	TEI-P5	http://www.tei-c.org/release/doc/tei-p5-doc/de/html/
Objektverzeichnis	Archäologie, Kunstgeschichte	MIDAS	http://www.heritage-standards.org.uk/midas/docs/
Objekte - Grabungen	Archäologie	ArchaeoML	
Flächen, geographische Daten	Archäologie	ADex	http://www.landesarchaeologen.de/fileadmin/Dokumente/Dokumente_Kommissionen/Dokumente_Archaeologie-Informationssysteme/Dokumente_AIS_ADeX/ADeX_2-0_Doku.pdf
Audio, Mulitmedia	Musikwissenschaft, Multimedia	MPEG-7	http://mpeg.chiariglione.org/standards/mpeg-7
Umfragedaten	Sozialwissenschaften, empirische Forschung	DDI	http://www.ddialliance.org/Specification/
Kontrollierte Vokabulare	Alle	XML	Getty-Thesaurus, Personennamendatei, FoF, etc

Tools und Verfahren für die digitalen Geisteswissenschaften

Übersicht über eine Liste von Kriterien für geeignete DH-Software

Anmerkung zu Tools und Verfahren: Die Diskussion beim ersten Arbeitstreffen des Stakeholdergremium Fachgesellschaften hat ergeben, dass es eine Differenz zwischen theoretisch angestrebten Faktoren wie Datenaustauschbarkeit, Standardisierung, Langzeitarchivierung und Publikation und praktisch angewendeten Tools und Verfahren gibt. In der praktischen Arbeit werden oben genannte Faktoren eher als einengend wahrgenommen und verwendet werden **Tools, die greifbar sind oder deren Bedienung bereits bekannt** und erprobt ist. Unten stehende Liste versucht also eher Tools anzuführen, die in der DH Community tatsächlich genutzt werden als Tools, die aufgrund der Umsetzung angestrebter Faktoren vielleicht eher genutzt werden sollten.

Disziplin	Verfahren	Beobachtetes Tool	Empfohlenes Tool
Editionswissenschaften, Geschichtswissenschaften, Kunstwissenschaften	Transkribierungstools	Transkribus (beta)	
Editionswissenschaften	Erstellen von Editionen und Vergleichen von Textversionen	CollateX, TextGrid	
Handschriftenforschung	Erstellen von Editionen und Digitalisierung von Manuskripten sowie Analyse digitaler Handschriftensammlungen	eCodicology, DigiPal,	
Alle	XML Editoren (zur Erzeugung inhaltlicher und struktureller Metadaten)	Oxygen, XMLSpy	Oxygen, XMLSpy
Sprachwissenschaften, Literaturwissenschaften, Editionswissenschaften, Geschichtswissenschaften	Annotationstools (welche konkret TEI oder Open Annotation als Austauschformat implementieren und validieren)	Annotation Studio, CATMA, AnnotatorJS	
Kunstwissenschaften, Geschichtswissenschaften	visuelle Annotation von Bilddaten	Hyper Image,	
Musikwissenschaften	Annotation & Analyse von Musikdaten	MEISE, Augmented Notes, Sonic Visualizer	
Alle	Bildanzeige und -verarbeitungsmöglichkeiten	123D Catch, GIMP, Pixlr, Photoshop	GIMP
Geographie, Geschichte, Kulturwissenschaften, Literaturwissenschaften	geospatiale Darstellung und Verarbeitung von Daten	Geo-Browser, CartoDB, StoryMapJS, Google Earth, ~~Google Maps~~ http://leafletjs.com/ und http://www.openstreetmap.org/about	
Linguistik, Literaturwissenschaften, Geschichte	statistische Verfahren Text (Frequenzanalysen, Corpusvergleichende Analysen, Kollokationsanalysen)	Stylo für R, R, Textal, Textplot, TXM, Voyant Tools, Word2Vec, Textmechanic, Textometrica, Juxta Commons,	
Bildwissenschaften, Kulturwissenschaften, Medienwissenschaften	statistische Verfahren Bild	Image Plot	
Musikwissenschaften	statistische Verfahren Audio		
Linguistik	Lemmatisierung		
Linguistik	PoS-Tagging	CLAWS, Stanford Parser	
Linguistik	logikbasierte Analysen		
Linguistik	Spracherkennung		
Linguistik, Literaturwissenschaften	NER	Stanford NER, NEX,	
	Visualisierung solcher Verfahren und Daten	Visualizing Variation, ManyEyes, RAW, R, Tableau, yEd	
Literaturwissenschaften	Topic Modelling	Mallet, Topic Modelling Tool, In-Browser Topic Modelling, DFR-Browser, FACTORIE	
Literaturwissenschaften, Medienwissenschaften, Kulturwissenschaften, Geschichte	Netzwerkanalyse und -visualisierung	Gephi, Jigsaw, Netlytic, UCINet, Mallet-to-Gephi	
Film-, Medien- und Kulturwissenschaften	Film Anaylse	Cinemetric, ClipNotes, Image Plot	
	Verwendung externer Wissensbasen, also Ontologien und Taxonomien, die mit den entsprechenden Sprachen (RDF, RDF(S), OWL) in maschinenlesbarer Form vorliegen		

Quellen: http://dhresourcesforprojectbuilding.pbworks.com/w/page/69244319/Digital%20Humanities%20Tools

http://lab.softwarestudies.com/p/software-for-digital-humanities.html

http://www.digipal.eu/digipal/page/718/

Empfohlene Lizenzen

Die folgende Liste stellt eine (unvollständige) Übersicht über weit verbreitete und empfohlene Lizenzen im Bereich des Open Access dar.

Die Lizensierung unterschiedlicher Arten von Inhalt unterliegt häufig unterschiedlichen Bestimmungen, weswegen die lizenzdefinierenden Organisationen darauf Rücksicht genommen haben und entsprechend Lizenzen für unterschiedliche Inhaltstypen publiziert haben:

Man unterscheidet gemeinhin zwischen Lizenzen für Content (Texte, Musikstücke, Videos) und Code (Software, Softwarebibliotheken, Standards) und sogar Lizenzen für Dokumentation und weitere Inhaltstypen (Lehrmaterialien, Fonts...) s.u.

Lizenzen für Content

Für die Publikation von Inhalten sind die Lizenzen der Creative Commons weit verbreitet und werden auch von DARIAH-DE empfohlen, weil

- es sich hierbei um international anerkannte Lizenzen handelt, welche in den meisten Ländern in lokales Nutzungs- /Verbreitungsrecht überführt werden können
- diese für freie Verfügbarkeit der Inhalte sorgen
- gleichzeitig die Auswahl zwischen 6 verschiedenen Arten von vorgefertigen Lizenzverträgen besteht, die dem Produzenten der Inhalte eine große Wahlfreiheit lässt, wie genau die Inhalte weiter verwendet werden dürfen.

HINWEIS: Ein interessanter Artikel, der beschreibt, warum die Verwendung von CC-BY-NC (Creative Commons mit nicht kommerzieller Nutzung) mit Vorsicht zu verwenden ist, findet sich hier: http://irights.info/artikel/cc-lizenz-kommerziell-nein-danke/7193

Lizenzorganisation	Version, Art
CC unported	Attribution v1.0(CC BY)
	Attribution Share Alike v1.0(CC BY-SA)
	Attribution No Derivatives v1.0(CC BY-ND)
	Attribution Non-Commercial v1.0(CC BY-NC)
	Attribution Non-Commercial Share Alike v1.0(CC BY-NC-SA)
	Attribution Non-Commercial No Derivatives v1.0(CC BY-NC-ND)
	Attribution v2.0(CC BY)
	Attribution Share Alike v2.0(CC BY-SA)
	Attribution No Derivatives v2.0(CC BY-ND)
	Attribution Non-Commercial v2.0(CC BY-NC)
	Attribution Non-Commercial Share Alike v2.0(CC BY-NC-SA)
	Attribution Non-Commercial No Derivatives v2.0(CC BY-NC-ND)
	Attribution v2.5(CC BY)
	Attribution Share Alike v2.5(CC BY-SA)
	Attribution No Derivatives v2.5(CC BY-ND)
	Attribution Non-Commercial v2.5(CC BY-NC)
	Attribution Non-Commercial Share Alike v2.5(CC BY-NC-SA)
	Attribution Non-Commercial No Derivatives v2.5(CC BY-NC-ND)
	Attribution v3.0(CC BY)
	Attribution Share Alike v3.0(CC BY-SA)
	Attribution No Derivatives v3.0(CC BY-ND)
	Attribution Non-Commercial v3.0(CC BY-NC)
	Attribution Non-Commercial Share Alike v3.0(CC BY-NC-SA)
	Attribution Non-Commercial No Derivatives v3.0(CC BY-NC-ND)
GNU **Design** Science Licence	Design Science Licence (DSL)
Europeana	Europeana: Rights Reserved - Free Access
Free Art License	Free Art License 1.3 (FAL 1.3)
	Free Art License 1.3 (FAL 1.2)
Open Data Commons	Open Data Commons Attribution Licence v1.0(ODC-By)

	Open Data Commons Open Database Licence v1.0(ODC-ODbL)
	Open Data Commons Database Contents Licence v1.0(ODC-DbCL)
Open Government Licence (UK)	Open Government Licence for public sector information
	Non-Commercial Government Licence
Public Domain	CC0
	CC Public Domain Mark
	Open Data Commons Public Domain Dedication and Licence (PDDL)
	ODC Attribution-Sharealike Community Norms

Lizenzen für Code

Vgl: http://opensource.org/licenses/category

Populäre Lizenzen, die weit verbreitet sind oder von starken Communities unterstützt werden
Apache License, 2.0 (Apache-2.0)
BSD 2-Clause "Simplified" or "FreeBSD" license (BSD-2-Clause)
BSD 3-Clause "New" or "Revised" license (BSD-3-Clause)
GNU General Public License (GPL)
GNU Library or "Lesser" General Public License (LGPL)
MIT license (MIT)
Mozilla Public License 2.0 (MPL-2.0)
Common Development and Distribution License (CDDL-1.0)
Eclipse Public License (EPL-1.0)

Lizenzen für ganz bestimmte Zwecke
Educational Community License, Version 2.0 (ECL-2.0)
IPA Font License (IPA)
Open Font License 1.1 (OFL-1.1)

Lizenzen für Dokumentation

Lizenzorganisation	**Version, Art**
GNU Free **Documentation** License	GNU Free Documentation License v1.3(FDL)
	GNU Free Documentation License v1.2(FDL)
	GNU Free Documentation License v1.1(FDL)